Lineare Algebra I/II
2021/22
Universität Regensburg

Marc Hoyois

27. Juli 2022

Inhaltsverzeichnis

Einführung

Diese Vorlesung hat zwei allgemeine Ziele:

- Das erste Ziel ist natürlich die Lineare Algebra kennenzulernen. Die Lineare Algebra ist ein besonderer Bereich der Mathematik, indem sie in fast allen anderen mathematischen Bereichen verwendet wird, von der Analysis bis zu der Geometrie. Das steht zum Beispiel im Gegensatz zu der Analysis und der Geometrie, die nur für einen (wenn auch großen) Teil der Mathematik relevant sind. Die Lineare Algebra ist auch unerlässlich in der theoretischen Physik: Die beiden grundlegendsten Theorien der Physik, nämlich die Allgemeine Relativitätstheorie und die Quantenfeldtheorie, benutzen mehrere fortgeschrittene Begriffe aus der Linearen Algebra. Weiter hat die Lineare Algebra viele verschiedene Anwendungen in der heutigen Zeit, zum Beispiel für numerische Simulationen, Suchalgorithmen, maschinelles Lernen, usw.
- Das zweite Ziel, das sich auch mit der Vorlesung *Analysis* deckt, ist den Begriff des *mathematischen Beweises* kennenzulernern. Insbesondere werden Sie durch die Praxis lernen, was als mathematischer Beweis zählt, solche Beweise zu verstehen und selbst zu schreiben, sowie „einfache“ Beweise selbst herauszufinden.

Obwohl wir die Mathematik auf Deutsch besprechen, ist die mathematische Sprache ganz besonders, und man muss sich daran gewöhnen. Der wichtigste Unterschied zwischen der natürlichen Sprache und der mathematischen Sprache ist, dass in der mathematischen Sprache keine Mehrdeutigkeiten erlaubt sind: Jede Aussage muss eine ganz eindeutige Bedeutung haben, unabhängig von irgendeiner Auslegung.

Mathematische Texte wie dieses Skript bestehen aus folgenden Bausteinen:

- **Definitionen** führen neue Begriffe ein. Definitionen sind besonders wichtig in der Mathematik, denn sie notwendig sind, um den Rest des Textes überhaupt zu verstehen. Deswegen sollen Sie unbedingt alle Definitionen auswendig lernen.
- Es gibt verschiedene Arten von Aussagen:
  - **Sätze** sind Aussagen, die besonders wichtig oder schwierig sind.
  - **Propositionen** sind Aussagen, die nicht besonders schwierig sind.
  - **Lemmata** sind Hilfssätze, die in Beweisen weiterer Sätze verwendet werden.
  - **Korollare** sind Folgerungen vorheriger Sätze.
- Jede solche Aussage soll bewiesen werden, durch einen **Beweis**.
- Es gibt noch **Beispiele** und **Bemerkungen**, die auch wichtig sind.

In diesem Skript gibt es einige Sätze, die nicht bewiesen werden. Diesen Sätzen ist ein Stern vorangestellt: ***Satz**. Es gibt auch ein paar Sätze, die bewiesen werden, aber deren Beweise besonders kompliziert sind. Diesen Beweisen ist dann ein Stern vorangestellt: ***Beweis**. Solche Beweise müssen Sie nicht unbedingt lesen, aber Sie können es als Herausforderung nehmen, sie zu verstehen.

Überblick über die Vorlesung

In den Vorlesungen *Lineare Algebra I* und *II* werden wir folgende Themen studieren:

- Mengentheoretische Grundlagen: Mengen, Abbildungen und Relationen
- Grundlegende algebraische Strukturen: Gruppen und Körper
- Vektorräume
- Lineare Abbildungen
- Matrizen, Matrizenkalkül und die Determinante
- Lineare Gleichungssysteme
- Eigenwerte und Eigenvektoren
- Euklidische und unitäre Vektorräume
- Ringe und Moduln
- Endlich erzeugte Moduln über Hauptidealringen
- Normalformen linearer Abbildungen
- Tensorprodukte

Kapitel 1

Mengentheoretische Grundlagen

1.1 Logische Grundlagen

Eine vereinfachte Sichtweise der Mathematik ist, dass sie sich mit der Bestimmung der Wahrheit bzw. Falschheit von objektiven Aussagen beschäftigt. Beispiele von objektiven Aussagen sind „$1 + 1 = 3$“, „drei Punkte im Raum liegen gemeinsam auf mindestens einer Ebene“ und „jede gerade Zahl größer als 2 ist die Summe zweier Primzahlen“. Natürlich ist die erste Aussage falsch und die zweite wahr; die Wahrheit der dritten Aussage ist derzeit nicht bekannt.

Die mathematische Logik beschäftigt sich damit, grundlegende Begriffe wie „Aussage“ und „Wahrheit“ präzise zu definieren. In dieser Vorlesung versuchen wir nicht zu erläutern, was genau eine mathematische Aussage ist. Es gibt „atomare Aussagen“, die nicht aus kleineren Aussagen aufgebaut werden, z.B. „$1 + 1 = 3$“ oder „der Punkt P liegt auf der Ebene E“. Aus atomaren Aussagen können wir weitere Aussagen auf verschiedene Weise aufbauen:

- *Negation.* Jede Aussage φ besitzt eine gegenteilige Aussage „nicht φ“.
- *Logische Verknüpfungen.* Zu je zwei Aussagen φ und ψ kann man unter anderem die folgenden Aussagen bilden: „φ und ψ“, „φ oder ψ“, „φ impliziert ψ“ und „φ ist äquivalent zu ψ“.
- *Quantifizierung.* Wenn eine Aussage $\varphi(x)$ von einer Variablen x abhängt, dann kann man über diese Variable *quantifizieren*, um neue Aussagen zu erhalten: „für alle x gilt $\varphi(x)$“ (Allaussage) und „es existiert (mindestens) ein x, für das $\varphi(x)$ gilt“ (Existenzaussage).

In der folgenden Tabelle werden diese logischen Bausteine zusammengefasst:

symbolische Aussage	Bedeutung	Name	äquivalente Aussagen
$\neg\varphi$	nicht φ	Negation	
$\varphi \wedge \psi$	φ und ψ	Konjunktion	$\neg(\neg\varphi \vee \neg\psi)$
$\varphi \vee \psi$	φ oder ψ	Disjunktion	$\neg(\neg\varphi \wedge \neg\psi)$
$\varphi \Rightarrow \psi$	φ impliziert ψ, wenn φ, dann ψ, aus φ folgt ψ	Implikation	$\neg\varphi \vee \psi$
$\varphi \Leftrightarrow \psi$	φ ist äquivalent zu ψ, φ gilt genau dann, wenn ψ	Äquivalenz	$(\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \varphi)$
$\forall x\, \varphi(x)$	für alle/jedes x gilt $\varphi(x)$	Allaussage	$\neg\exists x\, \neg\varphi(x)$
$\exists x\, \varphi(x)$	es existiert/gibt ein x mit $\varphi(x)$	Existenzaussage	$\neg\forall x\, \neg\varphi(x)$

Die Symbole $\forall$ und $\exists$ heißen der *Allquantor* und der *Existenzquantor*. Man kann sich die Quantoren $\forall$ und $\exists$ als unendliche Verallgemeinerungen der logischen Verknüpfungen $\wedge$ und $\vee$ vorstellen. Manchmal verwendet man auch die Notation $\exists! x$ mit der Bedeutung „es existiert *genau ein* x“. Eigentlich ist $\exists! x\, \varphi(x)$ eine Abkürzung der Aussage

$$\exists x\, \varphi(x) \wedge \forall x \forall y((\varphi(x) \wedge \varphi(y)) \Rightarrow x = y).$$

Die Aussage $\forall x \forall y((\varphi(x) \wedge \varphi(y)) \Rightarrow x = y)$ heißt *Eindeutigkeitsaussage* für x in $\varphi(x)$: Sie bedeutet, dass *höchstens ein* x mit $\varphi(x)$ existiert.

Bemerkung 1.1.1. In mathematischen Texten (außer in der mathematischen Logik) schreibt man die logischen Symbole $\neg, \wedge, \vee, \forall, \exists$ sehr selten: Sie werden eher auf Deutsch ausgeschrieben. Im weiteren Verlauf dieses Skriptes werden wir also diese Symbole nicht benutzen. Die Symbole $\forall$ und $\exists$ sind trotzdem nützlich, wenn man mit der Hand schreibt.

Bemerkung 1.1.2 (Reihenfolge der Quantoren)**.** Viele mathematische Aussagen fangen mit mehreren Quantoren an. Bei denen muss man beachten, dass die Reihenfolge verschiedener Quantoren wichtig ist. Sei zum Beispiel $\varphi(x, y)$ die Aussage „wenn x ein Punkt im Raum ist, dann ist y eine Gerade im Raum, auf der x liegt“. Die Aussage

$$\forall x \exists y\, \varphi(x, y)$$

bedeutet, dass jeder Punkt im Raum auf mindestens einer Gerade liegt (was wahr ist), und die Aussage

$$\exists y \forall x\, \varphi(x, y)$$

bedeutet, dass alle Punkte im Raum auf derselben Gerade liegen (was falsch ist). Für eine beliebige Aussage $\varphi(x, y)$ gilt die Implikation

$$\exists y \forall x\, \varphi(x, y) \Rightarrow \forall x \exists y\, \varphi(x, y),$$

aber nicht unbedingt die umgekehrte Implikation.

Wie kann man entscheiden, ob eine gegebene Aussage wahr oder falsch ist? Der Wahrheitswert der Aussagen $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \Rightarrow \psi$ und $\varphi \Leftrightarrow \psi$ kann man einfach bestimmen, wenn die Wahrheitswerte von φ und ψ bekannt sind. Dazu verwendet man die folgenden Wahrheitstabellen:

φ	ψ	$\neg\varphi$	$\varphi \wedge \psi$	$\varphi \vee \psi$	$\varphi \Rightarrow \psi$	$\varphi \Leftrightarrow \psi$
w	w	f	w	w	w	w
w	f	f	f	w	f	f
f	w	w	f	w	w	f
f	f	w	f	f	w	w

Bemerkung 1.1.3. Bei den Wahrheitstabellen der Disjunktion $\vee$ und der Implikation $\Rightarrow$ muss man folgendes beachten:

- Wenn beide φ und ψ wahr sind, dann ist „φ oder ψ“ auch wahr. In der Mathematik ist „oder“ nie exklusiv, d.h., es bedeutet nicht „entweder φ oder ψ“, sondern „φ oder ψ oder beide“. Mit den obigen Symbolen kann das exklusive Oder als $\neg(\varphi \Leftrightarrow \psi)$ geschrieben werden.
- Wenn φ falsch ist, dann ist „φ impliziert ψ“ immer *wahr*. In der Mathematik ist die Implikation immer so verstanden. Anders gesagt: Aus etwas Falschem folgt alles.

Bemerkung 1.1.4. Die obige Liste von logischen Verknüpfungen ist nicht erschöpfend. Andere Beispiele sind das exklusive Oder $\neg(\varphi \Leftrightarrow \psi)$ und die umgekehrte Implikation $\varphi \Leftarrow \psi$. Man kann jedoch zeigen, dass alle möglichen logischen Verknüpfungen (d.h., alle Wahrheitstabellen) als Kombinationen von $\neg$ und $\vee$ erhalten werden können.

Definition 1.1.5 (Tautologie)**.** Eine Aussage heißt *Tautologie*, wenn sie aus Aussagen $\varphi_1, \ldots, \varphi_n$ und den logischen Verknüpfungen $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$ aufgebaut ist, so dass sich unter allen möglichen w/f-Belegungen der Aussagen φ_i der Wert w ergibt (gemäß den obigen Wahrheitstabellen).

Eine wichtige Eigenschaft des Begriffs der Tautologie ist seine *Berechenbarkeit*: Es ist immer möglich, automatisch und in endlich vielen Schritten nachzuprüfen, ob eine Aussage eine Tautologie ist oder nicht.

Beispiel 1.1.6 (Wichtige Tautologien)**.** Aussagen folgender Gestalt sind Tautologien:

(i) $\varphi \vee \neg\varphi$ (Satz vom ausgeschlossenen Dritten)

(ii) $\varphi \Leftrightarrow \neg\neg\varphi$ (Gesetz der doppelten Negation)

(iii) $\neg(\varphi \wedge \neg\varphi)$ (Satz vom ausgeschlossenen Widerspruch)

(iv) $(\varphi \Rightarrow \psi) \Leftrightarrow (\neg\psi \Rightarrow \neg\varphi)$ (Gesetz der Kontraposition)

(v) $(\varphi \wedge \neg\varphi) \Rightarrow \psi$ (Ex falso quodlibet/„aus Falschem [folgt] Beliebiges“)

(vi) $(\neg\varphi \Rightarrow (\psi \wedge \neg\psi)) \Rightarrow \varphi$ (Reductio ad absurdum/Widerspruchsbeweis)

(vii) $((\varphi \Rightarrow \psi) \wedge (\psi \Rightarrow \chi)) \Rightarrow (\varphi \Rightarrow \chi)$ (Syllogismus)

Als Beispiel überprüfen wir, dass das Gesetz der Kontraposition eine Tautologie ist:

φ	ψ	$\neg\varphi$	$\neg\psi$	$\varphi \Rightarrow \psi$	$\neg\psi \Rightarrow \neg\varphi$	$(\varphi \Rightarrow \psi) \Leftrightarrow (\neg\psi \Rightarrow \neg\varphi)$
w	w	f	f	w	w	w
w	f	f	w	f	f	w
f	w	w	f	w	w	w
f	f	w	w	w	w	w

Diese Tabelle zeigt, dass der Wahrheitswert von (iv) immer w ist, unabhängig von den Wahrheitswerten von φ und ψ. Das ist genau die Definition einer Tautologie.

Der Begriff der Wahrheit erweist sich als ziemlich subtil, wenn wir auch quantifizierte Aussagen betrachten wollen. Eigentlich sind wir in der Mathematik eher an dem konkreteren Begriff der *Beweisbarkeit* interessiert. Um den präzise zu definieren, benötigen wir *Axiome* und *Schlussregeln*. Axiome sind ausgewählte Aussagen, die als „wahr“ angenommen werden, und Schlussregeln schreiben vor, wie man Aussagen aus anderen Aussagen ableiten kann. Ein *Beweis* ist dann eine endliche Folge von Aussagen, in der jede Aussage entweder ein Axiom ist oder durch eine Schlussregel aus vorherigen Aussagen folgt. Eine Aussage φ heißt *beweisbar*, wenn ein Beweis existiert, dessen letzte Aussage φ ist.

Wir beschreiben jetzt ein solches System von Axiomen und Schlussregeln, das für die ganze Mathematik geeignet ist: Es ist die sogenannte *Prädikatenlogik erster Stufe mit Gleichheit*. Dies dient nur der Veranschaulichung, und es ist gar nicht wichtig, sich diese Axiome und Schlussregeln einzuprägen. Auf jeden Fall ist die folgende Beschreibung unvollständig, weil wir unter anderem nicht erläutert haben, was genau die atomaren Aussagen sind und was die „Variablen“ $x, y, \ldots$ sind.

- Die *aussagenlogischen Axiome* sind alle Tautologien.
- Die *quantorlogischen Axiome* sind:
  - $\forall x\, \varphi(x) \Rightarrow \varphi(a)$.
  - Falls x in ψ nicht vorkommt, $\forall x(\psi \Rightarrow \varphi(x)) \Rightarrow (\psi \Rightarrow \forall x\, \varphi(x))$.

- Die *Gleichheitsaxiome* sind:
  - $x = x$ (das *Identitätsaxiom*).
  - $x = y \Rightarrow (\varphi(x) \Leftrightarrow \varphi(y))$.
- Es gibt nur zwei Schlussregeln:
  - *Modus Ponens*: Aus φ und $\varphi \Rightarrow \psi$ kann man ψ herleiten.
  - *Allquantoreinführung*: Aus $\varphi(x)$ kann man $\forall x\, \varphi(x)$ herleiten.

In diesem System wird die Existenzaussage $\exists x\, \varphi(x)$ als Abkürzung von $\neg\forall x\, \neg\varphi(x)$ definiert, und sie erfordert keine weiteren Axiome.

Bemerkung 1.1.7. Die Schlussregel der Allquantoreinführung sieht vielleicht ein bisschen merkwürdig aus. Man soll sie wie folgt verstehen: Wenn man in diesem System eine Aussage $\varphi(x)$ mit einer freien Variablen x beweisen kann, bedeutet das, dass diese Aussage für ein *beliebiges* x gilt. Die selbständige unquantifizierte Aussage $\varphi(x)$ hat also dieselbe Bedeutung wie die Allaussage $\forall x\, \varphi(x)$.

Beispiel 1.1.8. Als Beispiel geben wir einen formalen Beweis der Aussage

$$\forall x \forall y (x \neq x \Rightarrow x = y),$$

wobei $x \neq x$ eine Abkürzung von $\neg(x = x)$ ist:

1. $x = x$	(Identitätsaxiom)
2. $x = x \Rightarrow (x \neq x \Rightarrow x = y)$	(Tautologie)
3. $x \neq x \Rightarrow x = y$	(Modus Ponens aus 1 und 2)
4. $\forall y (x \neq x \Rightarrow x = y)$	(Allquantoreinführung aus 3)
5. $\forall x \forall y (x \neq x \Rightarrow x = y)$	(Allquantoreinführung aus 4).

Beispiel 1.1.9. Ein berühmter Syllogismus lautet:

> Alle Menschen sind sterblich.
> Sokrates ist ein Mensch.
> Also ist Sokrates sterblich.

Sei $\mu(x)$ die Aussage „x ist ein Mensch“ und sei $\sigma(x)$ die Aussage „x ist sterblich“. In der Prädikatenlogik erster Stufe sieht dieser Syllogismus wie folgt aus:

1. $\forall x (\mu(x) \Rightarrow \sigma(x))$	(1. Annahme)
2. $\mu(\text{Sokrates})$	(2. Annahme)
3. $\forall x (\mu(x) \Rightarrow \sigma(x)) \Rightarrow (\mu(\text{Sokrates}) \Rightarrow \sigma(\text{Sokrates}))$	(quantorlogisches Axiom)
4. $\mu(\text{Sokrates}) \Rightarrow \sigma(\text{Sokrates})$	(Modus Ponens aus 1 und 3)
5. $\sigma(\text{Sokrates})$	(Modus Ponens aus 2 und 4).

Beispiel 1.1.10. Für den Exiztenzquantor $\exists$ gelten folgenden Aussagen:

- $\varphi(a) \Rightarrow \exists x\, \varphi(x)$
- Falls x in ψ nicht vorkommt, $\forall x (\varphi(x) \Rightarrow \psi) \Rightarrow (\exists x\, \varphi(x) \Rightarrow \psi)$

Diese Aussagen folgen aus den entsprechenden quantorlogischen Axiomen. Hier ist zum Beispiel ein Beweis der ersten Aussage:

1. $\forall x\, \neg\varphi(x) \Rightarrow \neg\varphi(a)$	(quantorlogisches Axiom)
2. $(\forall x\, \neg\varphi(x) \Rightarrow \neg\varphi(a)) \Rightarrow (\varphi(a) \Rightarrow \neg\forall x\, \neg\varphi(x))$	(Tautologie)
3. $\varphi(a) \Rightarrow \neg\forall x\, \neg\varphi(x)$	(Modus Ponens aus 1 und 2).

1.1.1 Beispiele von Beweisen

Zum Aufwärmen besprechen wir ein paar Beweise von Elementarsätzen in der Algebra: dem Satz von Euklid, dass unendlich viele Primzahlen existieren, und dem „Satz der Pythagoreer“, dass $\sqrt{2}$ (die Länge der Diagonale in einem Einheitsquadrat) keine rationale Zahl ist. Dazu werden wir einige Begriffe benutzen, die später in der Vorlesung genauer eingeführt werden.

Die *natürlichen Zahlen* sind die Zahlen 0, 1, 2, 3, und so weiter (siehe Abschnitt 1.2.1 für eine mengentheoretische Definition). Sind m und n natürliche Zahlen, so sagt man „m teilt n“ oder „n ist durch m teilbar“, wenn eine natürliche Zahl r mit $n = m \cdot r$ existiert. Eine *Primzahl* ist eine natürliche Zahl, die nicht gleich 1 ist und die nur durch 1 und sich selbst teilbar ist. Die kleinste Primzahl ist 2, dann kommen 3, 5, 7, 11, usw. (Die natürliche Zahl 0 ist durch jede andere natürliche Zahl teilbar, und damit keine Primzahl.)

Satz 1.1.11 (Euklid)**.** *Es gibt unendlich viele Primzahlen.*

Um diesen Satz zu beweisen brauchen wir das folgende Lemma:

Lemma 1.1.12. *Sei $n \geq 2$ eine natürliche Zahl. Dann existiert eine Primzahl p, die n teilt.*

Beweis. Wir verwenden das Prinzip der vollständigen Induktion (Korollar 1.2.23). Damit dürfen wir annehmen, dass jede natürliche Zahl m mit $2 \leq m < n$ durch eine Primzahl teilbar ist (diese Aussage heißt die *Induktionsvoraussetzung*). Wenn n schon eine Primzahl ist, können wir $p = n$ nehmen (da n durch sich selbst teilbar ist). Wir nehmen jetzt an, dass n keine Primzahl ist. Nach Definition von Primzahl existiert dann eine natürliche Zahl $m \neq 1, n$, die n teilt. Das heißt, es gilt $n = m \cdot r$ mit einer natürichen Zahl r. Da $n \neq 0, m$, ist $r \geq 2$ und damit ist $m < n$. Nach der Induktionsvoraussetzung existiert eine Primzahl p, die m teilt. Da n durch m teilbar ist und m durch p teilbar ist, ist auch n durch p teilbar, wie gewünscht. □

Bemerkung 1.1.13. Lemma 1.1.12 hat die folgende logische Gestalt:

$$\forall n(\varphi(n) \Rightarrow \exists p(\psi(p) \wedge \chi(p, n))),$$

wobei $\varphi(n)$, $\psi(p)$ und $\chi(p, n)$ die Aussagen „n ist eine natürliche Zahl und $n \geq 2$“, „p ist eine Primzahl“ und „p teilt n“ sind.

Beweis vom Satz 1.1.11. Wir verwenden einen Widerspruchsbeweis. Angenommen, es gäbe nur endlich viele Primzahlen $p_1, p_2, \ldots, p_k$. Sei

$$q = p_1 \cdot p_2 \cdot \ldots \cdot p_k$$

das Produkt aller Primzahlen und sei $n = q + 1$. Nach dem Lemma 1.1.12 existiert eine Primzahl p, die n teilt. Es existiert also eine natürliche Zahl m mit $n = p \cdot m$. Da q das Produkt aller Primzahlen ist, ist q durch p teilbar: Es existiert eine natürliche Zahl r mit $q = p \cdot r$. Dann

$$1 = n - q = p \cdot m - p \cdot r = p \cdot (m - r).$$

Insbesondere ist 1 durch p teilbar, und damit ist $p = 1$. Aber $p \neq 1$ nach Definition einer Primzahl, was ein Widerspruch ist. □

Der Beweis vom Lemma 1.1.12 liefert auch die folgende stärkere Aussage: Jede natürliche Zahl $n \geq 2$ ist ein Produkt von Primzahlen (das gilt auch für $n = 1$, wenn man 1 als das leere Produkt betrachtet). Denn im Beweis sind beide m und r Produkte von Primzahlen nach der Induktionsvoraussetzung, und damit ist auch $n = m \cdot r$ ein Produkt von Primzahlen. Diese stärkere Aussage ist die Existenzaussage im folgenden wichtigen Satz; die Eindeutigkeitsaussage ist schwieriger und wird in der Vorlesung *Lineare Algebra II* besprochen:

***Satz 1.1.14** (Fundamentalsatz der Arithmetik)**.** *Jede natürliche Zahl $n \geq 1$ lässt sich eindeutig als Produkt von Primzahlen darstellen. Genauer existiert zu jeder Primzahl p genau eine natürliche Zahl $v_p(n)$, so dass $v_p(n) \neq 0$ nur für endlich viele Primzahlen p und*

$$n = \prod_{p \text{ Primzahl}} p^{v_p(n)}.$$

Die rechte Seite der obigen Gleichheit ist eine Abkürzung des Produkts

$$p_1^{v_{p_1}(n)} \cdot p_2^{v_{p_2}(n)} \cdot p_3^{v_{p_3}(n)} \cdot \ldots,$$

wobei $p_1, p_2, p_3, \ldots$ alle Primzahlen in aufsteigender Reihenfolge sind. Das ist also ein Produkt unendlich vieler Zahlen (nach dem Satz von Euklid), aber es ist trotzdem sinnvoll, da nur endlich viele dieser Zahlen nicht gleich 1 sind. Zum Beispiel:

$$\begin{aligned} 10 &= 2 \cdot 5, \\ 56 &= 2^3 \cdot 7, \\ 60 &= 2^2 \cdot 3 \cdot 5. \end{aligned}$$

Die im obigen Sinne eindeutige Darstellung von n als Produkt von Primzahlen heißt die *Primfaktorzerlegung* von n. Der Exponent $v_p(n)$ ist die *Vielfachheit* der Primzahl p in n.

Wir wenden uns nun dem Thema der Irrationalität von $\sqrt{2}$ zu. Die *rationalen Zahlen* sind reellen Zahlen, die als Bruch a/b zweier ganzen Zahlen a, b mit $b \neq 0$ dargestellt werden können. In einer solchen Bruchdarstellung kann man immer annehmen, dass a und b *teilerfremd* sind, d.h., dass 1 die einzige natürliche Zahl ist, die beide a und b teilt (sonst kann man a und b durch einen gemeinsamen Teiler dividieren, ohne die Zahl a/b zu verändern).

In den reellen Zahlen besitzt jede Zahl $r \geq 0$ eine Quadratwurzel $\sqrt{r}$, die die einzige ≥ 0 reelle Zahl ist, deren Quadrat gleich r ist. Der folgende Satz bedeutet, dass die reelle Zahl $\sqrt{2}$ keine rationale Zahl ist.

Satz 1.1.15. *Es gibt keine rationale Zahl x mit $x^2 = 2$.*

Um diesen Satz zu beweisen brauchen wir wieder ein Lemma. Eine natürliche Zahl heißt bekanntlich *gerade*, wenn sie durch 2 teilbar ist.

Lemma 1.1.16. *Eine natürliche Zahle n ist genau dann gerade, wenn ihr Quadrat n^2 gerade ist.*

Beweis. Das ist eine Aussage der Gestalt $\varphi \Leftrightarrow \psi$. Um sie zu beweisen, brauchen wir beide Implikationen $\varphi \Rightarrow \psi$ und $\varphi \Leftarrow \psi$ zu beweisen.

Zu $\Rightarrow$. Sei n gerade. Das heißt, es existiert eine natürliche Zahl m, so dass $n = 2m$. Man berechnet:

$$n^2 = (2m)^2 = 2^2 m^2 = 2(2m^2).$$

Also ist n^2 auch gerade.

Zu $\Leftarrow$. Wir verwenden das Gesetz der Kontraposition: Um eine Aussage der Form $\varphi \Rightarrow \psi$ zu beweisen, genügt es die Kontraposition $\neg\psi \Rightarrow \neg\varphi$ zu beweisen. Es genügt also zu zeigen, dass n^2 *un*gerade ist, wenn n *un*gerade ist. Die Zahl $n - 1$ ist dann gerade, und hat somit die Form $2m$. Es gilt dann $n = 2m + 1$ und man berechnet:

$$\begin{aligned} n^2 = (2m+1)^2 &= (2m+1)(2m+1) = 2m(2m+1) + 1(2m+1) \\ &= 4m^2 + 2m + 2m + 1 = 4m^2 + 4m + 1 = 2(2m^2 + 2m) + 1. \end{aligned}$$

Da $2(2m^2 + 2m)$ gerade ist, ist n^2 ungerade. □

Beweis vom Satz 1.1.15. (Widerspruchsbeweis.) Angenommen, es existiert eine rationale Zahl x mit $x^2 = 2$. Nach Definition der rationalen Zahlen lässt sich x als Bruch $\pm a/b$ mit natürlichen Zahlen a, b darstellen, wobei a und b teilerfremd sind. Es gilt dann $(a/b)^2 = 2$, also $a^2 = 2b^2$. Insbesondere sind a^2 und daher a gerade, nach Lemma 1.1.16. Es gilt also $a = 2n$ mit einer natürlichen Zahl n. Dann ist $2b^2 = 4n^2$, also $b^2 = 2n^2$, und damit ist b^2 gerade. Nach Lemma 1.1.16 ist auch b gerade. Dass a und b beide gerade sind, steht im Widerspruch zur Annahme, dass a und b teilerfremd sind. □

1.2 Mengen

Ohne Axiome kann man nichts beweisen. Ebenfalls kann mann nichts aus nichts *definieren*. Deswegen werden grundlegende mathematische Objekte nicht direkt definiert; stattdessen muss man grundlegende mathematische Objekte *indirekt* durch ein Axiomsystem definieren. Die Axiome sagen uns nicht, *was* genau diese Objekte sind, sondern *wie* man mit diesen Objekten umgehen kann.

Es hat sich herausgestellt, dass *Mengen* gute primitive Objekte sind, auf denen fast alle die Mathematik beruhen kann. Der moderne mathematische Begriff von „Menge“ wurde von Georg Cantor am Ende des 19. Jahrhunderts eingeführt. Sein 1895 Artikel *Beiträge zur Begründung der transfiniten Mengenlehre* beginnt mit folgendem Absatz:

> Unter einer „Menge“ verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche Elemente von M genannt werden) zu einem Ganzen.

Wir werden mit so einem Begriff von „Menge“ arbeiten. Zum Beispiel, wenn man eine bestimmte Liste $a, b, c, \dots$ von Objekten betrachtet, darf man diese Objekte in einer Menge zusammenfassen. Diese Menge wird dann mit

$$\{a, b, c, \dots\}$$

bezeichnet. Wenn $\varphi(x)$ eine logische Aussage ist, darf man auch nach Cantors obiger Definition die Menge aller Objekte x betrachten, für die die Aussage $\varphi(x)$ gilt. Diese Menge wird mit

$$\{x \mid \varphi(x)\} \quad \text{oder} \quad \{x : \varphi(x)\}$$

bezeichnet.

Notation 1.2.1. Die Aussage „m ist ein Element von M“ wird mit der Notation $m \in M$ abgekürzt. Man sagt auch „m liegt in M“ oder „M enthält m“. Man schreibt $m \notin M$ für die gegenteilige Aussage, d.h., falls m kein Element von M ist.

Leider ist Cantors Definition einer Menge keine präzise mathematische Definition, und sie führt sehr schnell zu einem berühmten logischen Paradoxon:

Paradoxon 1.2.2 (Die Russelsche Antinomie)**.** Wir betrachten die Menge

$$R := \{x \mid x \text{ ist eine Menge und } x \notin x\}.$$

In Worten ist R die Menge aller Mengen, die kein Element von sich selbst sind. Nach Definition von R gilt also

$$x \in R \iff x \notin x$$

für alle Mengen x. Frage: Stimmt $R \in R$ oder nicht? Wenn wir x durch R in der obigen Äquivalenz ersetzen, erhalten wir

$$R \in R \iff R \notin R.$$

Das ist eine Aussage der Gestalt $\varphi \Leftrightarrow \neg\varphi$, also ein Widerspruch!

Die Russelsche Antinomie zeigt folgendes: Um einen widerspruchfreien Begriff von Menge zu erhalten, darf $\{x | x \notin x\}$ *nicht* eine Menge sein. Eine natürliche Frage ist dann: Für welche logischen Aussagen $\varphi(x)$ darf man die Menge $\{x \mid \varphi(x)\}$ bilden? Diese Frage ist schwierig, aber sie kann durch die *axiomatische Mengenlehre* ausführlich beantwortet werden.

In der Praxis der Mathematik ist es gar nicht wichtig, die genauen Axiome der Mengenlehre zu kennen; ein gutes intuitives Verständnis von Mengen und Mengenoperationen ist hinreichend. In diesem Abschnitt erklären wir einige Konstruktionen mit Mengen, die durch die axiomatische Mengenlehre begründet werden können, und die für den größten Teil der Mathematik ausreichen. Das unterliegende System von Axiomen, das wir implizit benutzen werden, heißt die *Zermelo-Fraenkel-Mengenlehre mit Auswahlaxiom* oder *ZFC* (das C steht für „choice“, das englische Wort für Auswahl).

Definition 1.2.3 (Gleichheit von Mengen)**.** Zwei Mengen A und B sind genau dann *gleich*, in Zeichen $X = Y$, wenn sie dieselben Elemente enthalten.

Zum Beispiel: $\{0, 0, 1\} = \{0, 1\} = \{1, 0\}$.

Definition 1.2.4 (Die leere Menge)**.** Die *leere Menge*, $\varnothing$ oder $\{\}$, ist die Menge, die keine Elemente enthält.

Die letzten zwei Definitionen entsprechen Axiomen der Mengenlehre, nämlich das *Extensionalitätsaxiom* und das *Leermengenaxiom*. Nach Definition 1.2.3 ist die leere Menge eindeutig, d.h.: Enthalten A und B keine Elemente, so gilt $A = B$. Deswegen darf man wirklich „*die* leere Menge“ sagen.

Definition 1.2.5 (Teilmenge)**.** Seien A, B Mengen. Die Menge A heißt *Teilmenge* von B, in Zeichen $A \subset B$, falls jedes Element von A ein Element von B ist. Man sagt auch „A ist in B enthalten“.

Bemerkung 1.2.6. Manche Quellen verwenden die Notation $A \subseteq B$, wenn A eine Teilmenge von B ist, und sie schreiben $A \subset B$, nur wenn $A \neq B$. In diesem Skript werden wir $\subseteq$ nicht verwenden.

Wie die Russelsche Antinomie zeigt, nicht alle logischen Aussagen $\varphi(x)$ bilden gültige Mengen $\{x \mid \varphi(x)\}$. Aber wenn man die Objekte x nur innerhalb einer bestimmten Menge A auswählen, dann gibt es keine Probleme, und man darf folgende Teilmenge von A immer betrachten:[1]

$$\{x \mid x \in A \text{ und } \varphi(x)\}.$$

Diese Menge wird auch mit

$$\{x \in A \mid \varphi(x)\}$$

bezeichnet.

Definition 1.2.7 (Vereinigung, Durchschnitt, Komplement)**.** Seien A, B Mengen.

- Die *Vereinigung* von A und B ist die Menge

  $$A \cup B := \{x \mid x \in A \text{ oder } x \in B\}$$

  (gelesen „A vereinigt mit B“).

- Der *Durchschnitt* oder die *Schnittmenge* von A und B ist die Menge

  $$A \cap B := \{x \mid x \in A \text{ und } x \in B\}$$

  (gelesen „A geschnitten mit B“). Die Mengen A und B heißen *disjunkt*, falls $A \cap B = \varnothing$.

[1] Das folgt aus dem *Aussonderungsaxiom* der Mengenlehre.

- Das *Komplement* von B in A, oder die *Differenz* von A und B, ist die Menge

$$A \setminus B := \{x \mid x \in A \text{ und } x \notin B\}$$

(gelesen „A ohne B“).

Bemerkung 1.2.8. In der obigen Definition haben wir das Symbol := geschrieben. Der Doppelpunkt betont, dass die linke Seite durch die rechte Seite definiert wird.

Proposition 1.2.9. *Seien A, B, C Mengen.*

(i) *Es gilt $A = B$ genau dann, wenn $A \subset B$ und $B \subset A$ gelten.*

(ii) *Ist $A \subset B$ und $B \subset C$, so ist $A \subset C$.*

(iii) *Die Operationen $\cup$ und $\cap$ sind* kommutativ*, d.h.,*

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A.$$

(iv) *Die Operationen $\cup$ und $\cap$ sind* assoziativ*, d.h.,*

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

(v) *Es gilt die* de Morganschen Gesetze

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

(vi) *Es gilt*

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

Beweis. Alle Aussagen folgen unmittelbar aus den Definitionen. Wir beweisen das erste de Morgansche Gesetz

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

als Beispiel. Nach (i) braucht man zu zeigen, dass beide Inklusionen $\subset$ und $\supset$ gelten.

Zu $\subset$. Sei $x \in A \cup (B \cap C)$. Das heißt, $x \in A$ oder $x \in B \cap C$. Falls $x \in A$, dann $x \in A \cup D$ für irgendeine Menge D; insbesondere gilt $x \in A \cup B$ und $x \in A \cup C$, d.h., $x \in (A \cup B) \cap (A \cup C)$. Falls $x \in B \cap C$, dann $x \in B$ und daher $x \in A \cup B$, und auch $x \in C$ und daher $x \in A \cup C$, also $x \in (A \cup B) \cap (A \cup C)$.

Zu $\supset$. Sei $x \in (A \cup B) \cap (A \cup C)$. Das heißt, $x \in A \cup B$ und $x \in A \cup C$. Ist $x \in A$, so folgt $x \in A \cup (B \cap C)$. Sonst muss es sein, dass $x \in B$ und $x \in C$, d.h., $x \in B \cap C$. In diesem Fall gilt dann auch $x \in A \cup (B \cap C)$. □

Bemerkung 1.2.10. Die erste Aussage in Proposition 1.2.9 ist trivial aber ganz wichtig in der Praxis: Um zu beweisen, dass zwei Mengen A und B gleich sind, muss man eigentlich zwei verschiedene Aussagen beweisen: dass jedes Element von A in B liegt, und umgekehrt dass jedes Element von B in A liegt.

Definition 1.2.11 (Potenzmenge)**.** Sei X eine Menge. Die *Potenzmenge* von X, $\mathcal{P}(X)$, ist die Menge aller Teilmengen von X:

$$\mathcal{P}(X) := \{A \mid A \subset X\}.$$

Die Existenz der Potenzmenge einer beliebigen Menge ist wieder ein Axiom der Mengenlehre, das *Potenzmengenaxiom*.

Bemerkung 1.2.12. Die leere Menge ist eine Teilmenge jeder Menge, weil jedes Element der leeren Menge ein Element aller anderen Mengen ist. Also für alle Mengen X gilt $\varnothing \in \mathcal{P}(X)$. Für alle Mengen X gilt auch $X \in \mathcal{P}(X)$, da $X \subset X$.

Seien a und b mathematische Objekte. Das *Paar* (a, b) ist ein neues Objekt mit folgender Eigenschaft: Zwei Paare (a, b) und (a', b') sind genau dann gleich, wenn $a = a'$ und $b = b'$. In der Mengenlehre kann man das Paar (a, b) als die Menge $\{\{a\}, \{a, b\}\}$ definieren. Allgemeiner, aus n Objekten $a_1, \ldots, a_n$ kann mann das *n-Tupel* $(a_1, \ldots, a_n)$ bilden.[2]

Bemerkung 1.2.13. Das Paar (a, b) ist nicht mit der Menge $\{a, b\}$ zu verwechseln. Zum Beispiel: Es gilt immer $\{a, b\} = \{b, a\}$, aber $(a, b) = (b, a)$ gilt nur, wenn $a = b$. Das heißt, in einem Paar (a, b) ist die Reihenfolge von a und b relevant. Eine ähnliche Bemerkung gilt für n-Tupel.

Definition 1.2.14 (Produkt, Summe)**.** Seien A, B Mengen. Das (kartesische) *Produkt* von A und B ist die Menge aller Paare (a, b) mit $a \in A$ und $b \in B$:

$$A \times B := \{(a, b) \mid a \in A \text{ und } b \in B\}.$$

Die *Summe* oder *disjunkte Vereinigung* von A und B ist die Menge

$$A \sqcup B := (\{1\} \times A) \cup (\{2\} \times B).$$

In der folgenden Tabelle sind die bisher eingeführten Notationen zusammengefasst:

Notation	Bedeutung
$a \in A$	a ist ein Element von A, a liegt in A
$A \subset B$	A ist eine Teilmenge von B
$\{x \in A \mid \varphi(x)\}$	die Menge aller $x \in A$, für die $\varphi(x)$ gilt
$\varnothing$	die leere Menge
$A \cup B$	die Vereinigung von A und B
$A \cap B$	der Durchschnitt von A und B
$A \setminus B$	das Komplement von B in A
$A \times B$	das Produkt von A und B
$A \sqcup B$	die Summe von A und B
$\mathcal{P}(A)$	die Potenzmenge von A

Definitionen 1.2.7 und 1.2.14 können auf mehr als zwei Mengen verallgemeinert werden. Zum Beispiel ist das n-fache Produkt

$$A_1 \times A_2 \times \cdots \times A_n$$

die Menge aller n-Tupel $(a_1, a_2, \ldots, a_n)$ mit $a_i \in A_i$ für jedes $i \in \{1, \ldots, n\}$. Um diese Konstruktionen auf sogar unendlich viele Mengen zu verallgemeinern, brauchen wir den Begriff der *Mengenfamilie*. Eine Mengenfamilie $(A_i)_{i \in I}$ mit Indexmenge I ordnet jedem Element $i \in I$ eine Menge A_i zu. Dieses „Zuordnen“ kann präziser als eine Abbildung $i \mapsto A_i$ mit Definitionsbereich I definiert werden (siehe Abschnitt 1.3).

Definition 1.2.15 (Vereinigung, Durchschnitt, Summe und Produkt von Familien)**.** Sei I eine Menge und $(A_i)_{i \in I}$ eine Mengenfamilie mit Indexmenge I.

- Die *Vereinigung* der Familie $(A_i)_{i \in I}$ ist die Menge

$$\bigcup_{i \in I} A_i := \{x \mid \text{es existiert } i \in I \text{ mit } x \in A_i\}.$$

[2] Man kann zum Beispiel dieses n-Tupel als iteriertes Paar $(a_1, (a_2, (\ldots, a_n) \ldots))$ definieren.

- Falls $I \neq \varnothing$, ist der *Durchschnitt* oder die *Schnittmenge* der Familie $(A_i)_{i\in I}$ die Menge

$$\bigcap_{i\in I} A_i := \{x \mid \text{für alle } i \in I \text{ gilt } x \in A_i\}.$$

- Die *Summe* oder *disjunkte Vereinigung* der Familie $(A_i)_{i\in I}$ ist die Menge

$$\coprod_{i\in I} A_i := \{(i, x) \mid i \in I \text{ und } x \in A_i\}.$$

- Das *Produkt* der Familie $(A_i)_{i\in I}$ ist die Menge

$$\prod_{i\in I} A_i := \{(a_i)_{i\in I} \mid \text{für alle } i \in I \text{ gilt } a_i \in A_i\}.$$

Die Definition der Vereinigung einer Familie wird durch das *Vereinigungsaxiom* der Mengenlehre begründet. Die anderen Konstruktionen in der obigen Definition können wie folgt begründet werden: $\bigcap_{i\in I} A_i$ ist eine Teilmenge von irgendeinem A_i, $\coprod_{i\in I} A_i$ ist gleich der Vereinigung $\bigcup_{i\in I} (\{i\} \times A_i)$, und $\prod_{i\in I} A_i$ kann präziser als Teilmenge von $\mathcal{P}\left(I \times \bigcup_{i\in I} A_i\right)$ definiert werden.

Bemerkung 1.2.16. In der Definition von $\bigcap_{i\in I} A_i$ ist es notwendig, $I \neq \varnothing$ vorauszusetzen. Sonst würde der Durchschnitt $\bigcap_{i\in I} A_i$ die „universelle Menge“ sein, d.h., die Menge *aller* Objekte. Aber die universelle Menge existiert nicht, sonst würde die Russelsche Menge als Teilmenge davon auch existieren, was bekanntlich nicht möglich ist.

Wenn man nur Familien von Teilmengen einer festen Menge X betrachtet (was fast immer der Fall ist), ist es sinnvoll ihre Vereinigungen und Durchschnitte als $\{x \in X \mid \dots\}$ zu definieren. Mit dieser Veränderung ist der Durchschnitt der Familie mit Indexmenge $\varnothing$ sinvoll und gleich X.

1.2.1 Die natürlichen Zahlen

Die *natürlichen Zahlen* sind die Zahlen 0, 1, 2, 3, usw. Die Menge aller natürlichen Zahlen wird mit $\mathbb{N}$ bezeichnet:

$$\mathbb{N} := \{0, 1, 2, 3, \dots\}.$$

Die genaue Beschaffenheit der natürlichen Zahlen ist nicht wichtig, und es gibt mehrere mögliche mengentheoretische Definitionen. Eine Standarddefinition ist:

$$0 := \varnothing,\ 1 := \{0\},\ 2 := \{0, 1\},\ 3 := \{0, 1, 2\},\ \text{usw.}$$

Mit dieser Definition ist die gewöhnliche Ordnungsrelation $\leq$ zwischen natürlichen Zahlen einfach die Teilmengenrelation $\subset$.

Übrigens würde die obige Definition von $\mathbb{N}$ in der formalen Mengenlehre nicht sinnvoll sein (was bedeutet denn „...“?). Dass eine solche Menge $\mathbb{N}$ trotzdem existiert folgt aus der *Unendlichkeitsaxiom* der Mengenlehre.

Bemerkung 1.2.17 (Ist 0 eine natürliche Zahl?)**.** In manchen Quellen werden die natürlichen Zahlen als $\mathbb{N} = \{1, 2, 3, \dots\}$ definiert, d.h., die Null wird nicht als natürliche Zahl betrachtet. Dann wird $\mathbb{N} \cup \{0\}$ auch mit $\mathbb{N}_0$ bezeichnet. Diese alternative Konvention ist populär in der Analysis, weil die Folge $(1/n)_{n\in\mathbb{N}}$ oft verwendet wird, in der 0 kein Element von $\mathbb{N}$ sein darf. In der Algebra benutzt man eher unsere Konvention, so dass z.B. $(\mathbb{N}, +)$ ein Monoid ist (siehe Bemerkung 2.1.13).

Bemerkung 1.2.18 (ganze, rationale, reelle, komplexe Zahlen)**.** Es gibt bekanntlich mehrere Erweiterungen der natürlichen Zahlen $\mathbb{N}$:

- die ganzen Zahlen $\mathbb{Z}$;
- die rationalen Zahlen $\mathbb{Q}$;
- die reellen Zahlen $\mathbb{R}$;
- die komplexen Zahlen $\mathbb{C}$.

Im Kapitel 2 werden wir erklären, wie die Mengen $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ konstruiert werden können.

Folgender Satz ist eine fundamentale Eigenschaft der natürlichen Zahlen. Wir nehmen diesen Satz ohne Beweis an, da ein Beweis eine detaillierte Beschreibung der unterliegenden Axiome erfordern würde.

***Satz 1.2.19** (Wohlordnungsprinzip)**.** *Jede nichtleere Teilmenge $A \subset \mathbb{N}$ besitzt ein* kleinstes Element*, d.h., ein Element $a \in A$, so dass $a \leq b$ für alle $b \in A$.*

Korollar 1.2.20 (Induktionsprinzip)**.** *Sei $A \subset \mathbb{N}$ eine Teilmenge mit folgenden Eigenschaften:*

(i) (Induktionsanfang) $0 \in A$.

(ii) (Induktionsschritt) *Für alle $n \in \mathbb{N}$ gilt:*

$$n \in A \implies n+1 \in A.$$

Dann ist $A = \mathbb{N}$.

Beweis. (Widerspruchsbeweis.) Angenommen, $A \neq \mathbb{N}$. Dann ist das Komplement $B := \mathbb{N} \setminus A$ nicht leer. Nach dem Wohlordnungsprinzip (Satz 1.2.19) hat B ein kleinstes Element $b \in B$. Da $0 \in A$ nach (i) ist $b \neq 0$. Also existiert $n \in \mathbb{N}$ mit $b = n+1$. Da b das kleinste Element von B ist, ist $n \notin B$, d.h., $n \in A$. Nach (ii) ist dann $b = n+1 \in A$. Also liegt b in $A \cap B = \varnothing$, im Widerspruch zur Definition von $\varnothing$. □

Das Induktionsprinzip wird häufig verwendet, um eine gegebene Aussage $\varphi(n)$ für alle natürlichen Zahlen $n \in \mathbb{N}$ zu beweisen. Dazu wenden wir das Induktionsprinzip auf die folgende Menge an:

$$A = \{n \in \mathbb{N} \mid \varphi(n)\}.$$

Es genügt also zu zeigen, dass $\varphi(0)$ gilt, und dass für alle $n \in \mathbb{N}$ die Implikation $\varphi(n) \Rightarrow \varphi(n+1)$ gilt. Dieses Beweisverfahren heißt *Beweis durch Induktion.* Während die Implikation $\varphi(n) \Rightarrow \varphi(n+1)$ im Induktionsschritt bewiesen wird, heißt die Aussage $\varphi(n)$ die *Induktionsvoraussetzung.*

Beispiel 1.2.21. Als Beispiel zum Induktionsprinzip beweisen wir folgende Aussage: Für alle $n \in \mathbb{N}$ ist die Summe aller natürlichen Zahlen $\leq n$ gleich $\frac{n(n+1)}{2}$. In Zeichen:

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2}. \tag{1.2.22}$$

Hierzu betrachten wir die Menge

$$A = \{n \in \mathbb{N} \mid (1.2.22) \text{ gilt}\} \subset \mathbb{N}.$$

- *Induktionsanfang.* Es gilt $0 \in A$, da $\sum_{k=0}^{0} k = 0 = \frac{0(0+1)}{2}$.
- *Induktionsschritt.* Sei $n \in A$. Dann gilt

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

  Dabei haben wir die Induktionsvoraussetzung in der zweiten Gleichung verwendet. Diese Berechnung zeigt, dass $n+1 \in A$.

Nach dem Induktionsprinzip gilt dann $A = \mathbb{N}$. Das heißt, (1.2.22) gilt für alle $n \in \mathbb{N}$, wie behauptet.

Folgendes Korollar ist eine stärkere Variante des Induktionsprinzips: Um zu beweisen, dass eine Aussage $\varphi(n)$ für alle $n \in \mathbb{N}$ gilt, darf man voraussetzen, dass $\varphi(m)$ für alle $m < n$ gilt.

Korollar 1.2.23 (Prinzip der vollständigen Induktion)**.** *Sei $A \subset \mathbb{N}$ eine Teilmenge, so dass folgendes gilt für alle $n \in \mathbb{N}$:*

$$\{m \in \mathbb{N} \mid m < n\} \subset A \implies n \in A.$$

Dann ist $A = \mathbb{N}$.

Beweis. Sei $\mathbb{N}_{<n} := \{m \in \mathbb{N} | m < n\}$. Wir wenden das Induktionsprinzip 1.2.20 mit folgender Menge an:

$$A' := \{n \in \mathbb{N} \mid \mathbb{N}_{<n} \subset A\} \subset \mathbb{N}.$$

- *Induktionsanfang.* Es gilt $0 \in A'$, da $\mathbb{N}_{<0} = \varnothing \subset A$.
- *Induktionsschritt.* Es sei $n \in A'$, d.h., $\mathbb{N}_{<n} \subset A$. Dann folgt aus der Voraussetzung, dass $n \in A$. Also $\mathbb{N}_{<n+1} = \mathbb{N}_{<n} \cup \{n\} \subset A$, d.h., $n + 1 \in A'$.

Aus dem Induktionsprinzip folgt, dass $A' = \mathbb{N}$. Insbesondere, für jedes $n \in \mathbb{N}$, liegt $n + 1$ in A', d.h., $\mathbb{N}_{<n+1} \subset A$, und daher $n \in A$. □

Der Beweis vom Lemma 1.1.12 war eine Anwendung des Prinzips der vollständigen Induktion.

1.3 Abbildungen

Seien X, Y Mengen. Eine *Abbildung* f von X nach Y soll etwas sein, das jedem Element x von X ein Element $f(x)$ von Y zuordnet. Folgende Definition ist der präzise mengentheoretische Ausdruck dieser Idee:

Definition 1.3.1 (Abbildung, Wert, Urbild, Definitionsmenge, Zielmenge, Graph)**.** Eine *Abbildung* ist ein Tripel $f = (X, Y, \Gamma)$, wobei X und Y Mengen sind und $\Gamma \subset X \times Y$ eine Teilmenge ihres kartesischen Produkts ist, mit folgender Eigenschaft:

$$\text{Zu jedem } x \in X \text{ gibt es } \textit{genau ein } y \in Y \text{ mit } (x, y) \in \Gamma.$$

Man sagt „f ist eine Abbildung *von X nach Y*“. Das einzige Element $y \in Y$ mit $(x, y) \in \Gamma$ heißt der *Wert* von f in x und wird mit $f(x)$ bezeichnet. Man sagt auch, dass x ein *Urbild* von y unter f ist.

Die Menge X heißt *Definitionsmenge* oder *Definitionsbereich* von f.
Die Menge Y heißt *Zielmenge* oder *Zielbereich* von f.
Die Menge Γ heißt *Graph* von f und wird auch mit Γ_f bezeichnet.

Notation 1.3.2. Die Notation $f\colon X \to Y$ bedeutet, dass f eine Abbildung von X nach Y ist. In diesem Zusammenhang, die Notation $x \mapsto y$ bedeutet, dass y der Wert von f in x ist, d.h., $y = f(x)$. Man sagt auch „f bildet x auf y ab“.

Notation 1.3.3. Wir bezeichnen die Menge aller Abbildungen von X nach Y mit $\mathrm{Abb}(X, Y)$ oder Y^X. Sie ist eine Teilmenge von $\{X\} \times \{Y\} \times \mathcal{P}(X \times Y)$.

Bemerkung 1.3.4. Abbildungen heißen auch *Funktionen*, aber das Wort „Funktion“ wird oft für Abbildungen nach $\mathbb{R}$ oder $\mathbb{C}$ vorbehalten.

Beispiel 1.3.5. Abbildungen können auf verschiedene Weise definiert werden.

(i) Wenn die Definitionsmenge endlich ist, kann man einfach alle Werte ausdrücklich geben. Zum Beispiel, folgende Liste definiert eine Abbildung f von $\{0,1,2\}$ nach $\mathbb{N}$:

$$\begin{aligned} f\colon \{0,1,2\} &\to \mathbb{N},\\ 0 &\mapsto 5,\\ 1 &\mapsto 1,\\ 2 &\mapsto 5. \end{aligned}$$

(ii) Man kann auch eine Abbildung durch eine „Formel“ definieren. Zum Beispiel:

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R},\\ x &\mapsto 2x^2+1. \end{aligned}$$

(iii) Man kann Abbildungen durch *Fallunterscheidung* definieren. Zum Beispiel:

$$f\colon \mathbb{R} \to \mathbb{R},\quad x \mapsto \begin{cases} 2x^2+1, & \text{falls } x \geq 0,\\ x+1, & \text{falls } x < 0. \end{cases}$$

Hier ist es wichtig, dass die beide Fälle „$x \geq 0$“ und „$x < 0$“ miteinander ausschließlich sind und den ganzen Definitionsbereich überdecken.

(iv) Ein weiteres Beispiel ist:

$$\begin{aligned} f\colon \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\} &\to \mathbb{N},\\ A &\mapsto \text{das kleinste Element von } A. \end{aligned}$$

Diese Definition ist sinnvoll, da jede nichtleere Teilmenge von $\mathbb{N}$ ein kleinstes Element enthält (nach dem Wohlordnungsprinzip 1.2.19), und dieses Element eindeutig ist.

(v) Die folgende Variante von (iv) ist *keine* wohldefinierte Abbildung, weil die leere Teilmenge kein kleinstes Element besitzt:

$$\begin{aligned} f\colon \mathcal{P}(\mathbb{N}) &\to \mathbb{N},\\ A &\mapsto \text{das kleinste Element von } A. \end{aligned}$$

Auch folgende Variante ist nicht sinnvoll, da die meisten Teilmengen von $\mathbb{N}$ mehr als ein Element enthalten:

$$\begin{aligned} f\colon \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\} &\to \mathbb{N},\\ A &\mapsto \text{ein Element von } A. \end{aligned}$$

Beispiel 1.3.6. Algebraische Operationen zwischen Zahlen, wie $+$, $-$ oder $\cdot$, können mithilfe des kartesischen Produkts als Abbildungen aufgefasst werden. Zum Beispiel, die Addition und Multiplikation von natürlichen Zahlen sind Abbildungen

$$\begin{aligned} +\colon \mathbb{N}\times\mathbb{N} &\to \mathbb{N}, \quad (n,m) \mapsto n+m,\\ \cdot\colon \mathbb{N}\times\mathbb{N} &\to \mathbb{N}, \quad (n,m) \mapsto n\cdot m. \end{aligned}$$

Bemerkung 1.3.7. Zu jeder Menge X gibt es genau eine Abbildung von $\varnothing$ nach X. Denn das kartesische Product $\varnothing \times X = \varnothing$ hat genau eine Teilmenge, nämlich $\varnothing$, und das Tripel $(\varnothing, X, \varnothing)$ ist eine Abbildung, weil eine Aussage der Gestalt „für alle $x \in \varnothing \ldots$“ immer wahr ist. Auf der anderen Seite gibt es eine Abbildung von X nach $\varnothing$, nur wenn X leer ist.

Bemerkung 1.3.8 (Gleichheit von Abbildungen)**.** Zwei Abbildungen f und g sind genau dann gleich, wenn sie dieselbe Definitionsmenge und dieselbe Zielmenge haben, und außerdem gilt $f(x) = g(x)$ für alle x aus der gemeinsamen Definitionsmenge.

Definition 1.3.9 (Bild, Urbild)**.** Sei $f\colon X \to Y$ eine Abbildung von X nach Y.

- Sei $A \subset X$ eine Teilmenge. Das *Bild* von A unter f ist die Menge

$$f(A) := \{f(x) \mid x \in A\} \subset Y.$$

  Das Bild von X selbst, $f(X)$, heißt die *Bildmenge* oder das *Bild* von f.

- Sei $B \subset Y$ eine Teilmenge. Das *Urbild* von B unter f ist die Menge

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

Bemerkung 1.3.10. Sei $f\colon X \to Y$ eine Abbildung. Die Notation $f(x)$ hat jetzt zwei verschiedene Bedeutungen: Wenn x ein Element von X ist, dann ist $f(x)$ ein Element von Y, nämlich der Wert von f in x. Aber wenn x eine Teilmenge von X ist, dann ist $f(x)$ eine Teilmenge von Y, nämlich das Bild von x unter f. Es könnte leider sein, dass x ein Element *sowie* eine Teilmenge von X ist (z.B. $X = \{\varnothing\}$ und $x = \varnothing$). In diesem Fall haben wir einen Problem, da $f(x)$ nicht wohldefiniert ist. Zum Glück ist das kein ernstes Problem: in der Praxis wird es immer klar sein, ob wir x als Element oder als Teilmenge von X auffassen. In der Mathematik sind solche harmlosen Zweideutigkeiten ziemlich häufig, weil es viel mehr mathematische Begriffe als verfügbare Symbole/Namen gibt. Ein anderes Beispiel: Das Wort „Urbild“ hat schon zwei verschiedene Bedeutungen (Definitionen 1.3.1 und 1.3.9)!

Proposition 1.3.11 (Eigenschaften des Bilds und des Urbilds)**.** *Sei $f\colon X \to Y$ eine Abbildung.*

(i) *Für jede Teilmenge $A \subset X$ gilt $A \subset f^{-1}(f(A))$.*

(ii) *Für jede Teilmenge $B \subset Y$ gilt $f(f^{-1}(B)) \subset B$.*

(iii) *Für jede Teilmengen $A, A' \subset X$ gelten:*

$$\begin{aligned} f(A \cup A') &= f(A) \cup f(A'), \\ f(A \cap A') &\subset f(A) \cap f(A'), \\ f(A \setminus A') &\supset f(A) \setminus f(A'). \end{aligned}$$

(iv) *Für jede Teilmengen $B, B' \subset Y$ gelten:*

$$\begin{aligned} f^{-1}(B \cup B') &= f^{-1}(B) \cup f^{-1}(B'), \\ f^{-1}(B \cap B') &= f^{-1}(B) \cap f^{-1}(B'), \\ f^{-1}(B \setminus B') &= f^{-1}(B) \setminus f^{-1}(B'). \end{aligned}$$

Beweis. Jede Aussage folgt unmittelbar aus den Definitionen. □

Definition 1.3.12 (Identität, Komposition)**.**

- Sei X eine Menge. Die *Identität* auf X ist die Abbildung

$$\begin{aligned} \mathrm{id}_X\colon X &\to X, \\ x &\mapsto x. \end{aligned}$$

  Das heißt: $\mathrm{id}_X = (X, X, \Delta_X)$, wobei $\Delta_X = \{(x, x) \mid x \in X\}$ die diagonale Teilmenge ist.

- Seien $f\colon X \to Y$ und $g\colon Y \to Z$ Abbildungen. Die *Komposition* oder *Verkettung* von f und g ist die Abbildung

$$\begin{aligned} g \circ f\colon X &\to Z, \\ x &\mapsto g(f(x)). \end{aligned}$$

  Die Notation $g \circ f$ wird als „g nach f“ gelesen.

Proposition 1.3.13 (Eigenschaften der Komposition)**.**

(i) *Sei $f\colon X \to Y$ eine Abbildung. Dann $f \circ \mathrm{id}_X = f$ und $\mathrm{id}_Y \circ f = f$.*

(ii) (Assoziativität der Komposition) *Seien $f\colon X \to Y$, $g\colon Y \to Z$ und $h\colon Z \to W$ drei Abbildungen. Dann gilt*

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

Beweis. Um zu beweisen, dass zwei Abbildungen mit derselben Definitions- und Zielmenge gleich sind, ist zu zeigen, dass sie denselben Wert in jedem Element ihrer Definitionsmenge nehmen (Bemerkung 1.3.8). Dies folgt unmittelbar aus den Definitionen von id und $\circ$. Zum Beispiel, für jedes $x \in X$ gilt:

$$\begin{aligned} (f \circ \mathrm{id}_X)(x) &= f(\mathrm{id}_X(x)) && \text{(Definition von } \circ) \\ &= f(x) && \text{(Definition von } \mathrm{id}_X). \end{aligned}$$ □

Definition 1.3.14 (Einschränkung)**.** Sei $f\colon X \to Y$ eine Abbildung und $A \subset X$ eine Teilmenge. Die *Einschränkung* von f auf A ist die wie folgt definierte Abbildung:

$$\begin{aligned} f|_A\colon A &\to Y, \\ x &\mapsto f(x). \end{aligned}$$

Das heißt: $f|_A = (A, Y, \Gamma_f \cap (A \times Y))$.

Definition 1.3.15 (Inklusionsabbildung)**.** Sei X eine Menge und $A \subset X$ eine Teilmenge. Die Abbildung

$$\begin{aligned} i_A\colon A &\to X, \\ x &\mapsto x, \end{aligned}$$

heißt die *Inklusionsabbildung* oder *Inklusion* von A in X.

Bemerkung 1.3.16. Es gilt $i_A = \mathrm{id}_X|_A$ und $f|_A = f \circ i_A$.

Definition 1.3.17 (kanonische Projektionen)**.** Seien A, B Mengen. Die Abbildungen

$$\begin{aligned} \pi_1\colon A \times B &\to A, \\ (a,b) &\mapsto a, \end{aligned} \qquad \begin{aligned} \pi_2\colon A \times B &\to B, \\ (a,b) &\mapsto b, \end{aligned}$$

heißen die *kanonischen Projektionen* aus $A \times B$ auf die Faktoren.

Allgemeiner, sei $(A_i)_{i \in I}$ eine Mengenfamilie mit Indexmenge I und sei $e \in I$. Die e-te kanonische Projektion ist die Abbildung

$$\begin{aligned} \pi_e\colon \prod_{i \in I} A_i &\to A_e, \\ (a_i)_{i \in I} &\mapsto a_e. \end{aligned}$$

Bemerkung 1.3.18. Das Wort „kanonisch“ hat keine präzise Bedeutung in der Mathematik, aber es ist trotzdem häufig verwendet. Es könnte viele verschiedene Abbildungen $A \times B \to A$ sein, aber ohne weitere Informationen gibt es nur eine, die „besonders“ ist, nämlich π_1. Deswegen ist π_1 die „kanonische Abbildung“ $A \times B \to A$. Ein anderes Beispiel: Wenn A eine Teilmenge von X ist, dann würde die kanonische Abbildung $A \to X$ die Inklusionsabbildung sein.

Definition 1.3.19 (injektiv, surjektiv, bijektiv)**.** Sei $f\colon X \to Y$ eine Abbildung.

- f heißt *injektiv*, wenn jedes Element von Y *höchstens ein* Urbild unter f besitzt.
- f heißt *surjektiv*, wenn jedes Element von Y *mindestens ein* Urbild unter f besitzt.
- f heißt *bijektiv*, wenn jedes Element von Y *genau ein* Urbild unter f besitzt.

Nach Definition gilt also: bijektiv $\iff$ injektiv und surjektiv. Eine injektive/surjektive/bijektive Abbildung wird auch als Injektion/Surjektion/Bijektion bezeichnet.

Beispiel 1.3.20.

(i) Die Abbildung $f\colon \{0,1,2\} \to \mathbb{N}$ aus Beispiel 1.3.5(i) ist nicht injektiv, da sie denselben Wert 5 in zwei verschiedenen Elementen nimmt (d.h., 5 hat zwei Urbilder unter f, nämlich 0 und 2). Sie ist nicht surjektiv, da $57 \notin f(\{0,1,2\})$ (d.h., 57 hat kein Urbild unter f).

(ii) Folgende Abbildung ist injektiv (und nicht surjektiv):

$$\begin{aligned} f\colon \{0,1,2\} &\to \mathbb{N}, \\ 0 &\mapsto 5, \\ 1 &\mapsto 1, \\ 2 &\mapsto 0. \end{aligned}$$

(iii) Die Abbildung $f\colon \mathbb{R} \to \mathbb{R}$ aus Beispiel 1.3.5(ii) ist weder injektiv noch surjektiv. Die aus Beispiel 1.3.5(iii) ist bijektiv.

(iv) Die Abbildung $f\colon \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\} \to \mathbb{N}$ aus Beispiel 1.3.5(iv) ist surjektiv aber nicht injektiv.

(v) Die Abbildung

$$\begin{aligned} &\mathbb{Z} \to \mathbb{N}, \\ &n \mapsto \begin{cases} 2n, & \text{falls } n \geq 0, \\ -(2n+1), & \text{falls } n < 0, \end{cases} \end{aligned}$$

ist bijektiv.

(vi) Die einzige Abbildung $\varnothing \to X$ ist injektiv (es gibt nichts zu zeigen!). Sie ist genau dann surjektiv, wenn X leer ist.

Beispiel 1.3.21. Seien A und B zwei Mengen. Es gibt eine kanonische Abbildung

$$\begin{aligned} A \sqcup B &\to A \cup B, \\ (1,a) &\mapsto a, \\ (2,b) &\mapsto b. \end{aligned}$$

(Nach Definition 1.2.14 ist $A \sqcup B$ eine Teilmenge von $\{1,2\} \times (A \cup B)$, und diese Abbildung ist die Einschränkung der zweiten kanonischen Projektion π_2.) Diese Abbildung ist immer surjektiv. Sie ist genau dann injektiv (und damit bijektiv), wenn A und B disjunkt sind.

Definition 1.3.22 (Umkehrabbildung)**.** Sei $f\colon X \to Y$ eine Abbildung. Eine Abbildung $g\colon Y \to X$ heißt *Umkehrabbildung* oder *inverse Abbildung* von f, falls

$$g \circ f = \mathrm{id}_X \quad \text{und} \quad f \circ g = \mathrm{id}_Y .$$

Wenn sie existiert, eine Umkehrabbildung von $f\colon X \to Y$ ist *eindeutig bestimmt*, denn: Sind g und g' zwei Umkehrabbildungen von f, so gilt

$$\begin{aligned}
g &= g \circ \mathrm{id}_Y && \text{(Proposition 1.3.13(i))}\\
&= g \circ (f \circ g') && (g' \text{ invers zu } f)\\
&= (g \circ f) \circ g' && \text{(Proposition 1.3.13(ii))}\\
&= \mathrm{id}_X \circ g' && (g \text{ invers zu } f)\\
&= g'. && \text{(Proposition 1.3.13(i)).}
\end{aligned}$$

Deswegen kann man von *der* Umkehrabbildung von f sprechen, und sie mit f^{-1} bezeichnen.

Satz 1.3.23. *Sei $f\colon X \to Y$ eine Abbildung. Die folgenden Aussagen sind äquivalent:*

(i) *f ist bijektiv.*

(ii) *f besitzt eine Umkehrabbildung.*

Beweis. Zu (i) $\Rightarrow$ (ii). Sei $f = (X, Y, \Gamma)$ bijektiv. Wir betrachten die Menge

$$\Gamma' := \{(y, x) \in Y \times X \mid (x, y) \in \Gamma\}.$$

Dann ist $g = (Y, X, \Gamma')$ eine Abbildung, denn: Sei $y \in Y$. Da f injektiv ist, gibt es höchstens ein $x \in X$ mit $(y, x) \in \Gamma'$. Da f surjetiv ist, gibt es mindestens ein $x \in X$ mit $(y, x) \in \Gamma'$. Also gibt es genau ein $x \in X$ mit $(y, x) \in \Gamma'$, d.h., g ist eine Abbildung. Nach Konstruktion gilt $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$, d.h., g ist die Umkehrabbildung von f.

Zu (ii) $\Rightarrow$ (i). Sei g eine Umkehrabbildung von f, und sei $y \in Y$. Zu zeigen ist, dass y genau ein Urbild unter f besitzt. Da $f \circ g = \mathrm{id}_Y$ ist $g(y)$ ein Urbild von y unter f, also gibt es mindestens ein Urbild. Seien x, x' zwei Urbilder von y unter f, d.h., $f(x) = y$ und $f(x') = y$. Da $g \circ f = \mathrm{id}_X$ gilt

$$x = g(f(x)) = g(y) = g(f(x')) = x'.$$

Also ist das Urbild von y eindeutig, wie gewünscht. □

Beispiel 1.3.24. Zu jeder Menge X gibt es eine bijektive Abbildung

$$\begin{aligned}
\mathrm{Abb}(X, \{0, 1\}) &\to \mathcal{P}(X),\\
f &\mapsto f^{-1}(\{1\}).
\end{aligned}$$

Die Umkehrabbildung bildet eine Teilmenge $A \in \mathcal{P}(X)$ auf die Abbildung $\chi_A\colon X \to \{0, 1\}$ ab, die durch

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

definiert wird. Die Abbildung χ_A heißt die *charakteristische Funktion* der Teilmenge A.

Proposition 1.3.25. *Seien X, Y, Z Mengen und seien $f\colon X \to Y$ und $g\colon Y \to Z$ Abbildungen.*

(i) *Sind f und g injektiv, so ist $g \circ f$ injektiv.*

(ii) *Sind f und g surjektiv, so ist $g \circ f$ surjektiv.*

(iii) *Ist $g \circ f$ injektiv, so ist f injektiv.*

(iv) *Ist $g \circ f$ surjektiv, so ist g surjektiv.*

Beweis. Wir beweisen stellvertretend (i) und (iv).

Zu (i). Zu zeigen ist, dass jedes $z \in Z$ höchstens ein Urbild unter $g \circ f$ besitzt. Seien $x, x' \in X$ zwei Urbilder von z, d.h., $g(f(x)) = g(f(x')) = z$. Aus der Injektivität von g folgt $f(x) = f(x')$, und aus der Injektivität von f folgt wiederum $x = x'$.

Zu (iv). Zu zeigen ist, dass jedes $z \in Z$ mindestens ein Urbild unter g besitzt. Da $g \circ f$ surketiv ist, gibt es ein $x \in X$ mit $g(f(x)) = z$. Insbesondere ist $f(x)$ ein Urbild von z unter g. □

Notation 1.3.26. Wenn wir die Injektivität bzw. Surjektivität einer Abbildung $f\colon X \to Y$ betonen wollen, schreiben wir manchmal $f\colon X \hookrightarrow Y$ bzw. $f\colon X \twoheadrightarrow Y$. Eine alternative Notation für injektive Abbildungen ist $f\colon X \rightarrowtail Y$. Wenn f bijektiv ist schreiben wir auch $f\colon X \xrightarrow{\sim} Y$.

Es ist oft nützlich, eine Abbildung $I \to X$ als eine „Familie von Elementen von X“ aufzufassen:

Definition 1.3.27 (Familie, Folge)**.** Seien I und X Mengen. Eine *Familie* $(x_i)_{i \in I}$ in X mit Indexmenge I ist einfach eine Abbildung

$$\begin{aligned} I &\to X, \\ i &\mapsto x_i. \end{aligned}$$

Eine *Folge* $(x_n)_{n \in \mathbb{N}}$ in X ist eine Familie in X mit Indexmenge $\mathbb{N}$.

1.3.1 Mächtigkeit

Definition 1.3.28 (Gleichmächtigkeit)**.** Zwei Mengen X and Y sind *gleichmächtig*, in Zeichen $|X| = |Y|$, wenn eine bijektive Abbildung $X \xrightarrow{\sim} Y$ existiert.

Definition 1.3.29 (Endlichkeit, Abzählbarkeit)**.** Sei X eine Menge.

- X heißt *endlich*, wenn eine natürliche Zahl n existiert, so dass X und $\{1, 2, \ldots, n\}$ gleichmächtig sind. Die natürliche Zahl n ist dann eindeutig bestimmt; sie heißt die *Mächtigkeit* oder *Kardinalität* von X, und wird mit $|X|$ bezeichnet.
- X heißt *unendlich*, wenn sie nicht endlich ist.
- X heißt *abzählbar*, wenn entweder X endlich ist oder X und $\mathbb{N}$ gleichmächtig sind.
- X heißt *überabzählbar*, wenn sie nicht abzählbar ist.

Bemerkung 1.3.30. In der Definition 1.3.29 betrachten wir die Menge $\{1, 2, \ldots, n\}$ mit n einer beliebigen natürlichen Zahl. Wenn $n = 0$ verstehen wir $\{1, 2, \ldots, n\}$ als die leere Menge. Insbesondere ist eine Menge genau dann leer, wenn ihre Mächtigkeit gleich Null ist.

Beispiel 1.3.31.

(i) $\mathbb{N}$, $\mathbb{Z}$, und $\mathbb{Q}$ sind paarweise gleichmächtig, und damit abzählbar. Eine bijektive Abbildung zwischen $\mathbb{N}$ und $\mathbb{Z}$ wurde im Beispiel 1.3.20(v) gegeben. Die Abzählbarkeit von $\mathbb{Q}$ folgt aus Cantors erstem Diagonalargument.

(ii) $\mathbb{R}$ und $\mathbb{C}$ sind gleichmächtig und überabzählbar. Die Unabzählbarkeit von $\mathbb{R}$ folgt aus Cantors zweitem Diagonalargument.

(iii) Eine Menge X und ihre Potenzmenge $\mathcal{P}(X)$ sind nie gleichmächtig, d.h., $\mathcal{P}(X)$ ist „wirklich größer“ als X. Denn es wäre eine surjektive Abbildung $f\colon X \to \mathcal{P}(X)$. Sei $M = \{x \in X \mid x \notin f(x)\}$. Da f surjektiv ist, existiert ein $m \in X$ mit $f(m) = M$. Nach Definition von M gilt dann: $m \in M \iff m \notin M$, was ein Widerspruch ist.

(iv) Man kann zeigen, dass die Menge $\mathbb{N}^{\mathbb{N}}$ aller Abbildungen von $\mathbb{N}$ nach $\mathbb{N}$ (alias Folgen in $\mathbb{N}$) zu $\mathbb{R}$ gleichmächtig ist. Die Teilmenge $\mathbb{N}^{(\mathbb{N})} \subset \mathbb{N}^{\mathbb{N}}$ aller Folgen, die schließlich null sind, ist aber abzählbar. Man kann eine bijektive Abbildung von $\mathbb{N}^{(\mathbb{N})}$ nach $\mathbb{N}$ explizit definieren:

$$\begin{aligned} \mathbb{N}^{(\mathbb{N})} &\to \mathbb{N}, \\ (a_n)_{n\in\mathbb{N}} &\mapsto (p_0^{a_0} p_1^{a_1} p_2^{a_2} \dots) - 1, \end{aligned}$$

wobei $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, usw. alle Primzahlen in aufsteigender Reihenfolge sind. Hierbei werden der *Satz von Euklid* und der *Fundamentalsatz der Arithmetik* verwendet (Sätze 1.1.11 und 1.1.14): Es gibt unendlich viele Primzahlen, und jede natürliche Zahl ≥ 1 lässt sich eindeutig als Produkt von Primzahlen darstellen.

Bemerkung 1.3.32. Seien A, B endliche Mengen und seien $a, b \in \mathbb{N}$ ihre jeweiligen Mächtigkeiten. Dann:

- Die Mächtigkeit der Summe $A \sqcup B$ ist $a + b$.
- Die Mächtigkeit des Produkts $A \times B$ ist ab.
- Die Mächtigkeit der Menge $\mathrm{Abb}(A, B)$ ist b^a. (Deswegen wird diese Menge auch mit B^A bezeichnet.)

Seien X und Y endliche Mengen. Es gilt $|X| \leq |Y|$ genau dann, wenn eine injektive Abbildung $X \hookrightarrow Y$ existiert. Also wenn injektive Abbildungen von X nach Y sowie von Y nach X existieren, dann sind X und Y gleichmächtig. Folgender Satz verallgemeinert diese Beobachtung auf unendliche Mengen:

Satz 1.3.33 (Satz von Cantor–Bernstein–Schröder)**.** *Seien X, Y Mengen. Wenn injektive Abbildungen $f\colon X \hookrightarrow Y$ und $g\colon Y \hookrightarrow X$ existieren, dann sind X und Y gleichmächtig.*

**Beweis.* Wir definieren eine Folge von Teilmengen $A_0, A_1, A_2, \dots \subset X$ wie folgt:

$$\begin{aligned} A_0 &= X \setminus g(Y), \\ A_{n+1} &= g(f(A_n)), \end{aligned}$$

und wir setzen

$$A := \bigcup_{n\in\mathbb{N}} A_n \subset X.$$

Nach Definition gilt

$$g(f(A)) = g\left(f\left(\bigcup_{n\in\mathbb{N}} A_n\right)\right) = \bigcup_{n\in\mathbb{N}} g(f(A_n)) = \bigcup_{n\in\mathbb{N}} A_{n+1},$$

und daher

$$A = A_0 \cup g(f(A)). \tag{1.3.34}$$

Es gilt $X \setminus g(Y) = A_0 \subset A$, und daher $X \setminus A \subset g(Y)$. Das heißt, jedes $x \in X \setminus A$ liegt im Bild von g. Da g injektiv ist, gibt es eigentlich *genau ein* Urbild von x unter g, das wir mit $g^{-1}(x)$ bezeichnen. Ebenso hat jedes $y \in f(A)$ genau ein Urbild $f^{-1}(y)$ unter f. Deswegen darf man definieren:

$$\begin{aligned} h\colon X &\to Y, & k\colon Y &\to X, \\ x &\mapsto \begin{cases} f(x), & \text{falls } x \in A, \\ g^{-1}(x), & \text{falls } x \in X \setminus A. \end{cases} & y &\mapsto \begin{cases} g(y), & \text{falls } y \in Y \setminus f(A), \\ f^{-1}(y), & \text{falls } y \in f(A). \end{cases} \end{aligned}$$

Wir behaupten, dass h bijektiv ist, mit Umkehrabbildung k.

- $k \circ h = \mathrm{id}_X$: Sei $x \in X$. Ist $x \in A$, so ist $f(x) \in f(A)$, und daher $k(h(x)) = k(f(x)) = f^{-1}(f(x)) = x$. Ist $x \notin A$, so ist $g^{-1}(x) \notin f(A)$, sonst wäre $x = g(g^{-1}(x)) \in g(f(A)) \subset A$ nach (1.3.34). Also gilt $k(h(x)) = k(g^{-1}(x)) = g(g^{-1}(x)) = x$.
- $h \circ k = \mathrm{id}_Y$: Sei $y \in Y$. Ist $y \in f(A)$, so ist $f^{-1}(y) \in A$, und daher $h(k(y)) = h(f^{-1}(y)) = f(f^{-1}(y)) = y$. Ist $y \notin f(A)$, so ist $g(y) \notin g(f(A))$ nach der Injektivität von g und $g(y) \notin A_0$ nach Definition von A_0, also $g(y) \notin A$ nach (1.3.34). Also gilt $h(k(y)) = h(g(y)) = g^{-1}(g(y)) = y$. □

Bemerkung 1.3.35. Es ist auch der Fall, dass X und Y gleichmächtig sind, wenn surjektive Abbildungen $f\colon X \twoheadrightarrow Y$ und $g\colon Y \twoheadrightarrow X$ existieren, oder wenn eine injektive Abbildung $f\colon X \hookrightarrow Y$ sowie eine surjektive Abbildung $g\colon X \twoheadrightarrow Y$ existieren. Diese Aussagen folgen aus dem Satz 1.3.33, weil jede surjektive Abbildung $f\colon X \twoheadrightarrow Y$ einen Schnitt $s\colon Y \to X$ besitzt, der eine injektive Abbildung ist (siehe Proposition 1.4.14(i)).

Nach dem Satz von Cantor–Bernstein–Schröder kann man von der Mächtigkeit unendlicher Mengen vernünftig sprechen. Man sagt zum Beispiel, dass die Mächtigkeit von X kleiner oder gleich der von Y ist, in Zeichen $|X| \leq |Y|$, wenn eine injektive Abbildung von X nach Y existiert. Dann sind zwei Mengen X und Y genau dann gleichmächtig, wenn $|X| \leq |Y|$ und $|Y| \leq |X|$. Eine Menge X ist genau dann abzählbar, wenn $|X| \leq |\mathbb{N}|$, und sie ist genau dann unendlich, wenn $|\mathbb{N}| \leq |X|$. Man kann auch zeigen, dass je zwei Mengen X, Y vergleichbare Mächtigkeiten haben, d.h., es gilt $|X| \leq |Y|$ oder $|Y| \leq |X|$ (dazu braucht man das Auswahlaxiom, siehe Abschnitt 1.4.2).

Man soll beachten, dass sich der Begriff der Mächtigkeit bei unendlichen Mengen manchmal anders als bei endlichen Mengen verhält. Zum Beispiel:

***Satz 1.3.36** (Mächtigkeit unendlicher Vereinigungen)**.** *Sei $(A_i)_{i\in I}$ eine Mengenfamilie mit Vereinigung $A = \bigcup_{i\in I} A_i$. Ist I unendlich und gilt $|A_i| \leq |I|$ für alle $i \in I$, so gilt auch $|A| \leq |I|$.*

Insbesondere: Eine abzählbare Vereinigung abzählbarer Mengen ist wieder abzählbar.

1.4 Relationen

Definition 1.4.1 (Relation, reflexiv, transitiv, symmetrisch, antisymmetrisch, total)**.** Sei X eine Menge. Eine *Relation* R auf X ist eine Teilmenge $R \subset X \times X$. Man sagt „x steht in Relation zu y bzgl. R“ und schreibt xRy, falls $(x, y) \in R$.

Eine Relation R auf X heißt:

- *reflexiv*, wenn xRx für alle $x \in X$;
- *transitiv*, wenn für alle $x, y, z \in X$,
$$xRy \text{ und } yRz \Longrightarrow xRz;$$
- *symmetrisch*, wenn für alle $x, y \in X$,
$$xRy \Longrightarrow yRx;$$
- *antisymmetrisch*, wenn für alle $x, y \in X$,
$$xRy \text{ und } yRx \Longrightarrow x = y;$$
- *total*, wenn für alle $x, y \in X$, xRy oder yRx.

Definition 1.4.2 (Äquivalenzrelation, partielle Ordnung, totale Ordnung)**.** Sei R eine Relation auf einer Menge X.

- R heißt *Äquivalenzrelation*, falls R reflexiv, transitiv und symmetrisch ist.
- R heißt *partielle Ordnung*, falls R reflexiv, transitiv und antisymmetrisch ist. Man sagt dann auch, dass das Paar (X, R) eine partiell geordnete Menge ist.
- R heißt *totale Ordnung*, falls R eine partielle Ordnung ist und außerdem total ist. Man sagt dann auch, dass das Paar (X, R) eine total geordnete Menge ist.

Beispiel 1.4.3.

(i) Die Identitäts- oder Gleichheitsrelation $=$ zwischen Elementen von X ist eine Äquivalenzrelation auf X. Sie entspricht der diagonalen Teilmenge

$$\Delta_X = \{(x, x) \mid x \in X\} \subset X \times X.$$

(ii) Die gewöhnliche Relation $\leq$ auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ bzw. $\mathbb{R}$ ist eine totale Ordnung.

(iii) Die Inklusionsrelation $\subset$ auf der Potenzmenge $\mathcal{P}(X)$ ist eine partielle Ordnung. Hat X mindestens zwei Elemente, so ist diese Ordnung *nicht* total.

(iv) Gleichmächtigkeit ist eine Äquivalenzrelation auf $\mathcal{P}(X)$. Denn seien A, B, C Teilmengen von X:

- *Zur Reflexivität:* Die Identität $\mathrm{id}_A \colon A \to A$ ist bijektiv, also ist A zu sich selbst gleichmächtig.
- *Zur Symmetrie:* Nach Satz 1.3.23 besitzt jede bijektive Abbildung eine Umkehrabbildung, die wieder bijektiv ist.
- *Zur Transitivität:* Nach Proposition 1.3.25(i,ii) ist die Komposition zweier bijektiven Abbildungen wieder bijektiv.

(v) Die Teilbarkeitsrelation $|$ auf $\mathbb{N}$ ist so definiert:

$$n|m \iff \text{es existiert } k \in \mathbb{N} \text{ mit } m = kn.$$

Sie ist eine partielle Ordnung auf $\mathbb{N}$, die nicht total ist.

Man kann ebenso die Teilbarkeitsrelation $|$ auf $\mathbb{Z}$ definieren. Auf $\mathbb{Z}$ ist diese Relation immer noch reflexiv und transitiv, aber sie ist nicht mehr antisymmetrisch, weil z.B. $2|-2$ und $-2|2$.

1.4.1 Quotient einer Menge modulo einer Äquivalenzrelation

Definition 1.4.4 (Äquivalenzklasse, Quotientenmenge, Quotientenabbildung)**.** Sei $\sim$ eine Äquivalenzrelation auf einer Menge X.

- Sei $x \in X$. Die Menge
$$[x] := \{y \in X \mid y \sim x\} \subset X$$
heißt die *Äquivalenzklasse* von x bzgl. $\sim$. Ein Element $y \in [x]$ heißt auch *Repräsentant* der Äquivalenzklasse von x.
- Die *Quotientenmenge* von X bzgl. $\sim$ ist die Menge aller Äquivalenzklassen
$$X/\sim := \{[x] \mid x \in X\} \subset \mathcal{P}(X)$$
(gelesen „X modulo $\sim$“ oder „X durch $\sim$“).

- Die surjektive Abbildung

$$X \twoheadrightarrow X/\!\sim,$$
$$x \mapsto [x],$$

heißt die *Quotientenabbildung* oder die *kanonische Abbildung.*

Definition 1.4.5 (Partition)**.** Sei X eine Menge. Eine *Partition* von X ist eine Menge $\mathcal{A}$ von nichtleeren Teilmengen von X, d.h., $\mathcal{A} \subset \mathcal{P}(X) \setminus \{\varnothing\}$, so dass jedes Element von X in *genau einem* Element von $\mathcal{A}$ liegt.

Proposition 1.4.6. *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X. Dann ist $X/\!\sim$ eine Partition von X.*

Beweis. Da $\sim$ reflexiv ist, liegt jedes x in seiner Äquivalenzklasse $[x]$. Insbesondere besteht $X/\!\sim$ aus nichtleeren Teilmengen von X. Seien $[x_1], [x_2] \in X/\!\sim$ zwei Äquivalenzklassen mit $x \in [x_1]$ und $x \in [x_2]$. Ziel ist zu zeigen, dass $[x_1] = [x_2]$. Da die Situation symmetrisch ist, genügt es zu zeigen, dass $[x_1] \subset [x_2]$. Sei $y \in [x_1]$. Nach Definition von $[-]$ gilt $x \sim x_1$, $x \sim x_2$ und $y \sim x_1$. Da $\sim$ symmetrisch ist, gilt $x_1 \sim x$. Da $\sim$ transitiv ist und $y \sim x_1 \sim x \sim x_2$, gilt $y \sim x_2$, d.h., $y \in [x_2]$. □

Bemerkung 1.4.7. Umgekehrt, wenn $\mathcal{A} \subset \mathcal{P}(X)$ eine Partition von X ist, ist dann die Relation

$$x \sim_{\mathcal{A}} y \iff \text{es existiert } A \in \mathcal{A} \text{ mit } x, y \in A$$

eine Äquivalenzrelation auf X, so dass $X/\!\sim_{\mathcal{A}} = \mathcal{A}$.

Die Abbildungen $\sim \mapsto X/\!\sim$ und $\mathcal{A} \mapsto \sim_{\mathcal{A}}$ bilden eigentlich zueinander inverse Bijektionen zwischen der Menge aller Äquivalenzrelationen auf X und der Menge aller Partitionen von X.

Definition 1.4.8 (Repräsentantensystem)**.** Sei $\sim$ eine Äquivalenzrelation auf einer Menge X. Ein *Repräsentantensystem* für $\sim$ ist eine Teilmenge $Y \subset X$, so dass die auf Y eingeschränkte Quotientenabbildung $Y \to X/\!\sim$ bijektiv ist (d.h., jede Äquivalenzklasse enthält genau ein Element von Y).

Beispiel 1.4.9 (Kongruenzrelation)**.** Sei n eine natürliche Zahl und sei

$$n\mathbb{Z} := \{nx \mid x \in \mathbb{Z}\}.$$

Man definiert eine Relation $\equiv_n$ auf $\mathbb{Z}$ wie folgt:

$$x \equiv_n y \iff x - y \in n\mathbb{Z}.$$

Man schreibt üblicherweise

$$x \equiv y \mod n$$

(gelesen „x ist kongruent zu y modulo n“) statt $x \equiv_n y$. Man kann leicht nachprüfen, dass $\equiv_n$ eine Äquivalenzrelation auf $\mathbb{Z}$ ist. Die Quotientenmenge $\mathbb{Z}/\!\equiv_n$ wird mit $\mathbb{Z}/n\mathbb{Z}$ bezeichnet, und ihre Elemente heißen *Restklassen modulo n*. Falls $n \neq 0$, ist die Menge $\mathbb{Z}/n\mathbb{Z}$ endlich der Mächtigkeit n: Die Komposition der Inklusionsabbildung und der Quotientenabbildung

$$\{0, \ldots, n-1\} \hookrightarrow \mathbb{Z} \twoheadrightarrow \mathbb{Z}/n\mathbb{Z}$$

ist bijektiv. Das heißt, $\{0, \ldots, n-1\}$ ist ein Repräsentantensystem für $\equiv_n$.

Satz 1.4.10 (universelle Eigenschaft der Quotientenmenge)**.** *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X, $q\colon X \to X/\!\sim$ die Quotientenabbildung, und $f\colon X \to Y$ eine beliebige Abbildung. Folgende Aussagen sind äquivalent:*

(i) *Für alle* $x, x' \in X$ *gilt*

$$x \sim x' \implies f(x) = f(x'). \tag{1.4.11}$$

(ii) *Es existiert eine Abbildung* $\bar{f}\colon X/\!\sim \to Y$ *mit* $f = \bar{f} \circ q$.

Außerdem ist die Abbildung $\bar{f}$ *eindeutig bestimmt (wenn sie existiert).*

Beweis. Wir müssen die beiden Implikationen (i) ⇒ (ii) und (ii) ⇒ (i) beweisen, und dann auch die letzte Aussage.

Zu (i) ⇒ (ii). Wir betrachten die Teilmenge

$$\bar{\Gamma} = \{(A, y) \in X/\!\sim \times\, Y \mid y \in f(A)\} \subset X/\!\sim \times\, Y,$$

und wir setzen $\bar{f} := (X/\!\sim, Y, \bar{\Gamma})$. Nach Definition liegt ein Paar (A, y) in $\bar{\Gamma}$ genau dann, wenn es ein Element $x \in A$ gibt, so dass $f(x) = y$. Sind $x, x' \in A$ zwei solche Elemente, so gilt $x \sim x'$ und daher $f(x) = f(x')$ nach (1.4.11). Dies zeigt, dass es zu jeder Äquivalenzklasse $A \in X/\!\sim$ genau einem $y \in Y$ mit $(A, y) \in \bar{\Gamma}$ gibt. Also ist $\bar{f}$ eine Abbildung, und es ist jetzt klar, dass $\bar{f} \circ q = f$.

Zu (i) ⇒ (ii). Angenommen, $\bar{f}$ existiert. Seien $x, x' \in X$ mit $x \sim x'$. Dann gilt:

$$f(x) = \bar{f}(q(x)) = \bar{f}(q(x')) = f(x').$$

Also ist die Bedingung (1.4.11) erfüllt.

Zur Eindeutigkeit. Seien $\bar{f}$ und $\tilde{f}$ zwei Abbildungen $X/\!\sim \to Y$, so dass $f = \bar{f} \circ q$ und $f = \tilde{f} \circ q$. Sei $[x] \in X/\!\sim$ eine beliebige Äquivalenzklasse. Dann gilt:

$$\bar{f}([x]) = \bar{f}(q(x)) = f(x) = \tilde{f}(q(x)) = \tilde{f}([x]).$$

Also gilt $\bar{f} = \tilde{f}$. □

Die folgende Situation kommt sehr häufig vor: man möchte eine Abbildung $f\colon X/\!\sim \to Y$ durch

$$f([x]) = g(x)$$

definieren, wobei g eine gewisse Abbildung von X nach Y ist. Bei einer solchen Definition muss man immer nachprüfen, dass g der Bedingung (1.4.11) genügt, d.h., dass $x \sim x' \Rightarrow g(x) = g(x')$. In diesem Fall sagt man, dass f durch die obige Gleichung *wohldefiniert* ist. Sonst wäre eine solche Definition nicht einmal sinnvoll.

Bemerkung 1.4.12. Die Aussage vom Satz 1.4.10 kann man auch auf folgende Weise formulieren. Die Abbildung

$$\begin{aligned} \mathrm{Abb}(X/\!\sim, Y) &\to \mathrm{Abb}(X, Y), \\ g &\mapsto g \circ q, \end{aligned}$$

ist injektiv, und ihr Bild besteht genau aus diesen Abbildungen $f\colon X \to Y$, die die Bedingung (1.4.11) erfüllen.

Bemerkung 1.4.13 (kommutative Diagramme)**.** Eine Situation mit mehreren Mengen und Abbildungen zwischen denen lässt sich oft gut durch ein Diagramm veranschaulichen. Zum Beispiel: Das Dreieck

$$\begin{array}{ccc} X & \xrightarrow{\ h\ } & Z \\ {\scriptstyle f}\downarrow & \nearrow_{g} & \\ Y & & \end{array}$$

stellt drei Mengen X, Y, Z und drei Abbildungen f, g, h mit gegebenen Definitions- und Zielmengen dar. Ein solches Dreieck heißt *kommutativ*, wenn $h = g \circ f$. Im Allgemeinen heißt

ein Diagramm von Mengen und Abbildungen *kommutativ*, wenn folgendes gilt: Für alle zwei Mengen X, Y im Diagramm stimmen alle möglichen Kompositionen von Abbildungen von X nach Y überein. Beispielsweise ist das Quadrat

$$\begin{array}{ccc} X & \xrightarrow{h} & Z \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle k} \\ Y & \xrightarrow[g]{} & W \end{array}$$

genau dann kommutativ, wenn $g \circ f = k \circ h$.

Die universelle Eigenschaft der Quotientenmenge lässt sich dann wie folgt formulieren: Wenn $f\colon X \to Y$ die Implikation $x \sim x' \Rightarrow f(x) = f(x')$ erfüllt, dann existiert genau eine Abbildung $\bar{f}\colon X/\!\sim \to Y$, so dass folgendes Dreieck kommutiert:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y. \\ {\scriptstyle q}\downarrow & \nearrow_{\exists!\, \bar{f}} & \\ X/\!\sim & & \end{array}$$

1.4.2 Das Auswahlaxiom und das Zornsche Lemma

Ein besonderes Axiom der Mengenlehre ist das *Auswahlaxiom*:

> Sei $(A_i)_{i\in I}$ eine Mengenfamilie nichtleerer Mengen. Dann existiert eine Abbildung $w\colon I \to \bigcup_{i\in I} A_i$ mit $w(i) \in A_i$ für alle $i \in I$.

Eine solche Abbildung w heißt *Auswahlfunktion* für die Mengenfamilie $(A_i)_{i\in I}$. Die Auswahlfunktion w wählt ein Element $w(i)$ aus jedem A_i aus. Dieses Axiom scheint sehr intuitiv zu sein, und eigentlich kann man es aus den anderen Axiomen herleiten, wenn die Indexmenge I endlich ist. Es gibt viele hilfreiche Folgerungen des Auswahlaxioms:

Proposition 1.4.14 (Folgerungen des Auswahlaxioms)**.**

(i) *Sei $f\colon X \to Y$ eine surjektive Abbildung. Dann existiert ein* Schnitt *von f, d.h., eine Abbildung $s\colon Y \to X$ mit $f \circ s = \mathrm{id}_Y$.*

(ii) *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X. Dann existiert ein Repräsentantensystem für $\sim$.*

(iii) *Sei $(A_i)_{i\in I}$ eine Mengenfamilie nichtleerer Mengen. Dann ist das Produkt $\prod_{i\in I} A_i$ nicht leer.*

Beweis. Zu (i). Wir betrachten die Mengenfamilie $(f^{-1}(\{y\}))_{y\in Y}$, deren Vereinigung gleich X ist. Da f surjektiv ist, ist kein Urbild $f^{-1}(\{y\})$ leer. Nach dem Auswahlaxiom existiert also eine Abbildung $s\colon Y \to X$ mit $s(y) \in f^{-1}(\{y\})$ für alle $y \in Y$, d.h., $f \circ s = \mathrm{id}_Y$.

Zu (ii). Die Quotientenabbildung $\pi\colon X \to X/\!\sim$ is surjektiv. Nach (i) existiert ein Schnitt s von π. Das Bild $s(X/\!\sim) \subset X$ ist dann ein Repräsentantensystem für $\sim$.

Zu (iii). Sei w eine Auswahlfunktion für die Familie $(A_i)_{i\in I}$. Dann $(w(i))_{i\in I}$ ist ein Element des Produkts $\prod_{i\in I} A_i$. □

In diesem Abschnitt erklären wir eine weitere Folgerung des Auswahlaxioms, das sogenannte *Zornsche Lemma*, das viele Anwendungen in der gesamten Mathematik hat. Dazu benötigen wir ein paar Definitionen zu geordneten Mengen.

Definition 1.4.15 (kleinstes/größtes Element, minimales/maximales Element, untere/obere Schranke)**.** Sei $(X, \leq)$ eine partiell geordnete Menge und $A \subset X$ eine Teilmenge.

- Ein *kleinstes Element* von A ist ein Element $a \in A$ mit $a \leq b$ für alle $b \in A$.

- Ein *größtes Element* von A ist ein Element $a \in A$ mit $b \leq a$ für alle $b \in A$.
- Ein *minimales Element* von A ist ein Element $a \in A$ mit folgender Eigenschaft: Für alle $b \in A$ mit $b \leq a$ gilt $b = a$.
- Ein *maximales Element* von A ist ein Element $a \in A$ mit folgender Eigenschaft: Für alle $b \in A$ mit $a \leq b$ gilt $b = a$.
- Eine *untere Schranke* von A ist ein Element $x \in X$ mit $x \leq a$ für alle $a \in A$.
- Eine *obere Schranke* von A ist ein Element $x \in X$ mit $a \leq x$ für alle $a \in A$.

Bemerkung 1.4.16. Wenn A ein kleinstes Element a besitzt, dann ist auch a das eindeutige minimale Element von A sowie eine untere Schranke von A. Aber im Allgemeinen ist ein minimales Element kein kleinstes Element, und A kann mehrere minimale Elemente enthalten. Falls $(X, \leq)$ eine *total* geordnete Menge ist, gibt es aber keinen Unterschied zwischen „kleinstes Element“ und „minimales Element“.

Definition 1.4.17 (Kette)**.** Sei $(X, \leq)$ eine partiell geordnete Menge. Eine *Kette* in X ist eine Teilmenge $K \subset X$, so dass die Einschränkung von $\leq$ auf K eine totale Ordnung ist.

Definition 1.4.18 (Wohlordnung)**.** Eine *Wohlordnung* auf einer Menge X ist eine partielle Ordnung $\leq$ mit folgender Eigenschaft: Jede nichtleere Teilmenge $A \subset X$ besitzt ein kleinstes Element. Man sagt dann auch, dass $(X, \leq)$ eine wohlgeordnete Menge ist.

Beispiel 1.4.19. Das Wohlordnungsprinzip 1.2.19 ist genau die Aussage, dass $(\mathbb{N}, \leq)$ eine wohlgeordnete Menge ist. Wenn man $\mathbb{N}$ ein neues Element ∞ hinzufügt, so dass $n \leq \infty$ für alle $n \in \mathbb{N}$, dann ist $(\mathbb{N} \cup \{\infty\}, \leq)$ auch eine wohlgeordnete Menge.

Bemerkung 1.4.20. Eine Wohlordnung ist insbesondere eine totale Ordnung, da jede Teilmenge der Gestalt $\{x, y\}$ ein kleinstes Element besitzt. Jede Teilmenge einer wohlgeordneten Menge ist wieder wohlgeordnet.

Beispiel 1.4.21. Die total geordneten Mengen $(\mathbb{Z}, \leq)$, $(\mathbb{Q}, \leq)$ und $(\mathbb{R}, \leq)$ sind *nicht* wohlgeordnet, weil sie selbst kein kleinstes Element besitzen. Die total geordneten Mengen $(\mathbb{Q}_{\geq 0}, \leq)$ und $(\mathbb{R}_{\geq 0}, \leq)$ sind auch nicht wohlgeordnet. Zum Beispiel besitzen die Teilmengen $\mathbb{Q}_{>0}$ und $\mathbb{R}_{>0}$ kein kleinstes Element.

Satz 1.4.22 (das Zornsche Lemma)**.** *Sei $(X, \leq)$ eine partiell geordnete Menge, in der jede Kette eine obere Schranke besitzt. Dann besitzt $(X, \leq)$ ein maximales Element.*

Der Beweis ist ziemlich kompliziert, und es ist jetzt nicht wichtig, ihn zu verstehen.

**Beweis.* (Widerspruchsbeweis.) Wir nehmen an, dass X kein maximales Element besitzt. Sei $\mathcal{W} \subset \mathcal{P}(X)$ die Menge aller Teilmengen von X, die bzgl. $\leq$ wohlgeordnet sind. Insbesondere ist jedes $K \in \mathcal{W}$ eine Kette. Zu jedem $K \in \mathcal{W}$ sei $X_{\geq K}$ die Menge aller oberen Schranken von K, d.h., aller $x \in X$ mit $k \leq x$ für alle $k \in K$.

Nach Voraussetzung ist $X_{\geq K}$ nicht leer. Behauptung: $X_{\geq K} \setminus K$ ist nicht leer. Sonst wäre jede obere Schranke von K in K liegen. Da K total geordnet ist, würde es daraus folgen, dass jedes $x \in X_{\geq K}$ ein maximales Element von X ist, im Widerspruch zur Annahme.

Also besteht die Mengenfamilie $(X_{\geq K} \setminus K)_{K \in \mathcal{W}}$ aus nichtleeren Mengen. Nach dem Auswahlaxiom existiert eine Abbildung $s \colon \mathcal{W} \to X$ mit $s(K) \in X_{\geq K} \setminus K$ für alle $K \in \mathcal{W}$. Anders gesagt, s wählt zu jedem $K \in \mathcal{W}$ eine obere Schranke $s(K)$, die außerhalb von K liegt.

Eine Teilmenge $K \subset X$ nennen wir *s-induktiv*, falls $K \in \mathcal{W}$ und $x = s(K_{<x})$ für alle $x \in K$, wobei

$$K_{<x} := \{y \in K \mid y \leq x \text{ und } y \neq x\}.$$

Zum Beispiel, die leere Teilmenge ist s-induktiv, $\{s(\varnothing)\}$ ist s-induktiv, und ist K s-induktiv, so ist $K \cup \{s(K)\}$ s-induktiv. Sei $V \subset X$ die Vereinigung aller s-induktiven Teilmengen von X. Wir zeigen im Folgenden, dass V s-induktiv ist. Dann ist $V \cup \{s(V)\}$ eine s-induktive Menge, die keine Teilmenge von V ist, im Widerspruch zur Definition von V. Damit wird der Satz bewiesen.

Ist A eine Teilmenge von $K \in \mathcal{W}$, so schreibt man $A \prec K$, falls A nach unten abgeschlossen ist, das heißt: Für alle $a \in A$ und $k \in K$, ist $k \leq a$, so ist $k \in A$.

Behauptung. Seien $K, L \subset X$ s-induktive Teilmengen. Dann gilt $K \prec L$ oder $L \prec K$.

Mithilfe dieser Behauptung können wir beweisen, dass V s-induktiv ist:

- *V ist wohlgeordnet.* Sei $A \subset V$ eine nichtleere Teilmenge. Nach Definition von V existiert eine s-induktive Menge K, so dass $A \cap K \neq \varnothing$. Da K wohlgeordnet ist, existiert ein kleinstes Element $a \in A \cap K$. Wir beweisen, dass a ein kleinstes Element von A ist. Sei $b \in A$ ein beliebiges Element. Liegt b auch in K, so ist $a \leq b$. Andernfalls, sei L eine s-induktive Menge mit $b \in L \setminus K$. Nach der Behauptung gilt dann $K \prec L$, und daher $a \leq b$.

- *V ist s-induktiv.* Sei $x \in V$ und sei K eine s-induktive Teilmenge mit $x \in K$. Aus der Behauptung folgt $V_{<x} = K_{<x}$. Also $s(V_{<x}) = s(K_{<x}) = x$.

Beweis der Behauptung. Nach Definition von $\prec$ ist eine beliebiege Vereinigung von $\prec K$ Teilmengen wieder $\prec K$. Sei A die Vereinigung aller Mengen, die $\prec K$ *und* $\prec L$ sind. Dann ist A wieder $\prec K$ und $\prec L$. Zu zeigen ist, dass $A = K$ oder $A = L$. Angenommen, $A \neq K$ und $A \neq L$. Da K wohlgeordnet ist, existiert ein kleinstes Element $k \in K \setminus A$; insbesondere gilt $K_{<k} \subset A$. Da A in K nach unten abgeschlossen ist, gelten auch $A \cup \{k\} \prec K$ und $A \subset K_{<k}$, also $A = K_{<k}$. Insbesondere ist $s(A) = s(K_{<k}) = k$, und daher $A \cup \{s(A)\} \prec K$. Wenn wir in diesem Argument K durch L ersetzen, erhalten wir $A \cup \{s(A)\} \prec L$. Das steht aber im Widerspruch zur Definition von A, da $A \cup \{s(A)\}$ keine Teilmenge von A ist. □

Korollar 1.4.23 (Wohlordnungssatz)**.** *Sei X eine beliebige Menge. Dann existiert auf X eine Wohlordnung.*

**Beweis.* Dieser Beweis ist eine typische Anwendung des Zornschen Lemmas. Wir betrachten folgende Menge:

$$W = \{(Y, R) \mid Y \subset X \text{ und } R \subset Y \times Y \text{ ist eine Wohlordnung auf } Y\} \subset \mathcal{P}(X) \times \mathcal{P}(X \times X),$$

und definieren auf W die folgende Relation $\prec$:

$$(Y, R) \prec (Y', R') \iff Y \subset Y',\ R = R' \cap (Y \times Y) \\ \text{und } Y \text{ ist in } Y' \text{ bzgl. } R' \text{ nach unten abgeschlossen.}$$

Man kann leicht nachprüfen, dass $\prec$ eine partielle Ordnung auf W ist. Als nächstes überprüfen wir, dass $(W, \prec)$ der Bedingung des Zornschen Lemmas genügt, d.h., dass jede Kette in W eine obere Schranke besitzt. Sei $K \subset W$ eine Kette. Wir setzen

$$Y_\infty = \bigcup_{(Y,R) \in K} Y \quad \text{und} \quad R_\infty = \bigcup_{(Y,R) \in K} R.$$

Dann ist das Paar (Y_∞, R_∞) ein Element von W, und es ist eine obere Schranke von K:

- *R_∞ ist eine Wohlordnung auf Y_∞.* Sei $A \subset Y_\infty$ eine nichtleere Teilmenge. Es gibt $(Y, R) \in K$, so dass $A \cap Y \neq \varnothing$. Sei a das kleinste Element von $A \cap Y$ bzgl. R. Dann ist a eigentlich das kleinste Element von A bzgl. R_∞. Denn sei $b \in A$, und sei $(Y', R') \in K$ mit $b \in Y'$. Falls b auch in Y liegt, gilt $aR_\infty b$. Andernfalls, da K eine Kette ist, gilt $(Y, R) \prec (Y', R')$ und $b \in Y' \setminus Y$. Da Y in Y' nach unten abgeschlossen ist, gilt dann auch $aR_\infty b$.

- *Für jedes* $(Y, R) \in K$ *gilt* $(Y, R) \prec (Y_\infty, R_\infty)$. Die Inklusion $R \subset R_\infty \cap (Y \times Y)$ ist klar. Zur anderen Inklusion, sei $(y, z) \in R_\infty \cap (Y \times Y)$. Es existiert dann $(Y', R') \in K$ mit $yR'z$. Da K eine Kette ist, gilt $R' \subset R$ oder $R = R' \cap (Y \times Y)$. In beiden Fällen folgt yRz. Es bleibt zu zeigen, dass Y in Y_∞ nach unten abgeschlossen ist. Sei $y \in Y$ und $z \in Y_\infty$ mit $zR_\infty y$, und sei $(Y', R') \in K$ mit $z \in Y'$. Falls $(Y', R') \prec (Y, R)$ ist $z \in Y$. Andernfalls, Y ist in Y' enthalten und nach unten abgeschlossen, und somit ist auch $z \in Y$.

Nach dem Zornschen Lemma besitzt W ein maximales Element (Y, R). Es bleibt zu zeigen, dass $Y = X$; dann ist R eine Wohlordnung auf ganz X, wie gewünscht. Sei also $x \in X$. Auf $Y \cup \{x\}$ kann man folgende Relation R' definieren:

$$(y, z) \in R' \iff (y, z) \in R \text{ oder } (z \notin Y \text{ und } z = x).$$

Dann ist R' eine Wohlordnung auf $Y \cup \{x\}$, und $(Y, R) \prec (Y \cup \{x\}, R')$. Da (Y, R) maximal bzgl. $\prec$ ist, erhalten wir $Y = Y \cup \{x\}$, d.h., $x \in Y$. □

Kapitel 2

Gruppen und Körper

2.1 Gruppen

Definition 2.1.1 (Gruppe)**.** Eine *Gruppe* ist ein Paar $(G, \cdot)$, bestehend aus einer Menge G und einer Abbildung

$$\cdot : G \times G \to G, \quad (g, h) \mapsto g \cdot h$$

(die *Verknüpfung* der Gruppe), mit folgenden Eigenschaften:

(i) Die Verknüpfung ist *assoziativ*, d.h., für alle $g, h, k \in G$ gilt

$$g \cdot (h \cdot k) = (g \cdot h) \cdot k.$$

(ii) Es existiert ein *neutrales Element* bzgl. $\cdot$, d.h., ein Element $e \in G$ so dass

$$e \cdot g = g \quad \text{und} \quad g \cdot e = g$$

für alle $g \in G$.

(iii) Jedes $g \in G$ besitzt ein *inverses Element*, d.h., ein Element $h \in G$ so dass

$$g \cdot h \quad \text{und} \quad h \cdot g$$

neutrale Elemente sind.

Beispiel 2.1.2. $(\mathbb{Z}, +)$ ist eine Gruppe: Die Addition von ganzen Zahlen ist assoziativ, $0 \in \mathbb{Z}$ ist ein neutrales Element bzgl. $+$, und $-n$ ist ein inverses Element von n. In ähnlicher Weise, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ und $(\mathbb{C}, +)$ sind Gruppen.

Andererseits ist $(\mathbb{N}, +)$ keine Gruppe. Die Bedingungen (i) und (ii) der Definition 2.1.1 sind erfüllt, aber (iii) nicht: Eine positive natürliche Zahl hat kein inverses Element bzgl. $+$ in $\mathbb{N}$.

Proposition 2.1.3 (Eindeutigkeit von neutralen und inversen Elementen)**.** *Sei $(G, \cdot)$ eine Gruppe.*

(i) *Sind $e, e' \in G$ neutrale Elemente bzgl. $\cdot$, so gilt $e = e'$.*

(ii) *Sei $g \in G$. Sind $h, h' \in G$ inverse Elemente von g, so gilt $h = h'$.*

Beweis. Zu (i). Seien $e, e' \in G$ neutrale Elemente. Dann

$$e = e \cdot e' = e',$$

wobei die erste Gleichung gilt, weil e' ein neutrales Element ist, und die zweite Gleichung gilt, weil e ein neutrales Element ist.

Zu (ii). Seien $h, h' \in G$ inverse Elemente von g, und sei $e \in G$ das eindeutige neutrale Element (nach (i)). Dann

$$\begin{aligned} h &= h \cdot e && (e \text{ neutral}) \\ &= h \cdot (g \cdot h') && (h' \text{ invers zu } g) \\ &= (h \cdot g) \cdot h' && (\cdot \text{ assoziativ}) \\ &= e \cdot h' && (h \text{ invers zu } g) \\ &= h' && (e \text{ neutral}). \end{aligned}$$ □

Notation 2.1.4. Wegen Proposition 2.1.3 darf man „*das* neutrale Element“ und „*das* inverse Element von g“ sagen, da die entsprechenden mathematischen Objekte eindeutig sind. Das neutrale Element einer Gruppe wird mit e oder 1 bezeichnet, und das inverse Element von einem Element g wird mit g^{-1} bezeichnet. Das Symbol $\cdot$ schreibt man normalerweise gar nicht, d.h., man schreibt eher gh statt $g \cdot h$.

Bemerkung 2.1.5. In einer Gruppe gilt $(gh)^{-1} = h^{-1}g^{-1}$ (Reihenfolge beachten!). Nach der Eindeutigkeit des inversen Element genügt es zu zeigen, dass $h^{-1}g^{-1}$ ein inverses Element von gh ist. Tatsächlich gilt:

$$(h^{-1}g^{-1})(gh) = h^{-1}((g^{-1}g)h) = h^{-1}(eh) = h^{-1}h = e,$$

und ebenso $(gh)(h^{-1}g^{-1}) = e$.

Notation 2.1.6. Wegen der Assoziativität der Verknupfung in einer Gruppe $(G, \cdot)$, darf man unmissverständlich $g \cdot h \cdot k$ schreiben: Es macht keinen Unterschied, ob wir die Klammern um $g \cdot h$ oder um $h \cdot k$ setzen. Allgemeiner, für jede endliche Liste von Elementen $g_1, g_2, \ldots, g_n \in G$, darf man das Produkt $g_1 \cdot g_2 \cdot \ldots \cdot g_n \in G$ ohne Klammern schreiben. Man schreibt auch

$$\prod_{i=1}^{n} g_i := g_1 \cdot g_2 \cdot \ldots \cdot g_n.$$

Wenn $n = 0$ verstehen wir das „leere Produkt“ als das neutrale Element $e \in G$. Wenn alle Elemente g_i dasselbe Element g sind, schreibt man einfach g^n für das n-fache Produkt von g mit sich selbst. Man setzt auch $g^0 := e$ und $g^{-n} := (g^n)^{-1}$. So wird die Potenz g^n für alle ganzen Zahlen $n \in \mathbb{Z}$ definiert.

Proposition 2.1.7 (Eigenschaften der Potenzen)**.** *Sei $(G, \cdot)$ eine Gruppe und sei $g \in G$.*

(i) *Für alle $m, n \in \mathbb{Z}$ gilt $g^m \cdot g^n = g^{m+n}$.*

(ii) *Für alle $m, n \in \mathbb{Z}$ gilt $(g^n)^m = g^{mn}$.*

Beweis. Zu (i). Wir betrachten zunächst den Fall $m \in \mathbb{N}$, in dem wir Induktion über m verwenden.

- *Induktionsanfang.* Wenn $m = 0$, gilt $g^0 \cdot g^n = e \cdot g^n = g^n = g^{0+n}$.
- *Induktionsschritt.* Es gilt

  $$g^{m+1} \cdot g^n = g \cdot g^m \cdot g^n = g \cdot g^{m+n} = g^{m+1+n},$$

  wobei die zweite Gleichung aus der Induktionsvoraussetzung folgt.

Damit ist (i) bewiesen für alle $m \geq 0$. Der Fall $n \geq 0$ wird mit einem ähnlichen Argument erledigt. Wenn $m < 0$, dann gilt

$$g^m \cdot g^n = (g^{-n} \cdot g^{-m})^{-1} = (g^{-(m+n)})^{-1} = g^{m+n}.$$

Dabei haben wir die Formel $(gh)^{-1} = h^{-1}g^{-1}$ in der ersten Gleichung und den Fall $n \geq 0$ (mit $-m$ anstelle von n) in der zweiten Gleichung verwendet.

Zu (ii). Wir beweisen zunächst den Fall $m \in \mathbb{N}$ durch Induktion über m.

- *Induktionsanfang.* Wenn $m = 0$, gilt $(g^n)^0 = e = g^0 = g^{0n}$.
- *Induktionsschritt.* Es gilt

$$(g^n)^{m+1} = g^n \cdot (g^n)^m = g^n \cdot g^{mn} = g^{n+mn} = g^{(m+1)n},$$

wobei die zweite Gleichung aus der Induktionsvoraussetzung folgt, und die dritte aus (i).

Damit ist (ii) bewiesen für alle $m \geq 0$. Wenn $m < 0$ ist der Beweis mit der folgenden Berechnung abgeschlossen:

$$(g^n)^m = ((g^n)^{-m})^{-1} = (g^{-mn})^{-1} = g^{mn}. \qquad \square$$

Bemerkung 2.1.8. Sei $(G, \cdot)$ eine Gruppe, $g, h \in G$ und $n \in \mathbb{Z}$. Im Allgemeinen gilt die Gleichung $(gh)^n = g^n h^n$ nur für $n = 0$ und $n = 1$. Zum Beispiel, $(gh)^2 = ghgh$ muss nicht gleich $g^2h^2 = gghh$ sein. Diese Gleichung gilt aber für alle $n \in \mathbb{Z}$, wenn die Gruppe abelsch ist (siehe Definition 2.1.9).

Definition 2.1.9 (abelsche Gruppe)**.** Eine Gruppe $(G, \cdot)$ heißt *abelsch*, falls ihre Verknüpfung *kommutativ* ist, d.h., für alle $g, h \in G$ gilt

$$g \cdot h = h \cdot g.$$

Beispiel 2.1.10. Die Gruppe $(\mathbb{Z}, +)$ ist abelsch, da $n + m = m + n$ für alle ganzen Zahlen n, m. In ähnlicher Weise, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ und $(\mathbb{C}, +)$ sind abelsche Gruppen.

Notation 2.1.11. Bei abelschen Gruppen verwendet man häufig die *additive Notation* statt der *multiplikativen* Notation (Notation 2.1.4): d.h., man schreibt $+$ für die Verknüpfung, 0 für das neutrale Element und $-a$ für das Inverse von a. Man schreibt auch $a-b$ als Abkürzung von $a + (-b)$. Ist $a_1, a_2, \ldots, a_n$ eine Liste von Elementen, so schreibt man

$$\sum_{i=1}^{n} a_i := a_1 + a_2 + \cdots + a_n.$$

Wenn $n = 0$ verstehen wir die „leere Summe“ als das neutrale Element 0. Wenn alle a_i gleich a sind, schreibt man einfach $n \cdot a$ oder na für diese Summe. Man setzt auch $0 \cdot a := 0$ und $(-n) \cdot a := -(n \cdot a)$. So wird $n \cdot a$ für alle ganzen Zahlen $n \in \mathbb{Z}$ definiert.

Notation 2.1.12. Oft unterdrückt man die Verknüpfung in der Notation für eine Gruppe $(G, \cdot)$. Das heißt, man sagt üblicherweise „Sei G eine Gruppe“ und nicht „Sei $(G, \cdot)$ eine Gruppe“. In diesem Fall wird die Verknüpfung standardmäßig mit $\cdot$ bezeichnet (oder mit $+$ im Fall einer abelschen Gruppe).

Bemerkung 2.1.13. Wenn wir in der Definition einer Gruppe (Definition 2.1.1) auf das dritte Axiom verzichten, erhalten wir den Begriff des *Monoids*. In einem Monoid ist das neutrale Element immer noch eindeutig bestimmt: Der Beweis von Proposition 2.1.3(i) benötigt keine inverse Elemente. Zum Beispiel ist $(\mathbb{N}, +)$ ein (abelsches) Monoid aber keine Gruppe.

2.2 Beispiele von Gruppen

2.2.1 Die ganzen Zahlen

Die natürlichen Zahlen $\mathbb{N}$ bilden *keine* Gruppe bezüglich Addition, denn die positiven natürlichen Zahlen besitzen keine inverse Elemente. Um $(\mathbb{N}, +)$ zu einer Gruppe zu erweitern, brauchen wir die negativen Zahlen hinzuzufügen. Dann erhalten wir die ganzen Zahlen $\mathbb{Z}$,

und $(\mathbb{Z}, +)$ ist eine abelsche Gruppe. In diesem Abschnitt erklären wir, wie $\mathbb{Z}$ aus $\mathbb{N}$ eigentlich konstruiert werden kann.

Die Idee ist, dass jede ganze Zahl als Differenz zweier natürlichen Zahlen n, m dargestellt werden kann. Auf diese Weise bestimmt jedes Paar $(n, m) \in \mathbb{N} \times \mathbb{N}$ eine ganze Zahl, nämlich $n - m$. Aber diese Darstellung ist nicht eindeutig: Es gilt $n - m = n' - m'$ genau dann, wenn $n + m' = n' + m$. Deswegen führen wir folgende Relation $\sim$ auf $\mathbb{N} \times \mathbb{N}$ ein:

$$(n, m) \sim (n', m') \iff n + m' = n' + m.$$

Man kann leicht nachprüfen, dass $\sim$ eine Äquivalenzrelation auf $\mathbb{N} \times \mathbb{N}$ ist. Als Beispiel überprüfen wir die Transitivität: Falls $(n, m) \sim (n', m')$ und $(n', m') \sim (n'', m'')$, dann gilt

$$\begin{aligned}
(n + m'') + m' &= (n + m') + m'' && \text{(Assoziativität und Kommutativität von } +) \\
&= (n' + m) + m'' && \text{(da } (n, m) \sim (n', m')) \\
&= (n' + m'') + m && \text{(Assoziativität und Kommutativität von } +) \\
&= (n'' + m') + m && \text{(da } (n', m') \sim (n'', m'')) \\
&= (n'' + m) + m', && \text{(Assoziativität und Kommutativität von } +)
\end{aligned}$$

und daher $n + m'' = n'' + m$, d.h., $(n, m) \sim (n'', m'')$ (hierbei wurde benutzt, dass für naturliche Zahlen n, m, p gilt: $n + p = m + p \Rightarrow n = m$).

Wir setzen nun

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\sim.$$

Mit dieser Definition ist $\mathbb{N}$ keine Teilemenge von $\mathbb{Z}$, was eher unangenehm ist. In der Praxis möchten wir auf jeden Fall $\mathbb{N}$ als Teilmenge von $\mathbb{Z}$ auffassen. Hierzu betrachten wir die Abbildung

$$i \colon \mathbb{N} \to \mathbb{Z}, \quad n \mapsto [(n, 0)].$$

Die Abbildung i ist *injektiv*. Denn seien $n, n' \in \mathbb{N}$ mit $[(n, 0)] = [(n', 0)]$; dann gilt $(n, 0) \sim (n', 0)$, d.h. $n + 0 = n' + 0$, also $n = n'$. Insbesondere induziert i eine Bijektion zwischen $\mathbb{N}$ und seinem Bild $i(\mathbb{N}) \subset \mathbb{Z}$, und so können wir $\mathbb{N}$ mit einer Teilmenge von $\mathbb{Z}$ identifizieren. Dementsprechend führen wir folgende Notation ein:

Notation 2.2.1. Sei $n \in \mathbb{N}$. Die Äquivalenzklasse $[(n, 0)] \in \mathbb{Z}$ bezeichnen wir auch mit n. Die Äquivalenzklasse $[(0, n)] \in \mathbb{Z}$ bezeichnen wir mit $-n$.

Wir möchten jetzt die arithmetischen Operationen $+$ und $\cdot$ von $\mathbb{N}$ auf $\mathbb{Z}$ fortsetzen, sowie die totale Ordnungsrelation $\leq$. Die Definitionen sind klar, wenn man sich die Äquivalenzklasse $[(n, m)]$ als die Differenz $n - m$ vorstellt:

$$\begin{aligned}
[(n_1, m_1)] + [(n_2, m_2)] &= [(n_1 + n_2, m_1 + m_2)], \\
[(n_1, m_1)] \cdot [(n_2, m_2)] &= [(n_1 n_2 + m_1 m_2, n_1 m_2 + n_2 m_1)], \\
[(n_1, m_1)] \leq [(n_2, m_2)] &\iff n_1 + m_2 \leq n_2 + m_1.
\end{aligned}$$

Da es Äquivalenzklassen auf der linken Seite dieser Definitionen gibt, muss man hier nachprüfen, dass alles wohldefiniert ist. Zum Beispiel, um zu zeigen, dass $+$ auf $\mathbb{Z}$ wohldefiniert ist, ist folgendes zu beweisen:

$$(n_1, m_1) \sim (n_1', m_1') \text{ und } (n_2, m_2) \sim (n_2', m_2') \implies (n_1 + n_2, m_1 + m_2) \sim (n_1' + n_2', m_1' + m_2').$$

Dies folgt unmittelbar aus der Definition von $\sim$.

Bemerkung 2.2.2. Diese Konstruktion von $\mathbb{Z}$ aus $\mathbb{N}$ ist ein Sonderfall einer allgemeineren Konstruktion, die eine Gruppe aus irgendeinem Monoid liefert.

2.2.2 Symmetrische Gruppen

Definition 2.2.3 (Permutation)**.** Sei X eine Menge. Eine *Permutation* von X ist eine bijektive Abbildung von X nach X. Die Menge aller Permutationen von X wird mit S_X bezeichnet.

Proposition 2.2.4. *Sei X eine Menge. Dann ist $(S_X, \circ)$ eine Gruppe. Falls X mindestens drei verschiedene Elemente besitzt, ist diese Gruppe nicht abelsch.*

Beweis. Komposition von Abbildungen ist assoziativ und die Identität id_X ist ein neutrales Element (Proposition 1.3.13). Jedes Element von S_X hat ein inverses Element nach Satz 1.3.23. Also ist $(S_X, \circ)$ eine Gruppe.

Zu je zwei Elementen $a, b \in X$ gibt es eine bijektive Abbildung $\tau_{a,b} \in S_X$, die wie folgt definiert wird:

$$\tau_{a,b}(x) = \begin{cases} b, & \text{falls } x = a, \\ a, & \text{falls } x = b, \\ x, & \text{andernfalls.} \end{cases}$$

Sind $a, b, c \in X$ drei verschiedene Elemente, so gilt $\tau_{b,c} \circ \tau_{a,b} \neq \tau_{a,b} \circ \tau_{b,c}$, denn:

$$\tau_{b,c}(\tau_{a,b}(a)) = c \quad \text{aber} \quad \tau_{a,b}(\tau_{b,c}(a)) = b.$$

Also ist $(S_X, \circ)$ nicht abelsch. □

Definition 2.2.5 (symmetrische Gruppe)**.** Die Gruppe $(S_X, \circ)$ heißt die *symmetrische Gruppe* von X. Falls $X = \{1, \ldots, n\}$ mit $n \in \mathbb{N}$ schreibt man S_n (auch Σ_n, $\mathfrak{S}_n$) statt S_X. Die Gruppe $(S_n, \circ)$ heißt die *symmetrische Gruppe vom Grad n*.

Beispiel 2.2.6. Die Gruppe S_2 hat genau zwei Elemente: die Identität id und die Abbildung $\tau\colon \{1, 2\} \to \{1, 2\}$, die 1 und 2 austauscht. Es gilt $\tau \circ \tau = \mathrm{id}$.

Bemerkung 2.2.7. Durch Induktion kann man leicht nachprüfen, dass die Mächtigkeit von S_n gleich $n!$ ist (wobei $0! = 1$).

Bemerkung 2.2.8 (indizierte Summen in einer abelschen Gruppe)**.** Sei A eine abelsche Gruppe (allgemeiner, ein abelsches Monoid) und $a_1, \ldots, a_n \in A$. Nach der Kommutativität von $+$ ist die Summe $\sum_{k=1}^{n} a_i$ *unabhängig* von der Reihenfolge dieser Elemente. Genauer heißt das folgendes: Für jede Permutation $\sigma \in S_n$ der Indexmenge $\{1, \ldots, n\}$ gilt:

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} a_{\sigma(k)}.$$

Sei nun I eine endliche Menge und $(a_i)_{i \in I}$ eine Familie von Elementen von A mit Indexmenge I. Dann kann man die Summe $\sum_{i \in I} a_i \in A$ wie folgt definieren. Man wählt eine Bijektion $f\colon \{1, \ldots, n\} \to I$ und setzt

$$\sum_{i \in I} a_i := \sum_{k=1}^{n} a_{f(k)}.$$

Die rechte Seite ist unabhängig von der Wahl von f, denn: Seien $f, g\colon \{1, \ldots, n\} \to I$ zwei Bijektionen. Dann ist $\sigma = f^{-1} \circ g$ eine Permutation von $\{1, \ldots, n\}$, und es gilt

$$\sum_{k=1}^{n} a_{f(k)} = \sum_{k=1}^{n} a_{f(\sigma(k))} = \sum_{k=1}^{n} a_{g(k)}.$$

2.3 Körper

Der Begriff des Körpers ist eine Abstraktion der arithmetischen Struktur der rationalen, reellen und komplexen Zahlen.

Definition 2.3.1 (Körper)**.** Ein *Körper* ist ein Tripel $(K, +, \cdot)$, bestehend aus einer Menge K und Abbildungen

$$\begin{aligned} +\colon K \times K \to K, &\quad (x, y) \mapsto x + y, \\ \cdot\colon K \times K \to K, &\quad (x, y) \mapsto x \cdot y, \end{aligned}$$

die als *Addition* und *Multiplikation* bezeichnet werden, mit folgenden Eigenschaften:

(i) $(K, +)$ ist eine abelsche Gruppe. Das heißt, die Verknüpfung $+$ ist assoziativ und kommutativ, sie besitzt ein neutrales Element 0, und jedes Element von K besitzt ein inverses Element bzgl. $+$.

(ii) Die Verknüpfung $\cdot$ ist assoziativ, kommutativ, und besitzt ein neutrales Element 1.

(iii) Jedes Element von $K \setminus \{0\}$ besitzt ein inverses Element bzgl. $\cdot$.

(iv) Es gilt $0 \neq 1$.

(v) Es gilt das *Distributivgesetz*: Für alle $x, y, z \in K$,

$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

Beispiel 2.3.2. $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ und $(\mathbb{C}, +, \cdot)$ sind Körper. Siehe Abschnitt 2.4 für die formalen Konstruktionen von $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$.

Bemerkung 2.3.3. Da die Verknüpfung $\cdot$ kommutativ ist, gilt auch in einem Körper das umgekehrte Distributivgesetz

$$(x + y) \cdot z = x \cdot z + y \cdot z.$$

Durch Induktion kann man außerdem folgendes verallgemeinertes Distributivgesetz beweisen:

$$x \cdot \left(\sum_{i=1}^{n} y_i \right) = \sum_{i=1}^{n} x \cdot y_i.$$

Dies gilt auch für $n = 0$ nach Proposition 2.3.8(i).

Bemerkung 2.3.4. Nach Definition gilt in einem Körper $0 \neq 1$. Ein Körper enthält deshalb mindestens zwei Elemente. Eigentlich existiert ein Körper mit genau zwei Elementen, siehe Abschnitt 2.4.4.

Bemerkung 2.3.5. Wenn wir in der Definition 2.3.1 auf Axiome (iii) und (iv) verzichten, erhalten wir den Begriff des *kommutativen Ringes*. Zum Beispiel ist $(\mathbb{Z}, +, \cdot)$ ein kommutativer Ring, aber kein Körper. Kommutative Ringe spielen eine wichtige Rolle in mehreren Bereichen der Mathematik, und sie werden in der späteren Vorlesung *Kommutative Algebra* untersucht.

Notation 2.3.6. In einem Körper verwenden wir ausschließlich die additive Notation 2.1.11 für die erste Verknüpfung und die multiplikative Notation 2.1.4 für die zweite Verknüpfung. Insbesondere schreiben wir $-x$ für das inverse Element von x bzgl. $+$ und (falls $x \neq 0$) x^{-1} für das inverse Element bzgl. $\cdot$. Außerdem verwenden wir die Bruchnotation

$$\frac{x}{y} := x \cdot y^{-1},$$

falls $y \neq 0$.

Notation 2.3.7. Wie bei Gruppen unterdrückt man oft die Verknüpfungen in der Notation für einen Körper $(K, +, \cdot)$. Das heißt, man sagt üblicherweise „Sei K ein Körper“ und nicht „Sei $(K, +, \cdot)$ ein Körper“.

Proposition 2.3.8 (Rechnen in Körpern)**.** *Sei $(K, +, \cdot)$ ein Körper.*

(i) *Für alle $x \in K$ gilt $0 \cdot x = 0$.*

(ii) *Für alle $x \in K$ gilt $(-1) \cdot x = -x$. Insbesondere ist $(-1) \cdot (-1) = 1$.*

(iii) (Nullteilerfreiheit) *Sind $x, y \in K \setminus \{0\}$, so ist $x \cdot y \in K \setminus \{0\}$.*

Beweis. Zu (i). Nach dem Distributivgesetz gilt $(0 + 0) \cdot x = 0 \cdot x + 0 \cdot x$. Da $0 + 0 = 0$, erhalten wir

$$0 \cdot x = 0 \cdot x + 0 \cdot x.$$

Da $(K, +)$ eine Gruppe ist, dürfen wir $0 \cdot x$ von beiden Seiten subtrahieren. Also $0 = 0 \cdot x$.

Zu (ii). Nach der Eindeutigkeit des inversen Element (Proposition 2.1.3) genügt es zu zeigen, dass $(-1) \cdot x$ ein inverses Element von x bezüglich $+$ ist, d.h., dass $(-1) \cdot x + x = 0$. Wir berechnen:

$$\begin{aligned}
(-1) \cdot x + x &= (-1) \cdot x + 1 \cdot x && \text{(1 neutral)} \\
&= ((-1) + 1) \cdot x && \text{(Distributivität)} \\
&= 0 \cdot x \\
&= 0. && \text{(nach (i))}
\end{aligned}$$

Zu (iii). Wir beweisen die Kontraposition. Seien $x, y \in K$ mit $x \cdot y = 0$. Zu zeigen ist, dass $x = 0$ oder $y = 0$. Falls $y \neq 0$, existiert ein inverses Element y^{-1}, so dass $y \cdot y^{-1} = 1$. Also gilt

$$\begin{aligned}
x &= x \cdot 1 && \text{(1 neutral)} \\
&= x \cdot (y \cdot y^{-1}) \\
&= (x \cdot y) \cdot y^{-1} && \text{(Assoziativität)} \\
&= 0 \cdot y^{-1} && \text{(da } x \cdot y = 0\text{)} \\
&= 0, && \text{(nach (i))}
\end{aligned}$$

wie gewünscht. □

Korollar 2.3.9. *Ist $(K, +, \cdot)$ ein Körper, so ist $(K \setminus \{0\}, \cdot)$ eine abelsche Gruppe.*

Beweis. Zuerst muss man beobachten, dass die Multiplikation $(K \setminus \{0\}) \times (K \setminus \{0\})$ auf $K \setminus \{0\}$ abbildet. Dies folgt aus Proposition 2.3.8(iii). Nach Definition eines Körpers ist die Verknüpfung $\cdot$ assoziativ und kommutativ, und sie besitzt ein neutrales Element $1 \in K \setminus \{0\}$. Es bleibt zu zeigen, dass jedes $x \in K \setminus \{0\}$ ein inverses Element in $K \setminus \{0\}$ besitzt. Nach Definition eines Körpers gibt es ein inverses Element $x^{-1} \in K$. Da $x^{-1} \cdot x = 1 \neq 0$, folgt aus Proposition 2.3.8(i), dass $x^{-1} \neq 0$. □

Notation 2.3.10. Sei $(K, +, \cdot)$ ein Körper. Die Menge $K \setminus \{0\}$ wird oft mit $K^\times$ oder K^* bezeichnet, besonders wenn man diese Menge als abelsche Gruppe bzgl. $\cdot$ betrachtet.

Bemerkung 2.3.11. Wenn wir in der Definition 2.3.1 die Axiome (ii) und (iii) durch das einfachere Axiom „$(K, \cdot)$ ist eine abelsche Gruppe“ ersetzen, dann erhalten wir einen Begriff, der keine Beispiele besitzt. Denn in einem solchen K würde 0 ein Inverses 0^{-1} haben, so dass $0 \cdot 0^{-1} = 1$. Auf der anderen Seite ist $0 \cdot 0^{-1} = 0$ nach Proposition 2.3.8(i), und somit $0 = 1$, im Widerspruch zum Axiom (iv).

Zur Erinnerung (Notation 2.1.11) haben wir in der additiven Gruppe $(K, +)$ eines Körpers das ganzzahlige Vielfache $n \cdot x$ für alle $n \in \mathbb{Z}$ und $x \in K$ definiert. Insbesondere gibt es eine Abbildung

$$\begin{aligned} &\mathbb{Z} \to K, \\ &n \mapsto n \cdot 1, \end{aligned}$$

wobei $1 \in K$ das neutrale Element bzgl. $\cdot$ ist. Man schreibt oft einfach n für das Element $n \cdot 1$ von K. Wenn $K = \mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$, ist diese Abbildung injektiv, und zwar die gewöhnliche Inklusion von $\mathbb{Z}$ in diesen Körpern. Im Abschnitt 2.4.4 geben wir Beispiele von Körpern, in denen $n \cdot 1 = 0$ mit einem $n \neq 0$. Im Vorgriff auf diese Beispiele führen wir folgende Definition ein:

Definition 2.3.12 (Charakteristik eines Körpers)**.** Sei K ein Körper. Die *Charakteristik* von K, $\mathrm{char}(K)$, ist die wie folgt definierte natürliche Zahl:

- Falls $n \cdot 1 \neq 0$ für alle $n \in \mathbb{N} \setminus \{0\}$, ist $\mathrm{char}(K) = 0$.
- Andernfalls ist $\mathrm{char}(K)$ das kleinste $n \in \mathbb{N} \setminus \{0\}$ mit $n \cdot 1 = 0$.

In einem Körper K der Charakteristik $n \neq 0$ gilt $n \cdot x = 0$ für alle $x \in K$, denn:

$$n \cdot x = x + \cdots + x = (1 + \cdots + 1) \cdot x = (n \cdot 1) \cdot x = 0 \cdot x = 0.$$

Proposition 2.3.13. *Sei K ein Körper. Dann ist die Charakteristik von K entweder 0 oder eine Primzahl.*

Beweis. Sei $n = \mathrm{char}(K)$. Wir nehmen an, dass $n \neq 0$. Da $1 \neq 0$ in K, ist auch $n \neq 1$. Es sei $n = rs$ mit natürlichen Zahlen $r, s \in \mathbb{N}$. Zu zeigen ist, dass $r = n$ oder $s = n$. Es gilt

$$0 = n \cdot 1 = r \cdot (s \cdot 1) = (r \cdot 1) \cdot (s \cdot 1),$$

wobei die letzte Gleichung aus dem Distributivgesetz folgt. Aus der Nullteilerfreiheit von K (Proposition 2.3.8(iii)) folgt $r \cdot 1 = 0$ oder $s \cdot 1 = 0$. Aber n ist nach Definition die kleinste natürliche Zahl mit $n \cdot 1 = 0$. Es muss also $r = n$ oder $s = n$ sein, wie gewünscht. □

Beispiel 2.3.14. In einem Körper K der Charakteristik 2 gilt $(a + b)^2 = a^2 + b^2$ für alle $a, b \in K$. Denn in einem beliebigen Körper gilt $(a+b)^2 = a^2 + 2ab + b^2$, wobei $2ab = ab + ab$ (dies folgt aus der Distributivität von $\cdot$ über $+$ und der Kommutativität von $\cdot$), aber $2ab = 0$ in K.

Allgemeiner: Wenn die Charakteristik von K eine Primzahl p ist, folgt aus dem binomischen Lehrsatz, dass $(a + b)^p = a^p + b^p$ für alle $a, b \in K$.

2.4 Beispiele von Körpern

2.4.1 Die rationalen Zahlen

Die ganzen Zahlen $\mathbb{Z}$ bilden *keinen* Körper bezüglich Addition und Multiplikation, denn die ganzen Zahlen außer ± 1 besitzen keine inverse Elemente bezüglich Multiplikation. Um $(\mathbb{Z}, +, \cdot)$ zu einem Körper zu erweitern, brauchen wir Brüche ganzer Zahlen hinzuzufügen. Dann erhalten wir die rationalen Zahlen $\mathbb{Q}$, und $(\mathbb{Q}, +, \cdot)$ ist ein Körper. In diesem Abschnitt erklären wir, wie $\mathbb{Q}$ aus $\mathbb{Z}$ eigentlich konstruiert werden kann.

Die Konstruktion von $\mathbb{Q}$ aus $\mathbb{Z}$ ist ganz analog zu der Konstruktion von $\mathbb{Z}$ aus $\mathbb{N}$, wobei die Multiplikation die Rolle der Addition übernimmt. Jede rationale Zahl kann als Bruch zweier ganzen Zahlen a, b mit $b \neq 0$ dargestellt werden, aber diese Darstellung ist nicht

eindeutig: Es gilt $a/b = a'/b'$ genau dann, wenn $ab' = a'b$. Deswegen führen wir folgende Äquivalenzrelation $\sim$ auf $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ ein:

$$(a, b) \sim (a', b') \iff ab' = a'b,$$

und setzen wir:

$$\mathbb{Q} := (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\sim.$$

Durch die injektive Abbildung

$$\begin{aligned} &\mathbb{Z} \to \mathbb{Q}, \\ &a \mapsto [(a, 1)], \end{aligned}$$

können wir $\mathbb{Z}$ mit einer Teilmenge von $\mathbb{Q}$ identifizieren.

Weiter definieren wir die arithmetischen Operationen $+$ und $\cdot$ und die totale Ordnungsrelation $\leq$ wie folgt (nach den üblichen Bruchregeln):

$$\begin{aligned} [(a_1, b_1)] + [(a_2, b_2)] &= [(a_1 b_2 + a_2 b_1, b_1 b_2)], \\ [(a_1, b_1)] \cdot [(a_2, b_2)] &= [(a_1 a_2, b_1 b_2)], \\ [(a_1, b_1)] \leq [(a_2, b_2)] &\iff \begin{cases} a_1 b_2 \leq a_2 b_1, & \text{falls } b_1 b_2 > 0, \\ a_1 b_2 \geq a_2 b_1, & \text{falls } b_1 b_2 < 0. \end{cases} \end{aligned}$$

Man sollte natürlich nachprüfen, dass $+$, $\cdot$ und $\leq$ durch diese Formeln wohldefiniert sind, was nicht schwierig ist. Außerdem setzen $+$, $\cdot$ und $\leq$ auf $\mathbb{Q}$ die entsprechenden Verknüpfungen bzw. Relation von $\mathbb{Z}$ fort.

2.4.2 Die reellen Zahlen

Im Vergleich zu den Konstruktionen von $\mathbb{Z}$ und $\mathbb{Q}$ ist die Konstruktion der reellen Zahlen $\mathbb{R}$ deutlich komplizierter, und sie benötigt etwas *analytisch*.

Die Idee ist, dass es in der Menge $\mathbb{Q}$ der rationalen Zahlen „Löcher“ bezüglich der gewöhnlichen Ordnung $\leq$ gibt, und dass wir die reellen Zahlen erhalten, indem wir diese Löcher ausfüllen. Zum Beispiel gibt es keine rationale Zahl x mit $x^2 = 2$ (Satz 1.1.15). Trotzdem kann man die rationalen Zahlen in zwei Teilmengen unterteilen, je nachdem x^2 kleiner als oder größer als 2 ist:

$$\begin{aligned} \mathbb{Q}_{<\sqrt{2}} &:= \{x \in \mathbb{Q} \mid x < 0 \text{ oder } x^2 < 2\}, \\ \mathbb{Q}_{>\sqrt{2}} &:= \{x \in \mathbb{Q} \mid x \geq 0 \text{ und } x^2 > 2\}. \end{aligned}$$

Dann ist jede Zahl in $\mathbb{Q}_{<\sqrt{2}}$ kleiner als jede Zahl in $\mathbb{Q}_{>\sqrt{2}}$, und es gilt:

$$\mathbb{Q}_{<\sqrt{2}} \cup \mathbb{Q}_{>\sqrt{2}} = \mathbb{Q}, \quad \mathbb{Q}_{<\sqrt{2}} \cap \mathbb{Q}_{>\sqrt{2}} = \varnothing.$$

Aber $\mathbb{Q}_{<\sqrt{2}}$ hat kein größtes Element, und $\mathbb{Q}_{>\sqrt{2}}$ hat kein kleinstes Element. In diesem Sinne gibt es ein „Loch“ zwischen $\mathbb{Q}_{<\sqrt{2}}$ und $\mathbb{Q}_{>\sqrt{2}}$, und dieses Loch wird durch die reelle Zahl $\sqrt{2}$ ausgefüllt.

Wir erklären jetzt die zwei häufigsten Konstruktionen von $\mathbb{R}$: die „Dedekindschen“ reellen Zahlen $\mathbb{R}_{\text{Dedekind}}$ und die „Cauchyschen“ reellen Zahlen $\mathbb{R}_{\text{Cauchy}}$. Die zweite Konstruktion wird in der Volesung *Analysis I* ausführlicher behandelt. Es gibt noch weitere Konstruktionen von $\mathbb{R}$, aber die gewählte Konstruktion ist nicht wichtig, solange die Ausgabe ein *vollständiger angeordneter Körper* ist.

Definition 2.4.1 (Dedekindscher Schnitt)**.** Ein *Dedekindscher Schnitt* auf $\mathbb{Q}$ ist eine Teilmenge $A \subset \mathbb{Q}$ mit folgenden Eigenschaften:

- $A \neq \varnothing$ und $A \neq \mathbb{Q}$.

- A ist in $\mathbb{Q}$ nach unten abgeschlossen, d.h., ist $a \in A$ und ist $b \leq a$, so ist $b \in A$.
- A enthält kein größtes Element.

Ein Dedekindscher Schnitt heißt *rational*, falls das Komplement von A ein kleinstes Element besitzt.

Man kann die reellen Zahlen als die Menge aller Dedekindschen Schnitte auf $\mathbb{Q}$ definieren:

$$\mathbb{R}_{\text{Dedekind}} := \{A \mid A \text{ ist ein Dedekindscher Schnitt auf } \mathbb{Q}\} \subset \mathcal{P}(\mathbb{Q}).$$

Mit dieser Definition gibt es eine injektive Abbildung

$$\mathbb{Q} \to \mathbb{R}_{\text{Dedekind}}, \quad q \mapsto \mathbb{Q}_{<q} = \{x \in \mathbb{Q} \mid x < q\},$$

deren Bild genau aus den rationalen Schnitten besteht. Die Addition auf $\mathbb{R}_{\text{Dedekind}}$, sowie die Ordnungsrelation $\leq$, kann man auch leicht definieren:

$$\begin{aligned} A + B &= \{a + b \mid a \in A \text{ und } b \in B\}, \\ A \leq B &\iff A \subset B, \end{aligned}$$

wobei $+$ auf der rechten Seite die gewöhnliche Verknüpfung auf $\mathbb{Q}$ ist. Die Multiplikation ist etwas mühsamer zu definieren. Falls $\mathbb{Q}_{<0} \subset A$ und $\mathbb{Q}_{<0} \subset B$ (d.h., A und B entsprechen ≥ 0 reellen Zahlen), setzen wir

$$A \cdot B = \mathbb{Q}_{<0} \cup \{a \cdot b \mid a \in A \cap \mathbb{Q}_{\geq 0} \text{ und } b \in B \cap \mathbb{Q}_{\geq 0}\}.$$

Mithilfe der üblichen Vorzeichenregeln kann man diese Multiplikation auf ganz $\mathbb{R}_{\text{Dedekind}}$ fortsetzen. Man kann dann beweisen, dass $(\mathbb{R}_{\text{Dedekind}}, +, \cdot)$ ein Körper ist. Die obige injektive Abbildung $\mathbb{Q} \to \mathbb{R}_{\text{Dedekind}}$ identifiziert zudem $\mathbb{Q}$ mit einem Teilkörper von $\mathbb{R}_{\text{Dedekind}}$.

Definition 2.4.2 (Cauchyfolge in $\mathbb{Q}$)**.** Eine Folge $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}$ heißt *Cauchyfolge* wenn folgendes gilt: Zu jeder rationalen Zahl $\varepsilon > 0$ gibt es eine natürliche Zahl $N \in \mathbb{N}$, so dass für alle $n, m \geq N$ gilt $|x_n - x_m| < \varepsilon$. Sei Cauchy($\mathbb{Q}$) die Menge aller Cauchyfolgen in $\mathbb{Q}$.

Zum Beispiel ist die konstante Folge $(q)_{n\in\mathbb{N}}$ mit $q \in \mathbb{Q}$ eine Cauchyfolge. Man definiert eine Äquivalenzrelation $\sim$ auf Cauchy($\mathbb{Q}$) wie folgt: $(x_n)_{n\in\mathbb{N}} \sim (y_n)_{n\in\mathbb{N}}$ genau dann, wenn folgendes gilt: Zu jeder rationalen Zahl $\varepsilon > 0$ gibt es eine natürliche Zahl $N \in \mathbb{N}$, so dass für alle $n \geq N$ gilt $|x_n - y_n| < \varepsilon$. Man kann dann die reellen Zahlen als die Quotientenmenge

$$\mathbb{R}_{\text{Cauchy}} := \text{Cauchy}(\mathbb{Q})/\sim$$

definieren. Mit dieser Definition gibt es eine injektive Abbildung

$$\mathbb{Q} \to \mathbb{R}_{\text{Cauchy}}, \quad q \mapsto [(q)_{n\in\mathbb{N}}],$$

(Zur Injektivität: Falls $q \neq q'$, sei $\varepsilon = |q - q'|/2 \in \mathbb{Q}$. Dann $|q - q'| \not< \varepsilon$. Aus der Definition von $\sim$ folgt, dass $(q)_{n\in\mathbb{N}} \not\sim (q')_{n\in N}$.) Die Addition und Multiplikation auf $\mathbb{R}_{\text{Cauchy}}$ werden folgendermaße definiert:

$$\begin{aligned} [(x_n)_{n\in\mathbb{N}}] + [(y_n)_{n\in\mathbb{N}}] &= [(x_n + y_n)_{n\in\mathbb{N}}], \\ [(x_n)_{n\in\mathbb{N}}] \cdot [(y_n)_{n\in\mathbb{N}}] &= [(x_n \cdot y_n)_{n\in\mathbb{N}}]. \end{aligned}$$

Hier ist es notwendig, ein paar Tatsachen nachzuprüfen; erstens, dass $(x_n + y_n)_{n\in\mathbb{N}}$ bzw. $(x_n \cdot y_n)_{n\in\mathbb{N}}$ eine Cauchyfolge ist, wenn $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ Cauchyfolgen sind, und zweitens, dass $+$ und $\cdot$ auf $\sim$-Äquivalenzklassen wohldefiniert sind. Man kann auch die strenge Ordnungrelation $<$ wie folgt definieren: $[(x_n)_{n\in\mathbb{N}}] < [(y_n)_{n\in\mathbb{N}}]$ genau dann, wenn es ein $N \in \mathbb{N}$ und ein $\varepsilon \in \mathbb{Q}_{>0}$ gibt, so dass $x_n + \varepsilon < y_n$ für alle $n \geq N$.

Die Mengen $\mathbb{R}_{\text{Dedekind}}$ und $\mathbb{R}_{\text{Cauchy}}$ sehen ganz verschieden aus. Um sie zu vergleichen, definieren wir eine Abbildung

$$\begin{aligned} v\colon \text{Cauchy}(\mathbb{Q}) &\to \mathbb{R}_{\text{Dedekind}}, \\ (x_n)_{n\in N} &\mapsto \{x \in \mathbb{Q} \mid \text{es existiert } N \in \mathbb{N}, \text{ so dass } x < x_n \text{ für alle } n \geq N\}. \end{aligned}$$

Man kann leicht nachprüfen, dass

$$(x_n)_{n\in N} \sim (y_n)_{n\in N} \Longrightarrow v((x_n)_{n\in N}) = v((y_n)_{n\in N}).$$

Nach der universellen Eigenschaft der Quotientenmenge (Satz 1.4.10) erhalten wir eine induzierte Abbildung

$$\bar{v}\colon \text{Cauchy}(\mathbb{Q})/\!\sim\; = \mathbb{R}_{\text{Cauchy}} \to \mathbb{R}_{\text{Dedekind}}.$$

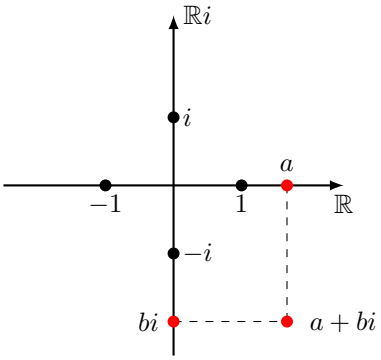
Man kann dann zeigen, dass $\bar{v}$ bijektiv ist, und zwar ein *Isomorphismus* von angeordneten Körpern, d.h., $\bar{v}$ ist kompatibel mit den arithmetischen Operationen und den Ordnungsrelationen, die wir auf beiden Seiten definiert haben. Es gilt zum Beispiel $\bar{v}(x+y) = \bar{v}(x)+\bar{v}(y)$, wobei $+$ auf der linken Seite die Addition von Äquivalenzklassen von Cauchyfolgen ist, und $+$ auf der rechten Seite die Addition von Dedekindschen Schnitten ist.

2.4.3 Die komplexen Zahlen

Eine komplexe Zahl ist ein Ausdruck der Gestalt $a + bi$ mit $a, b \in \mathbb{R}$:

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}.$$

Die reelle Zahl a bzw. b heißt der *Realteil* bzw. der *Imaginärteil* der komplexen Zahl $a + bi$. Mengentheoretisch kann man einfach $\mathbb{C}$ als $\mathbb{R} \times \mathbb{R}$ definieren, wobei ein Paar (a, b) als die komplexe Zahl $a + bi$ aufgefasst wird. Geometrisch kann man sich also die komplexe Zahl $a + bi$ als einen Punkt auf der Ebene vorstellen:



Mit folgenden Definitionen ist $(\mathbb{C}, +, \cdot)$ ein Körper:

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i, \\ (a + bi) \cdot (c + di) &= (ac - bd) + (ad + bc)i. \end{aligned}$$

Insbesondere ist $i^2 = -1$, wobei $i := 0 + 1i$. Umgekehrt folgen die obigen Formeln für $+$ und $\cdot$ aus $i^2 = -1$ und den Körperaxiomen. Wir identifizieren $\mathbb{R}$ mit einem Teilkörper von $\mathbb{C}$ mit Hilfe der injektiven Abbildung

$$\mathbb{R} \to \mathbb{C}, \quad a \mapsto a + 0i.$$

Bemerkung 2.4.3. Im Gegensatz zu $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$, gibt es auf $\mathbb{C}$ keine vernünftige totale Ordnung $\leq$. Insbesondere gibt es keine „positive“ und „negative“ komplexe Zahlen.

Beispiel 2.4.4. Eine komplexe Zahl $a + bi$ heißt *rational*, falls $a, b \in \mathbb{Q}$. Die rationalen komplexen Zahlen bilden einen Teilkörper $\mathbb{Q}(i) \subset \mathbb{C}$.

Man kann die Konstruktion der Zahlenmengen $\mathbb{Z}$ und $\mathbb{Q}$ damit motivieren, dass man bestimmte algebraische Gleichungen lösen will. Gleichungen der Gestalt

$$x + a = b, \quad a, b \in \mathbb{N},$$

since nicht immer mit $x \in \mathbb{N}$ lösbar. Deswegen führen wir die ganzen Zahlen $\mathbb{Z}$ ein, und dann haben *alle* Gleichungen

$$x + a = b, \quad a, b \in \mathbb{Z},$$

eine Lösung $x \in \mathbb{Z}$. In ähnlicher Weise, Gleichungen der Gestalt

$$a \cdot x = b, \quad a, b \in \mathbb{Z}, \quad a \neq 0,$$

sind nicht immer mit $x \in \mathbb{Z}$ lösbar. Deswegen führen wir die rationalen Zahlen $\mathbb{Q}$ ein, mit denen alle solchen Gleichungen lösbar sind. Der Übergang von $\mathbb{Q}$ nach $\mathbb{R}$ ist von anderer Art: Es handelt sich um eine analytische und nicht algebraische Konstruktion (obwohl es auch algebraische Gleichungen gibt, die in $\mathbb{R}$ aber nicht in $\mathbb{Q}$ lösbar sind, z.B. $x^2 = 2$). Nun ist die Gleichung

$$x^2 = a, \quad a \in \mathbb{R},$$

nur mit $x \in \mathbb{R}$ lösbar, wenn $a \geq 0$. Dies motiviert die Einführung der komplexen Zahlen. Es stellt sich heraus, dass in $\mathbb{C}$ *alle* algebraische Gleichungen lösbar sind, und deswegen benötigen wir keine weitere Erweiterung von $\mathbb{C}$. Das ist der *Fundamentalsatz der Algebra*, den wir jetzt genauer formulieren.

Definition 2.4.5 (algebraisch abgeschlossener Körper)**.** Ein Körper K heißt *algebraisch abgeschlossen*, falls jede Gleichung der Gestalt

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

mit $n \geq 1$ und $a_i \in K$ eine Lösung $x \in K$ besitzt.

***Satz 2.4.6** (Fundamentalsatz der Algebra)**.** *Der Körper $\mathbb{C}$ ist algebraisch abgeschlossen.*

Im Gegensatz dazu sind $\mathbb{Q}$ und $\mathbb{R}$ nicht algebraisch abgeschlossen.

2.4.4 Endliche Körper

Sei $n \geq 1$ eine natürliche Zahl. Zur Erinnerung ist die Menge der Restklassen modulo n, $\mathbb{Z}/n\mathbb{Z}$, die Quotientenmenge von $\mathbb{Z}$ bezüglich der Kongruenzrelation $\equiv_n$ (Beispiel 1.4.9). Sie ist eine endliche Menge der Mächtigkeit n:

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \ldots, [n-1]\}.$$

Die Äquivalenzrelation $\equiv_n$ ist kompatibel mit den arithmetischen Operationen $+$ und $\cdot$ im folgenden Sinne: Sind $x \equiv_n x'$ und $y \equiv_n y'$, so sind $x + y \equiv_n x' + y'$ und $x \cdot y \equiv_n x' \cdot y'$. Daraus folgt, dass die wie folgt definierten Operationen auf $\mathbb{Z}/n\mathbb{Z}$ wohldefiniert sind:

$$[x] + [y] := [x + y],$$
$$[x] \cdot [y] := [x \cdot y].$$

Mann kann sogar zeigen, dass das Tripel $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ ein kommutativer Ring ist (siehe Bemerkung 2.3.5).

Bemerkung 2.4.7. Wenn $n = 24$, ist uns die Addition auf $\mathbb{Z}/24\mathbb{Z}$ vom Rechnen mit Uhrzeiten sehr bekannt. Zum Beispiel können wir die Gleichung $[5] - [8] = [21]$ in $\mathbb{Z}/24\mathbb{Z}$ als „8 Stunden vor 5 Uhr ist 21 Uhr“ verstehen.

***Satz 2.4.8.** $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ *ist genau dann ein Körper, wenn n eine Primzahl ist.*

Der Körper $\mathbb{Z}/p\mathbb{Z}$ mit $p \in \mathbb{N}$ eine Primzahl wird auch mit $\mathbb{F}_p$ bezeichnet. Die Charakteristik von $\mathbb{F}_p$ ist gleich p.

Beispiel 2.4.9. Der Körper $\mathbb{F}_2$ hat genau zwei Elemente, nämlich 0 und 1. Die Addition und Multiplikation werden in folgenden Tabellen explizit dargelegt:

$\mathbb{F}_2 = \{0, 1\}$

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Die Operationen + und · in diesem Körper können auch als logische Operationen aufgefasst werden: Wenn wir 0 als „falsch“ und 1 als „wahr“ interpretieren, dann entspricht + dem exklusiven Oder und · der Konjunktion $\wedge$.

Beispiel 2.4.10. Hier sind die Additions- und Multiplikationstabellen des Körpers $\mathbb{F}_3$:

$\mathbb{F}_3 = \{0, 1, -1\}$

$+$	0	1	-1
0	0	1	-1
1	1	-1	0
-1	-1	0	1

$\cdot$	0	1	-1
0	0	0	0
1	0	1	-1
-1	0	-1	1

Bemerkung 2.4.11. In der Vorlesung *Algebra* wird gezeigt, dass ein endlicher Körper mit q Elementen genau dann existiert, wenn q eine Primzahlpotenz ist, d.h., wenn $q = p^n$ mit einer Primzahl p und einer natürlichen Zahl $n \geq 1$. Außerdem ist ein solcher Körper eindeutig bis auf Isomorphie und wird mit $\mathbb{F}_q$ bezeichnet. Die Charakteristik von $\mathbb{F}_{p^n}$ ist gleich p. Zum Beispiel gibt es einen Körper $\mathbb{F}_4$ mit vier Elementen und folgender Addition bzw. Multiplikation:

$\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$

$+$	0	1	α	β
0	0	1	α	β
1	1	0	β	α
α	α	β	0	1
β	β	α	1	0

$\cdot$	0	1	α	β
0	0	0	0	0
1	0	1	α	β
α	0	α	β	1
β	0	β	1	α

Bemerkung 2.4.12. Die endlichen Körper $\mathbb{F}_q$ sind nicht algebraisch abgeschlossen. Es existiert aber auch algebraisch abgeschlossene Körper der Primcharakteristik p.

Kapitel 3

Vektorraüme

In diesem Kapitel legen wir einen Körper K fest. Der Körper K heißt der *Grundkörper*, und die Elemente von K heißen *Skalare*. Wir verwenden üblicherweise griechische Buchstaben $\lambda, \mu, \ldots$ für Skalare.

3.1 Das prototypische Beispiel

Sei $n \in \mathbb{N}$. Das prototypische Beispiel eines Vektorraums über K ist die Menge K^n aller n-Tupel von Elementen von K:

$$K^n = \underbrace{K \times \cdots \times K}_{n \text{ mal}} = \{(x_1, \ldots, x_n) \mid x_i \in K\}.$$

Wir werden oft ein n-Tupel $(x_1, x_2, \ldots, x_n) \in K^n$ als *Spaltenvektor*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

darstellen. Der Grund für eine solche Darstellung wird später im Zusammenhang mit der Multiplikation von Matrizen begründet werden (siehe Abschnitt 4.2.1)

Sei $i \in \{1, \ldots, n\}$. Die i-te kanonische Projektion von K^n auf K ist die Abbildung

$$\begin{aligned} \pi_i \colon K^n &\to K, \\ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &\mapsto x_i. \end{aligned}$$

Ist $x \in K^n$, so schreibt man üblicherweise x_i für $\pi_i(x)$, so dass

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Das Element $x_i \in K$ heißt die i-te *Koordinate* des n-Tupels $x \in K^n$.

Bemerkung 3.1.1. Wenn $n = 0$ ist K^n das Produkt einer Mengenfamilie mit leerer Indexmenge. Nach Definition 1.2.15(iv) besteht also K^0 aus genau einem Element, dem „leeren Spaltenvektor“. Wenn $n = 1$ ist $K^n = K$.

Definition 3.1.2 (Addition und Skalarmultiplikation auf K^n)**.**

- Die *Addition* auf K^n ist die wie folgt definierte Abbildung:

$$+\colon K^n \times K^n \to K^n,$$
$$\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right) \mapsto \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

- Die *Skalarmultiplikation* auf K^n die wie folgt definierte Abbildung:

$$\cdot\colon K \times K^n \to K^n,$$
$$\left(\lambda, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) \mapsto \begin{pmatrix} \lambda \cdot x_1 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}.$$

Lemma 3.1.3. *Seien $G_1, \dots, G_n$ Gruppen. Dann ist das Produkt $G_1 \times \cdots \times G_n$ eine Gruppe mit der komponentenweisen Verknüpfung*

$$(g_1, \dots, g_n) \cdot (h_1, \dots, h_n) := (g_1 \cdot h_1, \dots, g_n \cdot h_n).$$

Die Gruppe $G_1 \times \cdots \times G_n$ ist abelsch, falls alle Gruppen $G_1, \dots, G_n$ abelsch sind.

Beweis. Alle Gruppenaxiome für $G_1 \times \cdots \times G_n$ folgen unmittelbar aus den Gruppenaxiomen für die einzelnen Faktoren G_i. Zur Assoziativität gilt nach Definition der Verknüpfung:

$$\big((g_1, \dots, g_n)(h_1, \dots, h_n)\big)(k_1, \dots, k_n) = ((g_1h_1)k_1, \dots, (g_nh_n)k_n),$$
$$(g_1, \dots, g_n)\big((h_1, \dots, h_n)(k_1, \dots, k_n)\big) = (g_1(h_1k_1), \dots, g_n(h_nk_n)),$$

und die rechten Seiten sind gleich nach der Assoziativität in jedem G_i. Im abelschen Fall wird die Kommutativität auf ähnliche Weise nachgeprüft. Das neutrale Element ist das n-Tupel der neutralen Elemente $(e, \dots, e)$. Das Inverse von $(g_1, \dots, g_n)$ ist $(g_1^{-1}, \dots, g_n^{-1})$. □

Bemerkung 3.1.4. Keine ähnliche Aussage gilt für Körper: Das Produkt $K_1 \times K_2$ zweier Körper ist *kein* Körper bzgl. der komponentenweisen Verknüpfungen, da z.B. $(1, 0)$ kein multiplikatives Inverses besitzt (es ist jedoch ein kommutativer Ring).

Proposition 3.1.5 (Eigenschaften der Addition und der Skalarmultiplikation auf K^n)**.**

(i) *$(K^n, +)$ ist eine abelsche Gruppe.*

(ii) *Für alle $\lambda, \mu \in K$ und $x \in K^n$ gilt*

$$(\lambda \cdot \mu) \cdot x = \lambda \cdot (\mu \cdot x).$$

(iii) *Für alle $x \in K^n$ gilt*

$$1 \cdot x = x.$$

(iv) *Für alle $\lambda, \mu \in K$ und $x, y \in K^n$ gilt*

$$\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y,$$
$$(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x.$$

Beweis. Aussage (i) ist der Sonderfall von Lemma 3.1.3 mit $G_1 = \cdots = G_n = K$. Die Beweise der Eigenschaften (ii)–(iv) sind ähnlich: Sie lassen sich komponentweise nachrechnen und folgen aus den entsprechenden Eigenschaften der Verknüpfungen $+$ und $\cdot$ des Körpers K. Wir beweisen stellvertretend Aussage (ii):

$$\begin{aligned}
(\lambda \cdot \mu) \cdot x &= \begin{pmatrix} (\lambda \cdot \mu) \cdot x_1 \\ \vdots \\ (\lambda \cdot \mu) \cdot x_n \end{pmatrix} && \text{(Definition der Skalarmultiplikation)} \\
&= \begin{pmatrix} \lambda \cdot (\mu \cdot x_1) \\ \vdots \\ \lambda \cdot (\mu \cdot x_n) \end{pmatrix} && \text{(Assoziativität von } \cdot \text{ in } K) \\
&= \lambda \cdot \begin{pmatrix} \mu \cdot x_1 \\ \vdots \\ \mu \cdot x_n \end{pmatrix} && \text{(Definition der Skalarmultiplikation)} \\
&= \lambda \cdot (\mu \cdot x). && \text{(Definition der Skalarmultiplikation)}
\end{aligned}$$

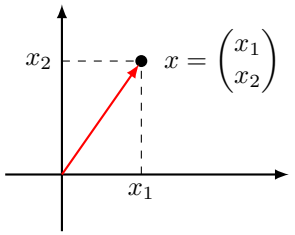
□

Definition 3.1.6 (Standardeinheitsvektoren)**.** Sei $i \in \{1, \ldots, n\}$. Der i-te *Standardeinheitsvektor* in K^n ist

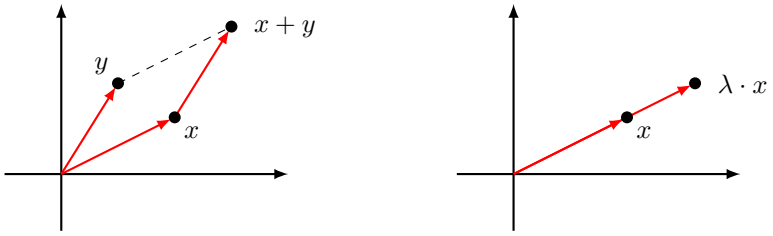
$$e_i := \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

wobei 1 in der i-ten Zeile liegt und alle anderen Koordinaten null sind.

Wenn $K = \mathbb{R}$ können wir $\mathbb{R}^2$ und $\mathbb{R}^3$ als die 2-dimensionale Ebene und den 3-dimensionalen Raum der Euklidischen Geometrie auffassen. Das heißt, wir können Elemente von $\mathbb{R}^2$ und $\mathbb{R}^3$ mit Punkten der Ebene und des Raums identifizieren. Manchmal stellt man auch ein Element x aus $\mathbb{R}^2$ oder $\mathbb{R}^3$ mit einem Pfeil von dem Nullpunkt nach dem Punkt x dar:



Diese Darstellung ist hilfreich, um die Addition und Skalarmultiplikation auf geometrische Weise zu beschreiben: Man erhält die Summe $x + y$ zweier Elemente x und y, indem man der Pfeil von y längs des von x verschiebt, und man erhält das λ-fache $\lambda \cdot x$ von x, indem man der Pfeil von x um den Faktor λ skaliert:



Diese geometrische Anschauung ist auch hilfreich in höherer Dimension oder bei anderen Körpern, selbst wenn man nicht zeichnen kann.

3.2 Vektorraüme

Wir abstrahieren jetzt die in Proposition 3.1.5 bewiesenen Eigenschaften von K^n zum Begriff des Vektorraums:

Definition 3.2.1 (Vektorraum)**.** Ein *Vektorraum* über K, oder K*-Vektorraum*, ist ein Tripel $(V, +, \cdot)$, bestehend aus einer Menge V und Abbildungen

$$\begin{aligned} +\colon V \times V \to V, &\quad (v, w) \mapsto v + w, \\ \cdot\colon K \times V \to V, &\quad (\lambda, v) \mapsto \lambda \cdot v, \end{aligned}$$

die als *Addition* und *Skalarmultiplikation* bezeichnet werden, mit folgenden Eigenschaften:

(i) $(V, +)$ ist eine abelsche Gruppe.

(ii) Für alle $\lambda, \mu \in K$ und $v \in V$ gilt

$$(\lambda \cdot \mu) \cdot v = \lambda \cdot (\mu \cdot v).$$

(iii) Für alle $v \in V$ gilt

$$1 \cdot v = v.$$

(iv) Für alle $\lambda, \mu \in K$ und $v, w \in V$ gilt

$$\begin{aligned} \lambda \cdot (v + w) &= \lambda \cdot v + \lambda \cdot w, \\ (\lambda + \mu) \cdot v &= \lambda \cdot v + \mu \cdot v. \end{aligned}$$

Elemente von V heißen *Vektoren*. Das neutrale Element 0 bzgl. $+$ heißt der *Nullvektor*.

Bemerkung 3.2.2. Im zweiten Axiom, $(\lambda \cdot \mu) \cdot v = \lambda \cdot (\mu \cdot v)$, das erste $\cdot$ ist die Multiplikation auf K, und die anderen drei $\cdot$ sind Skalarmultiplikation. Wegen diesem Axiom darf man einfach $\lambda \cdot \mu \cdot v$ schreiben. Allgemeiner, sind $\lambda_1, \ldots, \lambda_n$ Skalare, so darf man $\lambda_1 \cdot \ldots \cdot \lambda_n \cdot v$ ohne Klammern schreiben. Wie üblich wird das Symbol $\cdot$ oft ganz unterdrückt.

Beispiel 3.2.3 (Vektorraum der n-Tupel)**.** Für jedes $n \in \mathbb{N}$ ist $(K^n, +, \cdot)$ ein Vektorraum über K, wobei $+$ und $\cdot$ die Addition und Skalarmultiplikation von n-Tupeln sind (Definition 3.1.2). Dies folgt aus Proposition 3.1.5.

Beispiel 3.2.4 (Körpererweiterungen als Vektorräume)**.** Sei L ein Körper. Ein Teilmenge $K \subset L$ heißt *Teilkörper*, wenn sich die Addition und Multiplikation auf L zu K einschränken und K mit diesen eingeschränkten Verknüpfungen einen Körper bildet. Man sagt dann auch, dass L eine *Körpererweiterung* von K ist. Zum Beispiel: $\mathbb{R}$ und $\mathbb{C}$ sind Körpererweiterungen von $\mathbb{Q}$, und $\mathbb{C}$ ist auch eine Körpererweiterung von $\mathbb{R}$.

Wenn L eine Körpererweiterung von K ist, dann bildet L mit seiner Addition und seiner auf $K \times L$ eingeschränkten Multiplikation einen K-Vektorraum: Die Axiome (i)–(iv) der Definition 3.2.1 sind Sonderfälle der Körperaxiome für L.

Beispiel 3.2.5 (Vektorräume von Abbildungen)**.** Sei V ein K-Vektorraum und X eine beliebige Menge. Dann ist die Menge $\mathrm{Abb}(X, V)$ aller Abbildungen von X nach V ein K-Vektorraum bezüglich der punktweisen Addition bzw. Skalarmultiplikation:

$$\begin{aligned} +\colon \mathrm{Abb}(X, V) \times \mathrm{Abb}(X, V) &\to \mathrm{Abb}(X, V), \\ (f, g) &\mapsto (x \mapsto f(x) + g(x)), \end{aligned}$$

$$\begin{aligned} \cdot\colon K \times \mathrm{Abb}(X, V) &\to \mathrm{Abb}(X, V), \\ (\lambda, f) &\mapsto (x \mapsto \lambda \cdot f(x)). \end{aligned}$$

Alle Axiome der Definition 3.2.1 können punktweise nachgeprüft werden und folgen aus den entsprechenden Axiomen für V.

Insbesondere haben wir den K-Vektorraum $\mathrm{Abb}(X, K)$ aller Abbildungen von X nach K, der auch mit K^X bezeichnet wird. Der K-Vektorraum K^n kann als Sonderfall dieser Konstruktion aufgefasst werden, nämlich mit $X = \{1, \ldots, n\}$: Effektiv ist ein n-Tupel in K nichts anderes als eine Abbildung $\{1, \ldots, n\} \to K$.

Proposition 3.2.6 (Rechnen in Vektorräumen)**.** *Sei V ein Vektorraum über K.*

(i) *Für alle $\lambda \in K$ gilt $\lambda \cdot 0 = 0$, wobei $0 \in V$ der Nullvektor ist.*

(ii) *Für alle $v \in V$ gilt $0 \cdot v = 0$. Dabei bezeichnet 0 auf der linken Seite den Nullskalar und auf der rechten Seite den Nullvektor.*

(iii) *Für alle $\lambda \in K$ und $v \in V$ gilt $(-\lambda) \cdot v = -(\lambda \cdot v)$ und $\lambda \cdot (-v) = -(\lambda \cdot v)$.*

(iv) *Sind $\lambda \in K \setminus \{0\}$ und $v \in V \setminus \{0\}$, so ist $\lambda \cdot v \in V \setminus \{0\}$.*

Beweis. Zu (i). Nach Axiom (iv) gilt

$$\lambda \cdot 0 = \lambda \cdot (0 + 0) = \lambda \cdot 0 + \lambda \cdot 0.$$

Da $(V, +)$ eine Gruppe ist, können wir $\lambda \cdot 0$ von beiden Seiten subtrahieren, und erhalten wir $0 = \lambda \cdot 0$.

Zu (ii). Ähnlicher Beweis: Nach Axiom (iv) gilt

$$0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v,$$

und daher $0 = 0 \cdot v$.

Zu (iii). Nach der Eindeutigkeit von inversen Elementen genügt es zu zeigen, dass $(-\lambda) \cdot v$ und $\lambda \cdot (-v)$ inverse Elemente von $\lambda \cdot v$ bzgl. $+$ sind. Nach Axiom (iv) gilt

$$(-\lambda) \cdot v + \lambda \cdot v = ((-\lambda) + \lambda) \cdot v = 0 \cdot v,$$

und $0 \cdot v$ ist gleich Null nach (ii). Nach Axiom (iv) gilt ebenfalls

$$\lambda \cdot (-v) + \lambda \cdot v = \lambda \cdot ((-v) + v) = \lambda \cdot 0,$$

und $\lambda \cdot 0$ ist gleich Null nach (i).

Zu (iv). Wir beweisen die äquivalente Aussage: Ist $\lambda \cdot v = 0$ und $\lambda \neq 0$, so ist $v = 0$. Da K ein Körper ist hat λ ein Inverses λ^{-1} bzgl. $\cdot$. Dann gilt:

$$\begin{aligned}
v &= 1 \cdot v && \text{(Axiom (iii))} \\
&= (\lambda^{-1} \cdot \lambda) \cdot v && (\lambda^{-1} \text{ invers zu } \lambda) \\
&= \lambda^{-1} \cdot (\lambda \cdot v) && \text{(Axiom (ii))} \\
&= \lambda^{-1} \cdot 0 && \text{(Annahme)} \\
&= 0. && \text{(nach (i))}
\end{aligned}$$

□

3.2.1 Untervektorräume

Definition 3.2.7 (Untervektorraum)**.** Sei $(V, +, \cdot)$ ein Vektorraum über K. Eine Teilmenge $U \subset V$ heißt *Untervektorraum*, wenn sich die Abbildungen $+ \colon V \times V \to V$ und $\cdot \colon K \times V \to V$ zu Abbildungen $+ \colon U \times U \to U$ und $\cdot \colon K \times U \to U$ einschränken, und U mit diesen eingeschränkten Verknüpfungen ein K-Vektorraum ist.

Ist U ein Untervektorraum eines K-Vektorraums $(V, +, \cdot)$, so betrachten wir immer U als K-Vektorraum mit den eingeschränkten Verknüpfungen. Um Untervektorräume zu erkennen verwenden wir folgendes Kriterium:

Proposition 3.2.8 (Kriterium für Untervektorräume)**.** *Sei V ein Vektorraum über K. Eine Teilmenge $U \subset V$ ist genau dann ein Untervektorraum, wenn folgende drei Bedingungen erfüllt sind:*

(i) *U ist nicht leer.*

(ii) *Für alle $v, w \in U$ gilt $v + w \in U$.*

(iii) *Für alle $v \in U$ und $\lambda \in K$ gilt $\lambda \cdot v \in U$.*

Außerdem gilt in diesem Fall:

(iv) *Der Nullvektor $0 \in V$ liegt in U, und ist auch der Nullvektor von U.*

(v) *Für alle $v \in U$, der inverse Vektor $-v \in V$ liegt in U, und ist auch das Inverse von v in U.*

Beweis. Ist U ein Untervektorraum, so sind Bedingungen (ii) und (iii) nach Definition erfüllt. Außerdem ist U nicht leer, da es als Vektorraum einen Nullvektor enthält.

Umgekehrt, sei $U \subset V$ eine Teilmenge, die die Bedingungen (i)–(iii) erfüllt. Nach (ii) und (iii) schränken sich die Verknüpfungen $+$ und $\cdot$ auf U ein, und es ist zu zeigen, dass $(U, +, \cdot)$ ein K-Vektorraum ist. Die Axiome (ii)–(iv) in der Definition 3.2.1 eines Vektorraums, wie auch die Assoziativität und Kommutativität von $+$, gelten für alle $v, w \in V$ und $\lambda, \mu \in K$; insbesondere gelten sie für alle $v, w \in U$ und $\lambda, \mu \in K$. Es bleibt also zu zeigen, dass in U ein neutrales Element und inverse Elemente bzgl. $+$ existieren. Dazu genügt es (iv) und (v) zu beweisen. Nach Proposition 3.2.6(iii) und dem Vektorraumaxiom $1 \cdot v = v$ gilt

$$(-1) \cdot v = -(1 \cdot v) = -v$$

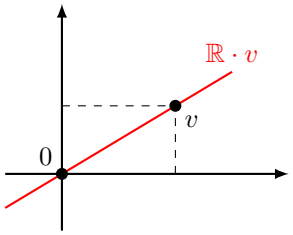
für alle $v \in V$. Aus (iii) folgt, dass $-v \in U$ für alle $v \in U$, d.h., es gilt (v). Nach (i) existiert ein $u \in U$. Da $u + (-u) = 0$, liegt 0 in U nach (v) und (ii), d.h., es gilt (iv). □

Beispiel 3.2.9. Sei V ein beliebiger K-Vektorraum. Dann sind $\{0\}$ und V Untervektorräume von V.

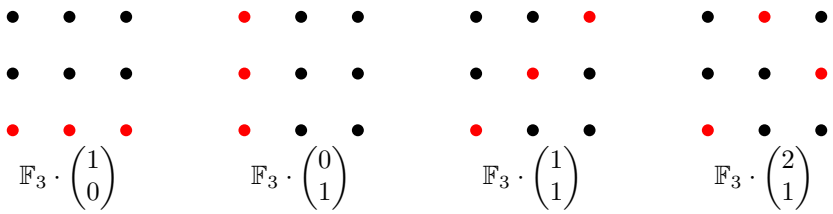
Beispiel 3.2.10 (Ursprungsgeraden)**.** Sei V ein K-Vektorraum und $v \in V \setminus \{0\}$. Dann ist die Teilmenge

$$K \cdot v := \{\lambda \cdot v \mid \lambda \in K\} \subset V$$

ein Untervektorraum. Dieser Untervektorraum heißt die *von v aufgespannte Gerade* in V. Wenn $K = \mathbb{R}$ und $V = \mathbb{R}^2$ oder $\mathbb{R}^3$, dann ist $\mathbb{R} \cdot v$ eine Gerade im üblichen Sinn: Sie ist nämlich die Gerade, die durch den Ursprung und v läuft:



Beispiel 3.2.11 (Ursprungsgeraden über endlichen Körpern)**.** Sei p eine Primzahl. Eine Ursprungsgerade in einem $\mathbb{F}_p$-Vektorraum besteht aus genau p Elementen. Folgendes Bild zeigt alle vier Ursprungsgeraden in $\mathbb{F}_3^2$:



Beispiel 3.2.12. Sei I eine beliebige Menge. Wir betrachten den K-Vektorraum $K^I = \mathrm{Abb}(I, K)$ aller Abbildungen von I nach K (siehe Beispiel 3.2.5). Sei $K^{(I)}$ seine Teilmenge bestehend aus aller Abbildungen, die außerhalb einer endlichen Teilmenge von I null sind:

$$K^{(I)} := \{f\colon I \to K \mid \text{es existiert } J \subset I \text{ endlich, so dass } f(I \setminus J) \subset \{0\}\}.$$

Dann ist $K^{(I)}$ ein Untervektorraum von K^I, denn: (i) Er enthält den Nullvektor; (ii) falls f_1 außerhalb J_1 und f_2 außerhalb J_2 null sind, dann ist $f_1 + f_2$ außerhalb $J_1 \cup J_2$ null; (iii) falls f außerhalb J null ist, dann ist $\lambda \cdot f$ ebenfalls außerhalb J null. Man beachte, dass $K^{(I)} = K^I$, wenn I endlich ist.

Beispiel 3.2.13 (konvergente Folgen)**.** Sei $\mathbb{R}^{\mathbb{N}}$ der $\mathbb{R}$-Vektorraum aller Folgen in $\mathbb{R}$. Dann ist die Teilmenge

$$\mathrm{Konv}(\mathbb{R}) := \{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid (x_n)_{n\in\mathbb{N}} \text{ konvergiert}\} \subset \mathbb{R}^{\mathbb{N}}$$

ein Untervektorraum. Dies folgt aus dem Kriterium 3.2.8, da die Summe zweier konvergenten Folgen konvergiert, wie auch die Skalarmultiplikation einer konvergenten Folge mit einer reellen Zahl.

Beispiel 3.2.14 (Funktionenräume)**.** Sei $K = \mathbb{R}$ und seien $a < b$ reelle Zahlen. Nach Beispiel 3.2.5 ist die Menge $\mathrm{Abb}([a, b], \mathbb{R})$ aller reellen Funktionen auf $[a, b]$, versehen mit der punktweisen Addition bzw. Skalarmultiplikation, ein $\mathbb{R}$-Vektorraum. In der Analysis betrachtet man viele verschiedene Sorten solcher Funktionen, und die entsprechenden Teilmengen von $\mathrm{Abb}([a, b], \mathbb{R})$ sind oft Untervektorräume. Das gilt zum Beispiel für stetige, differenzierbare, stetig differenzierbare, beliebig oft differenzierbare und Riemann-integrierbare Funktionen.

Bemerkung 3.2.15 (Abhängigkeit vom Grundkörper)**.** Sei $K \subset L$ eine Körpererweiterung (siehe Beispiel 3.2.4). Dann können wir jeden L-Vektorraum V als K-Vektorraum betrachten, indem wir die Skalarmultiplikation $\cdot\colon L \times V \to V$ auf $K \times V$ einschränken. Ob eine Teilmenge $U \subset V$ ein Untervektorraum ist, hängt davon ab, über welchem Körper wir V als Vektorraum betrachten. Zum Beispiel ist $\mathbb{R}$ ein Untervektorraum von $\mathbb{C}$, wenn wir $\mathbb{C}$ als $\mathbb{R}$-Vektorraum betrachten, aber nicht wenn wir $\mathbb{C}$ als $\mathbb{C}$-Vektorraum betrachten. Deswegen sollten wir eher von K-Untervektorräumen sprechen, wenn solche Mehrdeutigkeiten möglich sind.

Proposition 3.2.16 (Durchschnitt von Untervektorräumen)**.** *Sei V ein Vektorraum über K. Sind $U, W \subset V$ Untervektorräume, so ist $U \cap W \subset V$ ein Untervektorraum.*

Allgemeiner, ist $(U_i)_{i\in I}$ eine beliebige Familie von Untervektorräumen von V, so ist $\bigcap_{i\in I} U_i \subset V$ ein Untervektorraum.

Beweis. Wir beweisen die allgemeinere Aussage mithilfe der Proposition 3.2.8. Der Durchschnitt $\bigcap_{i\in I} U_i$ ist nicht leer, da jedes U_i den Nullvektor von V enthält. Die Bedingungen (ii) und (iii) für alle U_i implizieren dieselben Bedingungen für $\bigcap_{i\in I} U_i$. □

Definition 3.2.17 (erzeugter Untervektorraum, Erzeugendensystem)**.** Sei V ein Vektorraum über K und $E \subset V$ eine Teilmenge.

- Der *von E erzeugte Untervektorraum*, oder die *lineare Hülle* von E, ist

$$\mathrm{Span}_K(E) := \bigcap_{U\in\mathcal{U}(E)} U,$$

  wobei $\mathcal{U}(E)$ die Menge aller Untervektorräume $U \subset V$ mit $E \subset U$ ist. Nach Proposition 3.2.16 ist $\mathrm{Span}_K(E)$ ein Untervektorraum von V, und zwar der kleinste Untervektorraum, der E enthält.

- E heißt *Erzeugendensystem* von V, falls $\mathrm{Span}_K(E) = V$.

Ein Ausdruck der Gestalt

$$\sum_{i=1}^{n} \lambda_i \cdot v_i$$

mit $n \in \mathbb{N}$, $\lambda_i \in K$ und $v_i \in V$ heißt *Linearombination* der Vektoren v_i. Die folgende Proposition gibt eine explizite Beschreibung des Untervektorraums $\mathrm{Span}_K(E)$ als die Menge aller Linearkombinationen von Vektoren aus E. Insbesondere ist E genau dann ein Erzeugendensystem von V, wenn jeder Vektor aus V eine Linearkombination von Vektoren aus E ist.

Proposition 3.2.18. *Sei V ein Vektorraum über K und $E \subset V$ eine Teilmenge. Dann gilt*

$$\mathrm{Span}_K(E) = \left\{ \sum_{i=1}^{n} \lambda_i \cdot v_i \;\middle|\; n \in \mathbb{N},\ \lambda_i \in K,\ v_i \in E \right\}.$$

(Dabei ist die leere Summe $\sum_{i=1}^{0} \lambda_i \cdot v_i$ gleich 0, nach Notation 2.1.11.)

Beweis. Zu $\supset$. Nach Definition gilt $E \subset \mathrm{Span}_K(E)$. Da $\mathrm{Span}_K(E)$ ein Untervektorraum ist, jede Summe $\sum_{i=1}^{n} \lambda_i \cdot v_i$ auf der rechten Seite liegt in $\mathrm{Span}_K(E)$.

Zu $\subset$. Sei U die Menge auf der rechten Seite. Es gilt $E \subset U$. Nach Definition von $\mathrm{Span}_K(E)$ genügt es also zu zeigen, dass U ein Untervektorraum von V ist. Dazu verwenden wir das Kriterium 3.2.8. Es gilt $0 \in U$, insbesondere ist U nicht leer. Bedingung (ii) ist offensichtlich, und Bedingung (iii) folgt aus den Vektorraumaxiomen, da

$$\lambda \cdot \left(\sum_{i=1}^{n} \lambda_i \cdot v_i \right) = \sum_{i=1}^{n} (\lambda \cdot \lambda_i) \cdot v_i. \qquad \square$$

Beispiel 3.2.19.

(i) Es gilt $\mathrm{Span}_K(\varnothing) = \{0\}$, weil $\{0\}$ bereits ein Untervektorraum ist, der $\varnothing$ enthält.

(ii) Die Menge der Standardeinheitsvektoren $\{e_1, \ldots, e_n\} \subset K^n$ ist ein Erzeugendensystem von K^n, da jeder Vektor $x \in K^n$ als

$$x = x_1 \cdot e_1 + \cdots + x_n \cdot e_n$$

dargestellt werden kann.

(iii) Sei $E \subset \mathbb{Q}^n$ die Menge aller n-Tupel $(x_1, \ldots, x_n)$ mit $x_i > 7$ für alle i, d.h., $E = (\mathbb{Q}_{>7})^n$. Dann ist E ein Erzeugendensystem des $\mathbb{Q}$-Vektorraums $\mathbb{Q}^n$, denn: Sei $U \subset \mathbb{Q}^n$ ein Untervektorraum, der E enthält. Es gilt $(8, \ldots, 8) \in E$ und $e_i + (8, \ldots, 8) \in E$ für alle $i \in \{1, \ldots, n\}$, und damit $e_i \in U$. Aus (ii) folgt, dass $U = \mathbb{Q}^n$.

(iv) Sei I eine Menge und sei K^I der Vektorraum aller Abbildungen von I nach K (siehe Beispiel 3.2.5). Man kann jedem $i \in I$ einen „Standardeinheitsvektor“ $e_i \in K^I$ zuordnen, wobei

$$\begin{aligned} e_i \colon I &\to K, \\ j &\mapsto \begin{cases} 1, & \text{falls } j = i, \\ 0, & \text{falls } j \neq i. \end{cases} \end{aligned}$$

Im Gegensatz zu (ii), wenn I *unendlich* ist, ist die Menge $\{e_i \mid i \in I\}$ *kein* Erzeugendensystem von K^I. Denn jede Abbildung $f \colon I \to K$, die eine Linearkombination der Abbildungen e_i ist, ist gleich Null außerhalb einer endlichen Teilmenge von I. Es gilt also

$$\mathrm{Span}_K(\{e_i \mid i \in I\}) = K^{(I)} \subset K^I,$$

wobei $K^{(I)}$ der im Beispiel 3.2.12 definierte Untervektorraum ist.

Beispiel 3.2.20. Sei $K = \mathbb{R}$ und seien $u, v, w \in \mathbb{R}^3$ drei Vektoren. Es gibt vier verschiedenen geometrischen Möglichkeiten für $\mathrm{Span}_{\mathbb{R}}(\{u, v, w\})$:

- Falls $u = v = w = 0$, dann ist $\mathrm{Span}_{\mathbb{R}}(\{u, v, w\}) = \{0\}$.
- Falls die drei Vektoren auf derselben Ursprungsgerade G liegen, und nicht alle null sind, dann ist $\mathrm{Span}_{\mathbb{R}}(\{u, v, w\}) = G$.
- Falls die drei Vektoren auf derselben Ursprungsebene E liegen, aber nicht auf irgendeiner Ursprungsgerade, dann ist $\mathrm{Span}_{\mathbb{R}}(\{u, v, w\}) = E$.
- Falls die drei Vektoren auf keiner gemeinsamen Ursprungsebene liegen, dann ist $\mathrm{Span}_{\mathbb{R}}(\{u, v, w\}) = \mathbb{R}^3$, d.h., $\{u, v, w\}$ ist ein Erzeugendensystem.

Definition 3.2.21 (endlich erzeugt)**.** Ein K-Vektorraum V heißt *endlich erzeugt*, falls er ein endliches Erzeugendensystem besitzt.

Beispiel 3.2.22.

(i) Für alle $n \in \mathbb{N}$ ist K^n endlich erzeugt, da $\{e_1, \ldots, e_n\}$ ein endliches Erzeugendensystem von K^n ist (siehe Beispiel 3.2.19(ii)).

(ii) Ist I eine unendliche Menge, so sind K^I und $K^{(I)}$ *nicht* endlich erzeugt. Das werden wir später beweisen: Siehe Bemerkung 3.3.28.

3.2.2 Quotientenvektorräume

Im Abschnitt 1.4.1 haben wir den wichtigen Begriff des *Quotienten* einer Menge modulo einer Äquivalenzrelation eingeführt. Wenn $\sim$ eine Äquivalenzrelation auf einem Vektorraum V ist, die mit der Vektorraumstruktur in geeigneter Weise verträglich ist, dann vererbt sich die Vektorraumstruktur auf die Quotientenmenge $V/\!\sim$. Dies führt zum Begriff des *Quotientenvektorraum.*

Proposition 3.2.23. *Sei V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Sei $\sim_U$ die wie folgt definierte Relation auf V:*

$$v \sim_U w \iff v - w \in U.$$

(i) *$\sim_U$ ist eine Äquivalenzrelation auf V.*

(ii) *Die Äquivalenzklasse eines Vektors v bzgl. $\sim_U$ ist*

$$v + U := \{v + u \mid u \in U\}.$$

(iii) *Die Quotientenmenge $V/\!\sim_U$ hat die Struktur eines K-Vektorraums mit folgender Addition und Skalarmultiplikation:*

$$(v + U) + (w + U) = (v + w) + U,$$
$$\lambda \cdot (v + U) = \lambda v + U.$$

Das neutrale Element ist die Äquivalenzklasse von $0 \in V$, d.h., $0 + U = U$, und das inverse Element von $v + U$ ist $(-v) + U$.

Beweis. Zu (i). Die Reflexivität von $\sim_U$ folgt aus $0 \in U$ und die Symmetrie folgt aus der Implikation $u \in U \Rightarrow -u \in U$. Zur Transitivität: Es seien $v \sim_U w$ und $w \sim_U x$, d.h., $v - w \in U$ und $w - x \in U$. Da die Addition zweier Vektoren aus U wieder in U liegt, gilt $v - x = (v - w) + (w - x) \in U$, d.h., $v \sim_U x$.

Zu (ii). Nach Definition 1.4.4 ist die Äquivalenzklasse von v gleich

$$\{w \in V \mid w - v \in U\} = \{w \in V \mid \text{es existiert } u \in U \text{ mit } w = v + u\} = v + U.$$

Zu (iii). Wir zeigen zunächst, dass die Verknüpfungen $+$ und $\cdot$ auf $V/\sim_U$ wohldefiniert sind. Danach folgen die Vektorraumaxiome für $V/\sim_U$ unmittelbar aus den entsprechenden Axiomen für V. Seien $v \sim_U v'$ und $w \sim_U w'$. Zu zeigen ist, dass

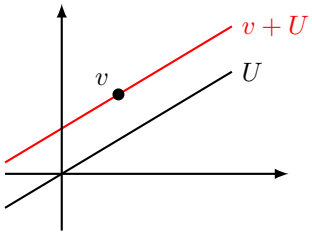
$$v + w \sim_U v' + w'$$
$$\text{und } \lambda v \sim_U \lambda v'.$$

Dies folgt aus den Gleichungen $(v + w) - (v' + w') = (v - v') + (w - w')$ und $\lambda v - \lambda v' = \lambda(v - v')$. □

Definition 3.2.24 (Quotientenvektorraum)**.** Sei V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Der in Proposition 3.2.23 definierte K-Vektorraum $V/\sim_U$ heißt der *Quotientenvektorraum* von V nach U und wird mit V/U (gelesen „V modulo U“ oder „V durch U“) bezeichnet.

Eine Teilmenge $A \subset V$ der Gestalt $A = v + U$ mit einem Untervektorraum U heißt *affiner Unterraum* von V. Affine Unterräume sind also verschobene Untervektorräume. Der Untervektorraum U ist eindeutig durch A bestimmt, da für jedes $v \in A$ gilt $U = A - v$. Der Quotientenvektorraum V/U ist nach Definition die Menge aller affinen Unterräume von V, die parallel zu U sind.

Folgendes Bild stellt einen Untervektorraum U von $\mathbb{R}^2$ und einen affinen Unterraum $v+U$ dar. Der Quotientenvektorraum $\mathbb{R}^2/U$ ist die Menge aller Geraden, die parallel zu U sind:



3.3 Basen und Dimension

Im K-Vektorraum K^n spielt die Familie der Standardeinheitsvektoren $(e_1, \ldots, e_n)$ aus Definition 3.1.6 eine besondere Rolle. Ein Grund dafür ist die folgende Eigenschaft: Jeder Vektor $x \in K^n$ kann als eine Summe

$$x = x_1 \cdot e_1 + \cdots + x_n \cdot e_n$$

dargestellt werden, mit *eindeutig bestimmten* Skalaren $x_1, \ldots, x_n \in K$. Wegen dieser Eigenschaft sagt man, dass die Familie $(e_1, \ldots, e_n)$ eine *Basis* von K^n ist. In diesem Abschnitt werden wir Basen in allgemeinen Vektorräumen definieren. Wir werden beweisen, dass jeder Vektorraum V eine Basis besitzt, und außerdem dass je zwei Basen von V dieselbe Länge haben. Die Länge einer Basis von V heißt dann die *Dimension* von V. Beispielsweise ist die Dimension von K^n gleich n.

3.3.1 Lineare Unabhängigkeit

Definition 3.3.1 (lineare Unabhängigkeit)**.** Sei V ein Vektorraum über K und sei $(v_i)_{i \in I}$ eine Familie von Elementen von V (d.h., eine Abbildung $I \to V$, $i \mapsto v_i$).

- Die Familie $(v_i)_{i\in I}$ heißt *linear unabhängig*, wenn folgendes gilt: Für jede endliche Teilmenge $J \subset I$ und Skalarfamilie $(\lambda_j)_{j\in J}$, ist $\sum_{j\in J} \lambda_j \cdot v_j = 0$, so folgt bereits $\lambda_j = 0$ für alle $j \in J$.
- Die Familie $(v_i)_{i\in I}$ heißt *linear abhängig*, wenn sie nicht linear unabhängig ist, d.h., wenn es eine endliche Teilmenge $J \subset I$ und eine Skalarfamilie $(\lambda_j)_{j\in J}$ existiert, so dass $\sum_{j\in J} \lambda_j \cdot v_j = 0$ und $\lambda_j \neq 0$ für mindestens ein $j \in J$.

Ein Ausdruck der Gestalt

$$\sum_{j\in J} \lambda_j \cdot v_j$$

mit $J \subset I$ einer endlichen Teilmenge heißt *Linearkombination* der Familie $(v_i)_{i\in I}$. Die Definition der linearen Unabhängigkeit wird oft folgendermaßen formuliert: Es gibt keine *nicht-triviale* Linearkombination der Familie $(v_i)_{i\in I}$, die gleich Null ist. Dabei heißt eine Linearkombination $\sum_{j\in J} \lambda_j \cdot v_j$ nicht-trivial, falls $\lambda_j \neq 0$ für mindestens ein $j \in J$.

Beispiel 3.3.2. Die leere Familie $(v_i)_{i\in\varnothing}$ ist immer linear unabhängig.

Beispiel 3.3.3. Sei V ein K-Vektorraum und $(v_i)_{i\in I}$ eine Familie von Vektoren aus V.

(i) Gibt es $i \in I$ mit $v_i = 0$, so ist die Familie $(v_i)_{i\in I}$ linear abhängig: Die Linearkombination $1 \cdot v_i$ ist nicht-trivial aber gleich Null.

(ii) Gibt es $i \neq j$ mit $v_i = v_j$, so ist die Familie $(v_i)_{i\in I}$ linear abhängig, da $v_i + (-1)\cdot v_j = 0$.

(iii) Folgendes Beispiel ist eine Verallgemeinerung von (i) und (ii). Es gebe einen Index $i \in I$ und eine endliche Teilmenge $J \subset I \setminus \{i\}$, so dass v_i eine Linearkombination von $(v_j)_{j\in J}$ ist, d.h., $v_i = \sum_{j\in J} \lambda_j \cdot v_j$ mit $\lambda_j \in K$. Dann ist die Familie $(v_i)_{i\in I}$ linear abhängig.

Beispiel 3.3.4. Sei $n \in \mathbb{N}$. Die Familie der Standardeinheitsvektoren $(e_i)_{i\in\{1,\dots,n\}}$ in K^n ist linear unabhängig. Denn seien $\lambda_1, \dots, \lambda_n \in K$ mit $\sum_{i=1}^n \lambda_i \cdot e_i = 0$. Nach Definition von e_i gilt

$$\sum_{i=1}^n \lambda_i \cdot e_i = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Ein n-Tupel ist genau dann gleich Null, wenn alle seinen Koordinaten gleich Null sind. Es folgt also $\lambda_1 = \cdots = \lambda_n = 0$.

Allgemeiner, die Familie $(e_i)_{i\in I}$ im K-Vektorraum K^I ist linear unabhängig (siehe Beispiel 3.2.19(iv)).

Bemerkung 3.3.5. Ob eine Familie $(v_i)_{i\in I}$ linear unabhängig ist kann nicht „paarweise“ überprüft werden. Zum Beispiel sind alle drei Familien (e_1, e_2), $(e_1, e_1 + e_2)$ und $(e_2, e_1 + e_2)$ in K^2 linear unabhängig, aber die Familie $(e_1, e_2, e_1 + e_2)$ ist linear abhängig.

Beispiel 3.3.6. Sei $K = \mathbb{R}$ und $V = \mathbb{R}^3$.

- Ein einzelner Vektor $v \in \mathbb{R}^3$ ist genau dann linear unabhängig, wenn $v \neq 0$, d.h., wenn der von $\{v\}$ erzeugter Untervektorraum eine Gerade ist.
- Zwei Vektoren $v, w \in \mathbb{R}^3$ sind genau dann linear unabhängig, wenn der von $\{v, w\}$ erzeugter Untervektorraum eine Ebene ist.
- Drei Vektoren $u, v, w \in \mathbb{R}^3$ sind genau dann linear unabhängig, wenn der von $\{u, v, w\}$ erzeugter Untervektorraum der ganze Raum $\mathbb{R}^3$ ist.
- Ein Familie bestehend aus mehr als drei Vektoren in $\mathbb{R}^3$ kann nicht linear unabhängig sein.

Proposition 3.3.7 (Charakterisierung der linearen Abhängigkeit)**.** *Sei $(v_i)_{i\in I}$ eine Familie in einem K-Vektorraum V. Folgende Aussagen sind äquivalent:*

(i) *$(v_i)_{i\in I}$ ist linear abhängig.*

(ii) *Einer der Vektoren in $(v_i)_{i\in I}$ ist eine Linearkombination der anderen. Das heißt: Es existiert $k \in I$, eine endliche Teilmenge $J \subset I \setminus \{k\}$ und eine Familie $(\lambda_j)_{j\in J}$ von Skalaren, so dass $v_k = \sum_{j\in J} \lambda_j \cdot v_j$.*

(iii) *Es existiert eine Teilmenge $J \subsetneq I$, so dass $\mathrm{Span}_K(\{v_i \mid i \in I\}) = \mathrm{Span}_K(\{v_i \mid i \in J\})$.*

Beweis. Wir beweisen die Implikationen (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

Zu (i) $\Rightarrow$ (ii). Nach Definition gibt es eine endliche Teilmenge $J \subset I$ und eine Familie von Skalaren $(\lambda_j)_{j\in J}$, die nicht alle null sind, so dass $\sum_{j\in J} \lambda_j \cdot v_j = 0$. Sei $k \in J$ ein Index mit $\lambda_k \neq 0$. Dann gilt

$$v_k = \sum_{j\in J\setminus\{k\}} \frac{-\lambda_j}{\lambda_k} \cdot v_j.$$

Zu (ii) $\Rightarrow$ (iii). Es seien k, J und $(\lambda_j)_{j\in J}$ wie in (ii). Wir zeigen, dass

$$\mathrm{Span}_K(\{v_i \mid i \in I\}) = \mathrm{Span}_K(\{v_i \mid i \in I \setminus \{k\}\}).$$

Zur Erinnerung ist $\mathrm{Span}_K(E)$ der Durchschnitt aller Untervektorräume von V, die E enthalten. Die Inklusion $\supset$ ist klar. Um die Inklusion $\subset$ zu überprüfen, ist also zu zeigen, dass folgendes für alle Untervektorräume U gilt:

$$\{v_i \mid i \in I \setminus \{k\}\} \subset U \implies v_k \in U.$$

Nach Voraussetzung ist $v_k = \sum_{j\in J} \lambda_j \cdot v_j$. Da $J \subset I \setminus \{k\}$, jedes v_j mit $j \in J$ liegt in U. Daraus folgt, dass $v_k \in U$.

Zu (iii) $\Rightarrow$ (i). Sei $k \in I\setminus J$. Da $v_k \in \mathrm{Span}_K(\{v_i | i \in J\})$ kann man nach Proposition 3.2.18 schreiben:

$$v_k = \sum_{s\in S} \lambda_s \cdot v_s,$$

wobei $S \subset J$ eine endliche Teilmenge ist und die λ_s Skalare sind. Dann gilt

$$(-1) \cdot v_k + \sum_{s\in S} \lambda_s \cdot v_s = 0.$$

Da $-1 \neq 0$ ist diese Linearkombination nicht trivial, also ist $(v_i)_{i\in I}$ linear abhängig. □

Um eine Charakterisierung der linearen *Un*abhängigkeit zu erhalten, brauchen wir eine wichtige Konstruktion.

Konstruktion 3.3.8. Sei $F = (v_i)_{i\in I}$ eine Familie von Vektoren in einem Vektorraum V über K. Diese Familie induziert eine Abbildung

$$\begin{aligned} \varphi_F \colon K^{(I)} &\to V, \\ (\lambda_i)_{i\in I} &\mapsto \sum_{i\in I} \lambda_i \cdot v_i. \end{aligned}$$

Dabei ist $K^{(I)}$ die Menge aller Abbildungen $I \to K$, die außerhalb einer endlichen Teilmenge von I null sind (siehe Beispiel 3.2.12). Obwohl I eine unendliche Menge sein könnte, ist die obige Summe $\sum_{i\in I} \lambda_i \cdot v_i$ sinnvoll, da nur endlich viele Summanden nicht null sind. Genauer könnte man schreiben:

$$\varphi_F((\lambda_i)_{i\in I}) = \sum_{i\in\{i\in I \mid \lambda_i\neq 0\}} \lambda_i \cdot v_i.$$

Man bemerkt, dass man die Familie F aus der Abbildung φ_F zurückbekommen kann, da $v_i = \varphi_F(e_i)$.

Proposition 3.3.9 (Charakterisierung der linearen Unabhängigkeit)**.** *Sei $F = (v_i)_{i\in I}$ eine Familie in einem K-Vektorraum V. Folgende Aussagen sind äquivalent:*

(i) *F ist linear unabhängig.*

(ii) *Die von F induzierte Abbildung $\varphi_F\colon K^{(I)} \to V$ ist injektiv.*

Beweis. Zu (i) $\Rightarrow$ (ii). Es seien $(\lambda_i)_{i\in I}$ und $(\mu_i)_{i\in I}$ zwei Elemente von $K^{(I)}$ mit

$$\sum_{i\in I} \lambda_i \cdot v_i = \sum_{i\in I} \mu_i \cdot v_i.$$

Diese Gleichung bedeutet, dass

$$\sum_{i\in J} \lambda_i \cdot v_i = \sum_{i\in J} \mu_i \cdot v_i,$$

wobei $J \subset I$ eine endliche Teilmenge ist, so dass $\lambda_i = \mu_i = 0$ für alle $i \in I \setminus J$. Daraus folgt:

$$\sum_{i\in J} (\lambda_i - \mu_i) \cdot v_i = 0.$$

Aus der linearen Unabhängigkeit von $(v_i)_{i\in I}$ folgt jetzt $\lambda_i - \mu_i = 0$ für alle $i \in J$. Also gilt $\lambda_i = \mu_i$ für alle $i \in I$, d.h., $(\lambda_i)_{i\in I} = (\mu_i)_{i\in I}$.

Zu (ii) $\Rightarrow$ (i). Sei $J \subset I$ eine endliche Teilmenge und $\sum_{j\in J} \lambda_j \cdot v_j$ eine Linearkombination, die gleich Null ist. Zu zeigen ist, dass $\lambda_j = 0$ für alle $j \in J$. Man kann die Familie $(\lambda_j)_{j\in J}$ zu einer Familie $(\lambda_i)_{i\in I} \in K^{(I)}$ fortsetzen, indem man $\lambda_i = 0$ setzt, falls $i \in I \setminus J$. Die Abbildung $\varphi_F\colon K^{(I)} \to V$ bildet dann $(\lambda_i)_{i\in I}$ auf 0 ab. Sie bildet auch die Nullfamilie $(0)_{i\in I}$ auf 0 ab. Aus der Injektivität von φ_F folgt, dass $(\lambda_i)_{i\in I} = (0)_{i\in I}$. □

3.3.2 Basen

Definition 3.3.10 (erzeugende Familie)**.** Sei V ein Vektorraum über K. Eine Familie $(v_i)_{i\in I}$ in V heißt *erzeugend*, wenn $\{v_i \mid i \in I\}$ ein Erzeugendensystem ist. Man sagt auch, dass die Familie $(v_i)_{i\in I}$ selbst ein Erzeugendensystem ist.

Definition 3.3.11 (Basis)**.** Sei V ein Vektorraum über K. Eine *Basis* von V ist eine erzeugende linear unabhängige Familie in V.

Beispiel 3.3.12. Die Familie der Standardeinheitsvektoren $(e_1, \ldots, e_n)$ ist eine Basis von K^n: Sie ist erzeugend (Beispiel 3.2.19(ii)) und linear unabhängig (Beispiel 3.3.4). Diese Basis heißt *Standardbasis* von K^n.

Beispiel 3.3.13. Sei I eine beliebige Menge. Die Familie der Standardeinheitsvektoren $(e_i)_{i\in I}$ in K^I ist linear unabhängig (Beispiel 3.3.4) und erzeugt den Untervektorraum $K^{(I)}$ (Beispiel 3.2.19(iv)). Sie ist also eine Basis von $K^{(I)}$.

Bemerkung 3.3.14 (Unabhängigkeit der Reihenfolge)**.** Sei $(v_i)_{i\in I}$ eine Basis von V und $\sigma\colon I \to I$ eine Permutation von I (Definition 2.2.5). Dann ist $(v_{\sigma(i)})_{i\in I}$ wieder eine Basis von V. Insbesondere, für alle $\sigma \in S_n$, ist die Familie $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$ eine Basis von K^n.

Beispiel 3.3.15.

(i) Die folgenden Familien sind Basen von K^2: (e_2, e_1), $(-e_1, e_2)$, $(e_1, e_1 + e_2)$.

(ii) Ob die Familie $(e_1 + e_2, e_1 - e_2)$ eine Basis von K^2 ist, hängt von der Characteristik von K ab. Falls die Charakteristik von K gleich 2 ist, dann gilt $e_1 + e_2 = e_1 - e_2$, und

damit ist die Familie linear abhängig. Falls die Charakteristik von K nicht gleich 2 ist, also falls $2 \neq 0$ in K, dann existiert $1/2$ in K, und es gilt

$$e_1 = \frac{1}{2}((e_1 + e_2) + (e_1 - e_2)) \quad \text{und} \quad e_2 = \frac{1}{2}((e_1 + e_2) - (e_1 - e_2)).$$

Dies zeigt, dass die Familie $(e_1 + e_2, e_1 - e_2)$ erzeugend ist. Sie ist auch linear unabhängig, denn: Sei $\lambda, \mu \in K$ mit

$$\lambda(e_1 + e_2) + \mu(e_1 - e_2) = 0, \quad \text{d.h.,} \quad (\lambda + \mu)e_1 + (\lambda - \mu)e_2 = 0.$$

Nach der linearen Unabhängigkeit von (e_1, e_2) gilt $\lambda + \mu = 0$ und $\lambda - \mu = 0$. Aus der zweiten Gleichung folgt $\lambda = \mu$, und aus der ersten $2\lambda = 0$. Da $2 \neq 0$ in K, folgt $\lambda = 0$.

(iii) Die folgenden Familien sind Basen von K^3: (e_2, e_3, e_1), $(e_1, e_1 + e_2, e_1 + e_2 + e_3)$, $(e_1 + e_2, e_1 + e_3, e_1 + e_2 + e_3)$.

(iv) Die Familie $(e_1 + e_2, e_1 + e_3, e_2 + e_3)$ ist genau dann eine Basis von K^3, wenn die Charakteristik von K nicht gleich 2 ist.

Proposition 3.3.16. *Sei $(v_i)_{i\in I}$ eine Basis von einem K-Vektorraums V. Zu jedem Vektor $v \in V$ gibt es eine eindeutige Familie $(\lambda_i)_{i\in I}$ von Skalaren, die null außerhalb einer endlichen Teilmenge von I sind, so dass $v = \sum_{i\in I} \lambda_i \cdot v_i$.*

Beweis. Da $v \in \mathrm{Span}_K(\{v_i \mid i \in I\})$, eine solche Familie $(\lambda_i)_{i\in I}$ existiert nach Proposition 3.2.18. Die Eindeutigkeit der Familie folgt aus Proposition 3.3.9. □

Definition 3.3.17 (Koordinatenvektor)**.** Sei $B = (v_i)_{i\in I}$ eine Basis eines K-Vektorraums V und sei $v \in V$. Die Familie $(\lambda_i)_{i\in I} \in K^{(I)}$ aus Proposition 3.3.16 heißt der *Koordinatenvektor* von v bzgl. der Basis B und wird mit $[v]_B$ bezeichnet.

Bemerkung 3.3.18. Sei $F = (v_i)_{i\in I}$ eine Familie von Vektoren in V. Ob diese Familie erzeugend, linear unabhängig oder eine Basis ist, lässt sich durch die induzierte Abbildung

$$\begin{aligned} \varphi_F \colon K^{(I)} &\to V, \\ (\lambda_i)_{i\in I} &\mapsto \sum_{i\in I} \lambda_i \cdot v_i, \end{aligned}$$

aus Konstruktion 3.3.8 durchsichtig ausdrücken. Es gilt nämlich:

(i) F ist genau dann erzeugend, wenn φ_F surjektiv ist (nach Proposition 3.2.18).

(ii) F ist genau dann linear unabhängig, wenn φ_F injektiv ist (nach Proposition 3.3.9).

(iii) F ist genau dann eine Basis, wenn φ_F bijektiv ist (nach (i) und (ii)).

Falls F eine Basis ist, ist der Koordinatenvektor von einem $v \in V$ bzgl. F gleich $\varphi_F^{-1}(v) \in K^{(I)}$.

Proposition 3.3.19 (Charakterisierung von Basen)**.** *Sei V ein Vektorraum über K und $(v_i)_{i\in I}$ eine Familie in V. Dann sind die folgenden Aussagen äquivalent:*

(i) *$(v_i)_{i\in I}$ ist eine Basis von V.*

(ii) *$(v_i)_{i\in I}$ ist eine* maximale linear unabhängige Familie*, d.h., sie ist linear unabhängig, und für jede linear unabhängige Familie $(v_j)_{j\in J}$ mit $I \subset J$ gilt $I = J$.*

(iii) *$(v_i)_{i\in I}$ ist eine* minimale erzeugende Familie*, d.h., sie ist erzeugend, und für jede erzeugende Familie $(v_j)_{j\in J}$ mit $J \subset I$ gilt $I = J$.*

Beweis. Wir beweisen die beide Äquivalenzen (i) $\Leftrightarrow$ (ii) und (i) $\Leftrightarrow$ (iii).

Zu (i) $\Rightarrow$ (ii). Sei $(v_i)_{i\in I}$ eine Basis und $(v_j)_{j\in J}$ eine Familie mit $I \subset J$. Falls $I \neq J$, dann ist $(v_j)_{j\in J}$ linear abhängig nach Proposition 3.3.7 (iii) $\Rightarrow$ (i).

Zu (ii) $\Rightarrow$ (i). Zu zeigen ist, dass die Familie $(v_i)_{i\in I}$ erzeugend ist. Sei $v \in V$ ein beliebiger Vektor. Wenn man v zu der Familie $(v_i)_{i\in I}$ hinzufügt, erhält man aufgrund der Maximalität von $(v_i)_{i\in I}$ eine Familie, die linear abhängig ist. Es gibt also eine nicht-triviale Linearkombination der Vektoren v_i und v, die gleich Null ist. Da $(v_i)_{i\in I}$ linear unabhängig ist, muss der Koeffizient von v in einer solchen Linearkombination nicht null sein. Daraus folgt, dass v eine Linearkombination der Vektoren v_i ist, wie gewünscht.

Zu (i) $\Rightarrow$ (iii). Sei $(v_i)_{i\in I}$ eine Basis und $J \subset I$ eine Teilmenge mit $J \neq I$. Zu zeigen ist, dass die Familie $(v_j)_{j\in J}$ nicht erzeugend ist. Aber wenn sie erzeugend wäre, dann wäre $(v_i)_{i\in I}$ linear abhängig sein, nach Proposition 3.3.7 (iii) $\Rightarrow$ (i).

Zu (iii) $\Rightarrow$ (i). Man hat zu zeigen, dass die Familie $(v_i)_{i\in I}$ linear unabhängig ist. Wenn nicht, dann existiert nach Proposition 3.3.7 (i) $\Rightarrow$ (iii) eine Teilmenge $J \subsetneq I$, so dass die Familie $(v_j)_{j\in J}$ erzeugend ist. Aber das steht im Widerspruch zur Minimalität von $(v_i)_{i\in I}$. □

Der folgende Satz ist einer der wichtigsten Struktursätze für Vektorräume. Wir werden ihn sehr häufig verwenden.

Satz 3.3.20 (Basisergänzungssatz)**.** *Sei V ein K-Vektorraum.*

(i) *Sei $(v_i)_{i\in I}$ eine erzeugende Familie in V und sei $J \subset I$ eine Teilmenge, so dass $(v_i)_{i\in J}$ linear unabhängig ist. Dann existiert eine Menge L mit $J \subset L \subset I$, so dass $(v_i)_{i\in L}$ eine Basis von V ist.*

Insbesondere:

(ii) *Jede erzeugende Familie $(v_i)_{i\in I}$ in V kann zu einer Basis eingeschränkt werden, d.h., es existiert eine Teilmenge $J \subset I$, so dass $(v_i)_{i\in J}$ eine Basis ist.*

(iii) *Jede linear unabhängige Familie $(v_i)_{i\in I}$ in V kann zu einer Basis ergänzt werden, d.h., es existiert eine Indexmenge $J \supset I$ und Vektoren v_j für $j \in J \setminus I$, so dass $(v_j)_{j\in J}$ eine Basis ist.*

Beweis. Aussage (ii) ist der Sonderfall von (i) mit $J = \varnothing$. Aussage (iii) folgt aus (i), indem wir zuerst die gegebene Familie $(v_i)_{i\in I}$ zu einer erzeugenden Familie ergänzen, z.B. zu einer Familie, die alle Vektoren aus V enthält.

Wir beweisen (i) zunächst im Spezialfall, dass I endlich ist, da der Beweis in diesem Fall einfacher ist. Wir beweisen die Existenz von L durch vollständige Induktion über die Mächtigkeit von $I \setminus J$ (Korollar 1.2.23). Falls $(v_i)_{i\in J}$ bereits erzeugend ist, kann man $L = J$ nehmen. Andernfalls gibt es einen Index $k \in I \setminus J$, so dass $v_k \notin \mathrm{Span}_K(\{v_i \mid i \in J\})$. Dann ist die Familie $(v_i)_{i\in J\cup\{k\}}$ linear unabhängig, denn: Sei

$$\sum_{i\in J\cup\{k\}} \lambda_i \cdot v_i = 0$$

mit $\lambda_i \in K$. Es gilt $\lambda_k = 0$, sonst wäre

$$v_k = -\sum_{i\in J} \frac{\lambda_i}{\lambda_k} \cdot v_i$$

und damit würde v_k in $\mathrm{Span}_K(\{v_i \mid i \in J\})$ liegen. Alle anderen λ_i sind dann auch null, da $(v_i)_{i\in J}$ linear unabhängig ist. Die Mächtigkeit von $I \setminus (J \cup \{k\})$ ist kleiner als die von $I \setminus J$. Nach der Induktionsvoraussetzung gibt es eine Menge L mit $J \cup \{k\} \subset L \subset I$, so dass $(v_i)_{i\in L}$ eine Basis von V ist. Damit ist (i) im Fall einer endlichen Menge I bewiesen.

Wir beweisen jetzt den allgemeinen Fall. Sei

$$\mathcal{A} := \{L \mid J \subset L \subset I \text{ und } (v_i)_{i\in L} \text{ ist linear unabhängig}\} \subset \mathcal{P}(I).$$

Wir betrachten $\mathcal{A}$ als partiell geordnete Menge bezüglich der Inklusionsrelation $\subset$. Wir behaupten, dass $\mathcal{A}$ ein maximales Element $L_{\max}$ besitzt, und dass die Familie $(v_i)_{i\in L_{\max}}$ eine Basis ist. Um zu zeigen, dass $\mathcal{A}$ ein maximales Element besitzt, verwenden wir das Zornsche Lemma 1.4.22: Es genügt zu zeigen, dass jede Kette $\mathcal{B} \subset \mathcal{A}$ eine obere Schranke besitzt. Falls $\mathcal{B} = \varnothing$, dann ist J eine obere Schranke von $\mathcal{B}$. Andernfalls betrachten wir die Vereinigung $B = \bigcup_{L\in\mathcal{B}} L$. Da $\mathcal{B}$ total geordnet und nicht leer ist, gibt es zu jeder endlichen Teilmenge $E \subset B$ ein $L \in \mathcal{B}$ mit $E \subset L$ (das kann mann leicht durch Induktion über $|E|$ nachprüfen). Für jede endliche Teilmenge $E \subset B$ ist also die Familie $(v_i)_{i\in E}$ linear unabhängig, und daher ist die ganze Familie $(v_i)_{i\in B}$ linear unabhängig. Das heißt: Es gilt $B \in \mathcal{A}$, und damit ist B eine obere Schranke von $\mathcal{B}$.

Nach dem Zornschen Lemma existiert also ein maximales Element $L_{\max} \in \mathcal{A}$. Es bleibt zu zeigen, dass $\{v_i \mid i \in L_{\max}\}$ ein Erzeugendensystem ist. Nach Voraussetzung ist $\{v_i \mid i \in I\}$ ein Erzeugendensystem. Deswegen genügt es zu zeigen, dass $v_k \in \mathrm{Span}_K(\{v_i \mid i \in L_{\max}\})$ für jedes $k \in I \setminus L_{\max}$. Da $L_{\max}$ maximal in $\mathcal{A}$ ist, ist die Familie $(v_i)_{i\in L_{\max}\cup\{k\}}$ linear abhängig: Es gibt Skalare λ_i, $i \in L_{\max} \cup \{k\}$, die nicht alle null sind, so dass

$$\sum_{i\in L_{\max}\cup\{k\}} \lambda_i \cdot v_i = 0.$$

Es muss eigentlich $\lambda_k \neq 0$ gelten, da $(v_i)_{i\in L_{\max}}$ linear unabhängig ist. Wir dürfen also durch λ_k dividieren, und erhalten

$$v_k = -\sum_{i\in L_{\max}} \frac{\lambda_i}{\lambda_k} \cdot v_i.$$

Also gilt $v_k \in \mathrm{Span}_K(\{v_i \mid i \in L_{\max}\})$, wie gewünscht. □

Bemerkung 3.3.21. Die Beweise des obigen Satzes im Fall einer endlichen Teilmenge I und im allgemeinen Fall waren sehr ähnlich. Der Unterschied war nur, dass wir im allgemeinen Fall das Induktionsprinzip durch das Zornsche Lemma ersetzen mussten. Das ist eigentlich ein typisches Anwendung des Zornschen Lemmas: Es ist oft der Fall, dass Aussagen, die bei endlichen Mengen durch Induktion bewiesen werden können, auch bei unendlichen Mengen mit dem Zornschen Lemma bewiesen werden können.

Beispiel 3.3.22. Die Familie $(e_1, e_2, e_3, e_1+e_2+e_3)$ in K^3 ist erzeugend, und die Teilfamilie $(e_1, e_1 + e_2 + e_3)$ ist linear unabhängig. Nach Satz 3.3.20(i) gibt es eine Basis zwischen den beiden Familien. In diesem Fall sind beide Familien $(e_1, e_2, e_1+e_2+e_3)$ und $(e_1, e_3, e_1+e_2+e_3)$ Basen von K^3.

Korollar 3.3.23 (Existenz von Basen)**.**

(i) *Jeder K-Vektorraum besitzt eine Basis.*

(ii) *Jeder endlich erzeugte K-Vektorraum besitzt eine endliche Basis.*

Beweis. Sei V ein K-Vektorraum und sei $(v_i)_{i\in I}$ eine erzeugende Familie, z.B. $(v)_{v\in V}$. Nach Satz 3.3.20(ii), kann diese Familie zu einer Basis eingeschränkt werden. Falls V endlich erzeugt ist, gibt es eine solche Familie mit einem endlichen I, and damit erhalten wir eine endliche Basis. □

Lemma 3.3.24 (Austauschlemma)**.** *Sei V ein K-Vektorraum, $(v_i)_{i\in I}$ eine Basis von V und $w \in V$. Sei $I' \subset I$ eine Teilmenge, so dass $w \notin \mathrm{Span}_K(\{v_i \mid i \in I'\})$; zum Beispiel, $I' = \varnothing$ und $w \neq 0$. Dann existiert ein Index $k \in I \setminus I'$, so dass die Familie, die sich aus $(v_i)_{i\in I}$ ergibt, wenn v_k gegen w ausgetauscht wird, wieder eine Basis von V ist.*

Beweis. Da $(v_i)_{i\in I}$ eine Basis ist, kann man schreiben

$$w = \sum_{i\in I} \lambda_i \cdot v_i, \tag{3.3.25}$$

wobei die Skalare $\lambda_i \in K$ alle null sind, außer endlich viele. Es gibt dann ein $k \in I \setminus I'$ mit $\lambda_k \neq 0$, sonst würde w in $\mathrm{Span}_K(\{v_i \mid i \in I'\})$ liegen. Sei $(\tilde{v}_i)_{i\in I}$ die Familie mit

$$\tilde{v}_i = \begin{cases} v_i, & \text{falls } i \neq k, \\ w, & \text{falls } i = k. \end{cases}$$

Die Famile $(\tilde{v}_i)_{i\in I}$ ist erzeugend, da

$$v_k = \frac{1}{\lambda_k}\left(w - \sum_{i\in I\setminus\{k\}} \lambda_i \cdot v_i\right) \in \mathrm{Span}_K(\{\tilde{v}_i \mid i \in I\}).$$

Die Familie $(\tilde{v}_i)_{i\in I}$ ist linear unabhängig, denn: Sei

$$\sum_{i\in I} \mu_i \cdot \tilde{v}_i = 0 \tag{3.3.26}$$

mit $(\mu_i)_{i\in I} \in K^{(I)}$. Aus (3.3.25) und (3.3.26) folgt:

$$0 = \mu_k \cdot w + \sum_{i\in I\setminus\{k\}} \mu_i \cdot v_i = (\mu_k \cdot \lambda_k) \cdot v_k + \sum_{i\in I\setminus\{k\}} (\mu_k \cdot \lambda_i + \mu_i) \cdot v_i.$$

Da $(v_i)_{i\in I}$ linear unabhängig ist, sind $\mu_k \cdot \lambda_k$ und $\mu_k \cdot \lambda_i + \mu_i$ mit $i \in I \setminus \{k\}$ alle null. Aus $\lambda_k \neq 0$ folgt jetzt, dass $\mu_k = 0$, und daher auch dass $\mu_i = 0$ für alle $i \in I \setminus \{k\}$. □

Satz 3.3.27 (alle Basen haben dieselbe Länge, endlicher Fall)**.** *Sei V ein endlich erzeugter K-Vektorraum.*

(i) *Ist $(v_i)_{i\in I}$ eine Basis von V, so ist die Menge I endlich.*

(ii) *Sind $(v_1, \ldots, v_n)$ und $(w_1, \ldots, w_m)$ zwei Basen von V, so gilt $n = m$.*

Beweis. Da V endlich erzeugt ist, existiert nach Korollar 3.3.23 eine Basis $(b_1, \ldots, b_n)$ von V mit $n \in \mathbb{N}$. Sei $(v_i)_{i\in I}$ eine beliebige Basis von V. Wir behaupten, dass es paarweise verschiedene Indizes $i_1, \ldots, i_n \in I$ gibt, so dass die Familie $(\tilde{v}_i)_{i\in I}$ mit

$$\tilde{v}_i = \begin{cases} b_k, & \text{falls } i = i_k \text{ mit } k \in \{1, \ldots, n\}, \\ v_i & \text{andernfalls,} \end{cases}$$

eine Basis von V ist. Dazu verwenden wir n-mal das Austauschlemma: Sind die Indizes $i_1, \ldots, i_{k-1}$ schon gefunden, wenden wir das Austauschlemma mit $I' = \{i_1, \ldots, i_{k-1}\}$ und $w = b_k$ an, um das Index i_k zu erhalten. Da $(b_1, \ldots, b_n)$ eine *maximale* linear unabhängige Familie ist (Proposition 3.3.19), folgt daraus, dass $I = \{i_1, \ldots, i_n\}$. Insbesondere ist I endlich der Mächtigkeit n, was beide (i) und (ii) beweist. □

Bemerkung 3.3.28. Ist I eine Menge, so hat $K^{(I)}$ die Basis $(e_i)_{i\in I}$ (Beispiel 3.3.13). Nach Satz 3.3.27(i), falls I unendlich ist, dann ist $K^{(I)}$ nicht endlich erzeugt. Da die Familie $(e_i)_{i\in I}$ zu einer Basis von K^I ergänzt werden kann (Satz 3.3.20(iii)), ist K^I ebenfalls nicht endlich erzeugt.

Man kann Satz 3.3.27(ii) auf beliebige Vektorräume verallgemeinern, mithilfe des Begriffs der Gleichmächtigkeit (Definition 1.3.28):

Satz 3.3.29 (alle Basen haben dieselbe Länge, allgemeiner Fall)**.** *Seien $(v_i)_{i\in I}$ und $(w_j)_{j\in J}$ zwei Basen eines K-Vektorraums V. Dann sind I und J gleichmächtig.*

Beweis. Falls V endlich erzeugt ist, folgt dies bereits aus dem Satz 3.3.27. Wir dürfen deshalb annehmen, dass I und J unendlich sind. Jedes v_i lässt sich eindeutig als Linearkombination der Vektoren w_j schreiben. Sei $J_i \subset J$ die endliche Teilmenge aller Indizes j, so dass der Koeffizient von w_j in dieser Linearkombination nicht null ist. Es gilt dann $v_i \in \mathrm{Span}_K(\{w_j \mid j \in J_i\})$ und daher

$$V = \mathrm{Span}_K(\{v_i \mid i \in I\}) \subset \mathrm{Span}_K\left(\left\{ w_j \;\middle|\; j \in \bigcup_{i\in I} J_i \right\}\right).$$

Da $(w_j)_{j\in J}$ eine Basis ist, und damit eine *minimale* erzeugende Familie (Proposition 3.3.19), gilt $\bigcup_{i\in I} J_i = J$. Aus dem Satz 1.3.36 folgt, dass $|J| \leq |I|$. Auf symmetrische Weise folgt $|I| \leq |J|$. Nach dem Satz von Cantor–Bernstein–Schröder (Satz 1.3.33) sind also I und J gleichmächtig. □

3.3.3 Dimension

Definition 3.3.30 (Dimension, endlich-dimensional, unendlich-dimensional)**.** Sei V ein K-Vektorraum. Die *Dimension* von V,

$$\dim_K(V) \in \mathbb{N} \cup \{\infty\},$$

wird wie folgt definiert:

- Falls V endlich erzeugt ist, ist $\dim_K(V) \in \mathbb{N}$ die Länge irgendeiner Basis von V, die nach Satz 3.3.27 eine wohldefinierte natürliche Zahl ist. In diesem Fall sagt man auch, dass V *endlich-dimensional* ist.
- Falls V nicht endlich erzeugt ist, setzt man $\dim_K(V) := \infty$. In diesem Fall heißt V *unendlich-dimensional*.

Bemerkung 3.3.31 (Dimension als Mächtigkeit)**.** Die Dimension eines unendlich-dimensionalen K-Vektorraums V kann genauer als die Mächtigkeit irgendeiner Basis von V definiert werden, die nach Satz 3.3.29 wohldefiniert ist.

Beispiel 3.3.32.

(i) Es gilt $\dim_K(K^n) = n$, da $(e_1, \ldots, e_n)$ eine Basis von K^n ist. Insbesondere gilt $\dim_K(K) = 1$.

(ii) Es gilt $\dim_K(V) = 0$ genau dann, wenn $V = \{0\}$ (die leere Familie ist eine Basis des trivialen Vektorraums $\{0\}$).

(iii) Ist $(v_1, \ldots, v_n)$ eine linear unabhängige Familie in einem K-Vektorraum V, so ist die Dimension des Untervektorraums $\mathrm{Span}_K(\{v_1, \ldots, v_n\})$ gleich n.

(iv) Sei I eine Menge. Der K-Vektorraum $K^{(I)}$ hat dann die Basis $(e_i)_{i\in I}$. Falls I unendlich ist, ist also $\dim_K(K^{(I)}) = \infty$. Im Sinne der Bemerkung 3.3.31 ist genauer die Dimension von $K^{(I)}$ gleich der Mächtigkeit von I.

Bemerkung 3.3.33 (Abhängigkeit vom Grundkörper)**.** Wenn $K \subset L$ eine Körpererweiterung ist, kann jeder L-Vektorraum V als K-Vektorraum betrachtet werden (siehe Bemerkung 3.2.15). Es gilt dann $\dim_L(V) \leq \dim_K(V)$, da jede L-Basis von V auch K-linear unabhängig ist, und daher zu einer K-Basis ergänzt werden kann (nach Satz 3.3.20(iii)). Im Allgemeinen ist aber $\dim_L(V) \neq \dim_V(K)$. Zum Beispiel, $\dim_{\mathbb{C}}(\mathbb{C}) = 1$ aber $\dim_{\mathbb{R}}(\mathbb{C}) = 2$: $(1, i)$ ist eine Basis von $\mathbb{C}$ als $\mathbb{R}$-Vektorraum.

Proposition 3.3.34 (Basen in endlich-dimensionalen Vektorräumen)**.** *Sei V ein Vektorraum über K mit $\dim_K(V) = n < \infty$, und sei $B = (v_1, \dots, v_n)$ eine Familie von n Vektoren aus V. Folgende Aussagen sind äquivalent:*

(i) *B ist eine Basis.*

(ii) *B ist erzeugend.*

(iii) *B ist linear unabhängig.*

Beweis. Nach Definition gelten (i) $\Rightarrow$ (ii) und (i) $\Rightarrow$ (iii). Falls B erzeugend bzw. linear unabhängig ist, dann kann B nach Satz 3.3.20 zu einer Basis B' eingeschränkt bzw. ergänzt werden. Die Basis B' muss nach Satz 3.3.27 aus n Vektoren bestehen, also gilt $B = B'$. Insbesondere war B bereits eine Basis. □

Proposition 3.3.35 (Dimension von Untervektorräumen)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann ist U auch endlich-dimensional und $\dim_K(U) \leq \dim_K(V)$. Falls $U \neq V$ gilt eigentlich $\dim_K(U) < \dim_K(V)$.*

Beweis. Nach Korollar 3.3.23 besitzt U eine Basis $(u_i)_{i \in I}$. Die Familie $(u_i)_{i \in I}$ ist insbesondere linear unabhängig, und kann nach Satz 3.3.20(iii) zu einer Basis von V ergänzt werden, die endlich ist nach Satz 3.3.27(i). Dies zeigt, dass I endlich ist und dass $\dim_K(U) \leq \dim_K(V)$.

Zur letzten Aussage beweisen wir die Kontraposition. Falls $\dim_K(V) = \dim_K(U)$, dann muss $(u_i)_{i \in I}$ bereits eine Basis von V sein, und daher muss $U = V$ sein. □

Definition 3.3.36 (Summe von Untervektorräumen)**.** Sei V ein K-Vektorraum und seien $U, W \subset V$ zwei Untervektorräume. Die *Summe* von U und W ist der Untervektorraum

$$U + W := \mathrm{Span}_K(U \cup W) \subset V.$$

Nach Proposition 3.2.18 gilt

$$U + W = \{u + w \mid u \in U \text{ und } w \in W\}.$$

Satz 3.3.37 (Dimensionsformel für Untervektorräume)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und seien $U, W \subset V$ Untervektorräume. Dann gilt*

$$\dim_K(U + W) = \dim_K(U) + \dim_K(W) - \dim_K(U \cap W).$$

Es ist hilfreich, diese Dimensionsformel im Fall $K = \mathbb{R}$ und $V = \mathbb{R}^3$ explizit zu untersuchen. Untervektorraüme von $\mathbb{R}^3$ sind $\{0\}$, Ursprungsgeraden, Ursprungsebenen, und $\mathbb{R}^3$ selbst. Seien zum Beispiel $U, W \subset \mathbb{R}^3$ zwei Ursprungsebenen. Falls $U \neq W$, dann ist $U \cap W$ eine Gerade und ist $U + W = \mathbb{R}^3$, und die Dimensionsformel lautet $3 = 2 + 2 - 1$. Falls $U = W$, dann gilt $U = W = U \cap W = U + W$, und die Dimensionsformel lautet $2 = 2 + 2 - 2$. Es ist natürlich unmöglich, dass $U \cap W = \{0\}$; das lässt sich auch aus der Dimensionsformel ableiten, da $2 + 2 - 0 = 4$ aber $\dim_{\mathbb{R}}(U + W) \leq \dim_{\mathbb{R}}(\mathbb{R}^3) = 3$. In höherer Dimension gibt es jedoch Ebenen, die nur in einem Punkt treffen, z.B. die Ebenen $\mathrm{Span}_{\mathbb{R}}(\{e_1, e_2\})$ und $\mathrm{Span}_{\mathbb{R}}(\{e_3, e_4\})$ in $\mathbb{R}^4$.

Beweis. Es folgt aus Proposition 3.3.35, dass U, V, $U + V$, und $U \cap V$ endlich-dimensional sind. Sei $(v_1, \dots, v_n)$ eine Basis von $U \cap W$. Nach Satz 3.3.20(iii) können wir diese Basis zu einer Basis $(v_1, \dots, v_n, u_1, \dots, u_p)$ von U und zu einer Basis $(v_1, \dots, v_n, w_1, \dots, w_q)$ von W ergänzen. Dann gilt $\dim_K(U) = n + p$, $\dim_K(W) = n + q$, $\dim_K(U \cap W) = n$. Wir müssen also zeigen, dass

$$\dim_K(U + W) = (n + p) + (n + q) - n = n + p + q.$$

Dazu zeigen wir, dass die Familie

$$(v_1, \dots, v_n, u_1, \dots, u_p, w_1, \dots, w_q)$$

eine Basis von $U + W$ ist. Sie erzeugt $U + W$, weil sie Basen von U sowie W enthält. Es bleibt die lineare Unabhängigkeit nachzuprüfen. Sei also

$$\underbrace{\sum_{i=1}^{n} \lambda_i \cdot v_i}_{\in U \cap W} + \underbrace{\sum_{j=1}^{p} \mu_j \cdot u_j}_{\in U} + \underbrace{\sum_{k=1}^{q} \nu_k \cdot w_k}_{\in W} = 0 \tag{3.3.38}$$

mit $\lambda_i, \mu_j, \nu_k \in K$. Aus dieser Gleichung folgt, dass die Summe $\sum_{j=1}^{p} \mu_j \cdot u_j$ in $U \cap W$ liegt. Da $(v_1, \ldots, v_n)$ eine erzeugende Familie von $U \cap V$ ist, gibt es Skalare $\lambda'_i \in K$, so dass

$$\sum_{j=1}^{p} \mu_j \cdot u_j = \sum_{i=1}^{n} \lambda'_i \cdot v_i.$$

Aus der linearen Unabhängigkeit von $(v_1, \ldots, v_n, u_1, \ldots, u_p)$ folgt insbesondere, dass $\mu_1 = \cdots = \mu_p = 0$. Wenn wir die Rollen von U und W vertauschen, erhalten wir gleichfalls $\nu_1 = \cdots = \nu_q = 0$. Von der Gleichung (3.3.38) bleibt übrig

$$\sum_{i=1}^{n} \lambda_i \cdot v_i = 0.$$

Da $(v_1, \ldots, v_n)$ linear unabhängig ist, folgt schließlich $\lambda_1 = \cdots = \lambda_n = 0$. □

Definition 3.3.39 (komplementäre Untervektorräume)**.** Sei V ein K-Vektorraum. Zwei Untervektorräume $U, W \subset V$ heißen *komplementär*, wenn folgende Bedingungen gelten:

$$U + W = V \quad \text{und} \quad U \cap W = \{0\}.$$

Man sagt auch, dass W zu U in V komplementär ist, oder dass W ein *direktes Komplement* von U in V ist.

Beispiel 3.3.40. Sei $(v_i)_{i \in I}$ eine Basis eines K-Vektorraums V und seien $I', I'' \subset I$ disjunkte Teilmengen mit $I = I' \cup I''$. Dann sind die Untervektorräume $\mathrm{Span}_K(\{v_i \,|\, i \in I'\})$ und $\mathrm{Span}_K(\{v_i | i \in I''\})$ komplementär. Beispielsweise sind $\mathrm{Span}_K(\{e_1, e_3\})$ und $\mathrm{Span}_K(\{e_2, e_4\})$ komplementäre Untervektorräume von K^4.

Proposition 3.3.41 (Charakterisierung von komplementären Untervektorräumen)**.** *Sei V ein endlich-dimensionaler K-Vektorraum, und seien $U, W \subset V$ Untervektorräume. Die folgenden Aussagen sind äquivalent:*

(i) *U und W sind komplementär.*

(ii) *$U + W = V$ und $\dim_K(V) = \dim_K(U) + \dim_K(W)$.*

(iii) *$U \cap W = \{0\}$ und $\dim_K(V) = \dim_K(U) + \dim_K(W)$.*

Beweis. Die Implikationen (i) ⇒ (ii) und (i) ⇒ (iii) folgen aus dem Satz 3.3.37 und der Definition von komplementären Untervektorräumen. Wir nehmen jetzt an, dass $\dim_K(V) = \dim_K(U) + \dim_K(W)$. Nach dem Satz 3.3.37 gilt dann

$$\dim_K(V) = \dim_K(U + W) + \dim_K(U \cap W).$$

Falls $U + W = V$, dann gilt $\dim_K(U \cap W) = 0$, d.h., $U \cap W = \{0\}$. Falls $U \cap W = \{0\}$, dann gilt $\dim_K(U + W) = \dim_K(V)$, und daher $U + W = V$ nach Proposition 3.3.35. □

Beispiel 3.3.42. Eine Ursprungsgerade G und eine Ursprungsebene E in $\mathbb{R}^3$ sind genau dann komplementär, wenn $G \cap E = \{0\}$.

Proposition 3.3.43 (Existenz von komplementären Untervektorräumen)**.** *Sei V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann existiert ein Untervektorraum $W \subset V$, der komplementär zu U in V ist.*

Beweis. Sei $(v_i)_{i\in I}$ eine Basis von U. Nach Satz 3.3.20(iii) gibt es eine Menge $J \supset I$ und Vektoren $v_j \in V$ für alle $j \in J \setminus I$, so dass die ergänzte Familie $(v_j)_{j\in J}$ eine Basis von V ist. Sei dann W der von $\{v_j \mid j \in J \setminus I\}$ erzeugte Untervektorraum von V. Nach Konstruktion gilt $U + W = V$, und es bleibt zu zeigen, dass $U \cap W = \{0\}$. Jedes $v \in U \cap W$ kann als

$$v = \sum_{j\in I} \lambda_j \cdot v_j \quad \text{und} \quad v = \sum_{j\in J\setminus I} \lambda_j \cdot v_j$$

dargestellt werden. Aus der linearen Unabhängigkeit von $(v_j)_{j\in J}$ folgt, dass alle λ_j null sind, und daher dass $v = 0$. □

Bemerkung 3.3.44. Im Allgemeinen hat ein Untervektorraum $U \subset V$ viele verschiedene komplementäre Untervektorräume. Zum Beispiel, wenn $U \subset K^2$ die von e_1 aufgespannte Gerade ist, dann ist jede andere Ursprungsgerade komplementär zu U. Man darf also nicht von *dem* komplementären Untervektorraum sprechen.

Es ist möglich, je zwei Vektorräume U und W als komplementäre Untervektoräume eines größeren Vektorraum $U \oplus W$ zu betrachten:

Definition 3.3.45 (direkte Summe)**.** Seien U und W Vektorräume über K. Die *direkte Summe* $U \oplus W$ von U und W ist das kartesische Produkt $U \times W$, versehen mit den komponentenweisen Addition und Skalarmultiplikation, d.h.:

$$(u, w) + (u', w') = (u + u', w + w')$$
$$\lambda \cdot (u, w) = (\lambda u, \lambda w).$$

Man kann leicht nachprüfen, dass $U \oplus W$ mit diesen Verknüpfungen ein K-Vektorraum ist (siehe Lemma 3.1.3 für einen ähnlichen Beweis). Die direkte Summe $U \oplus W$ heißt auch das *Produkt* von U und W, und kann auch mit $U \times W$ bezeichnet werden. Nach Proposition 3.2.8 sind $U \times \{0\}$ und $\{0\} \times W$ Untervektorräume von $U \oplus W$, und es ist klar, dass sie komplementäre Untervektorräume sind. Außerdem können U und W durch die injektiven Abbildungen

$$\begin{aligned} U &\to U \oplus W, & W &\to U \oplus W, \\ u &\mapsto (u, 0) & w &\mapsto (0, w) \end{aligned}$$

mit diesen Untervektorräumen von $U \oplus W$ identifiziert werden. Das heißt, wenn wir den Vektor $u \in U$ mit dem Paar $(u, 0)$ identifizieren, dann ist die Addition bzw. die Skalarmultiplikation auf U dieselbe wie die auf dem Untervektorraum $U \times \{0\} \subset U \oplus W$. Solche Identifizierungen werden wir später mit dem Begriff des Isomorphismus präziser machen (siehe Definition 4.1.14). Es gilt insbesondere

$$\dim_K(U \oplus W) = \dim_K(U) + \dim_K(W)$$

nach Satz 3.3.37.

Proposition 3.3.46 (Dimensionsformel für Quotientenvektorräume)**.** *Sei V ein endlichdimensionaler K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann gilt*

$$\dim_K(V) = \dim_K(U) + \dim_K(V/U).$$

Beweis. Nach Satz 3.3.20(iii) gibt es eine Basis $(v_i)_{i\in I}$ von V und eine Teilmenge $J \subset I$, so dass $(v_i)_{i\in J}$ eine Basis von U ist. Es genügt zu zeigen, dass $(v_i + U)_{i\in I\setminus J}$ eine Basis von V/U ist.

- $\{v_i + U \mid i \in I \setminus J\}$ *ist ein Erzeugendensystem.* Sei $v = \sum_{i \in I} \lambda_i v_i$ ein beliebiger Vektor aus V. Die Differenz $v - \sum_{i \in I \setminus J} \lambda_i v_i$ liegt in U, und daher gilt

$$v + U = \left(\sum_{i \in I \setminus J} \lambda_i \cdot v_i \right) + U = \sum_{i \in I \setminus J} \lambda_i \cdot (v_i + U)$$

  in V/U.

- $(v_i + U)_{i \in I \setminus J}$ *ist linear unabhängig.* Es sei

$$\sum_{i \in I \setminus J} \lambda_i \cdot (v_i + U) = 0 + U$$

  mit $\lambda_i \in K$. Diese Gleichung bedeutet, dass $\sum_{i \in I \setminus J} \lambda_i v_i \in U$. Es gibt also Skalare $\lambda_i \in K$ für alle $i \in J$, so dass

$$\sum_{i \in I \setminus J} \lambda_i \cdot v_i = \sum_{i \in J} \lambda_i \cdot v_i.$$

  Aus der linearen Unabhängigkeit von $(v_i)_{i \in I}$ folgt, dass $\lambda_i = 0$ für alle $i \in I$, und insbesondere für alle $i \in I \setminus J$, wie gewünscht. □

Kapitel 4

Lineare Abbildungen

In diesem Kapitel wird ein Grundkörper K immer wieder festgelegt.

4.1 Lineare Abbildungen

Wir fangen mit der Definition an:

Definition 4.1.1 (lineare Abbildung)**.** Seien V, W zwei Vektorräume über K. Eine *lineare Abbildung*, oder genauer *K-lineare Abbildung*, von V nach W ist eine Abbildung $f\colon V \to W$ mit folgenden Eigenschaften:

(i) Für alle $v, v' \in V$ gilt
$$f(v + v') = f(v) + f(v').$$

(ii) Für alle $v \in V$ und $\lambda \in K$ gilt
$$f(\lambda \cdot v) = \lambda \cdot f(v).$$

Lineare Abbildungen heißen auch *Vektorraumhomorphismen*. Die Menge aller K-linearen Abbildungen von V nach W wird mit $\mathrm{Hom}_K(V, W)$ bezeichnet.

Bemerkung 4.1.2. Die linearen Abbildungen zwischen Vektorräumen sind die Abbildungen, die mit der Vektorraumstruktur verträglich sind. Jedes Mal, wenn wir eine Art von „Mengen mit Struktur“ einführen, wie z.B. Gruppen, Körper, Vektorräume, partiell geordnete Mengen usw., gibt es normalerweise eine entsprechende Art von Abbildungen zwischen denen, die mit dieser Struktur verträglich sind. Zum Beispiel:

- Seien G und H Gruppen. Ein *Gruppenhomomorphismus* von G nach H ist eine Abbildung $f\colon G \to H$, so dass für alle $g, g' \in G$ gilt: $f(g \cdot g') = f(g) \cdot f(g')$.
- Seien K und L Körper. Ein *Körperhomomorphismus* von K nach L ist eine Abbildung $f\colon K \to L$, die mit beiden Verknüpfungen $+$ und $\cdot$ verträglich ist und außerdem 1 auf 1 abbildet.
- Seien X und Y partiell geordnete Mengen. Eine *monotone Abbildung* (oder *Ordnungshomomorphismus*) von X nach Y ist eine Abbildung $f\colon X \to Y$, so dass für alle $x, x' \in X$ gilt: $x \leq x' \Rightarrow f(x) \leq f(y)$.

Dieses Phänomen ist der Ausgangspunkt der *Kategorientheorie*. Man beachte, dass eine K-lineare Abbildung von V nach W insbesondere ein Gruppenhomomorphismus von $(V, +)$ nach $(W, +)$ ist.

Beispiel 4.1.3 (Skalierung und Verschiebung)**.** Sei V ein K-Vektorraum.

(i) Für jeden Skalar $\lambda_0 \in K$, die Skalierungsabbildung

$$\begin{aligned} V &\to V, \\ v &\mapsto \lambda_0 \cdot v, \end{aligned}$$

ist linear. Bedingungen (i) und (ii) der Definition 4.1.1 folgen aus Axiomen (iv) und (ii) der Definition 3.2.1.

(ii) Sei $v_0 \in V$ ein Vektor. Ist $v_0 \neq 0$, so ist die Verschiebungsabbildung

$$\begin{aligned} V &\to V, \\ v &\mapsto v + v_0, \end{aligned}$$

nicht linear. Zum Beispiel, $(v + v') + v_0 \neq (v + v_0) + (v' + v_0) = (v + v') + 2v_0$.

Beispiel 4.1.4 (generische Beispiele)**.** Seien V, W Vektorräume über K und sei $U \subset V$ ein Untervektorraum. Die folgenden Abbildungen sind K-linear:

(i) Die Identität $\mathrm{id}_V \colon V \to V$.

(ii) Die *Nullabbildung* $0 \colon V \to W$, $v \mapsto 0$.

(iii) Die Inklusionsabbildung $U \to V$, $u \mapsto u$.

(iv) Die Quotientenabbildung $V \to V/U$, $v \mapsto v + U$.

(v) Die kanonischen Abbildungen

$$\begin{aligned} \iota_1 &\colon V \to V \oplus W, \quad v \mapsto (v, 0), \\ \iota_2 &\colon W \to V \oplus W, \quad w \mapsto (0, w). \end{aligned}$$

(vi) Die kanonischen Abbildungen

$$\begin{aligned} \pi_1 &\colon V \oplus W \to V, \quad (v, w) \mapsto v, \\ \pi_2 &\colon V \oplus W \to W, \quad (v, w) \mapsto w. \end{aligned}$$

Beispiel 4.1.5. Sei $F = (v_i)_{i \in I}$ eine Familie von Vektoren in einem K-Vektorraum V. Dann ist die Abbildung

$$\begin{aligned} \varphi_F &\colon K^{(I)} \to V, \\ (\lambda_i)_{i \in I} &\mapsto \sum_{i \in I} \lambda_i \cdot v_i, \end{aligned}$$

aus Konstruktion 3.3.8 K-linear.

Beispiel 4.1.6. Ob die Abbildung

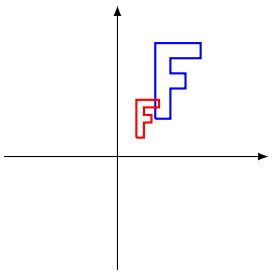
$$f \colon K \to K, \quad f(x) = x^2,$$

K-linear ist, hängt von dem Körper K ab. Es gilt $f(1 + 1) = 4$ und $f(1) + f(1) = 2$. Wenn die Charakteristik von K nicht 2 ist, dann ist $4 \neq 2$ (da $4 - 2 = 2 \neq 0$), und damit ist f auf jeden Fall keine lineare Abbildung. Aber ist $K = \mathbb{F}_2$, so ist f gleich der Identität (da $0^2 = 0$ und $1^2 = 1$), und insbesondere linear. Man kann jedoch zeigen, dass $\mathbb{F}_2$ der einzige Körper ist (bis auf Isomorphie), auf dem die Abbildung f K-linear ist. Zum Beispiel, wenn $K = \mathbb{F}_4$ (siehe Bemerkung 2.4.11), dann gilt $f(\alpha \cdot 1) = \alpha^2 = \beta \neq \alpha = \alpha \cdot f(1)$, und damit ist f nicht K-linear.

Beispiel 4.1.7 (lineare Abbildungen von $\mathbb{R}^2$ nach $\mathbb{R}^2$)**.** Beispiele von $\mathbb{R}$-linearen Abbildungen $\mathbb{R}^2 \to \mathbb{R}^2$ sind Skalierungen, Spiegelungen (an einer Ursprungsgerade), Drehungen (um den Ursprung), Scherungen, usw. In folgenden Beispielen, die rote Figur ist das Bild der blauen Figur unter der gegebenen linearen Abbildung.

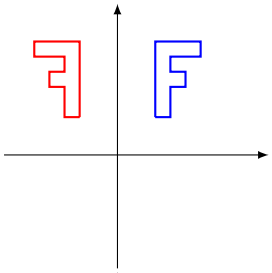
(i) Skalierung um $\frac{1}{2}$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



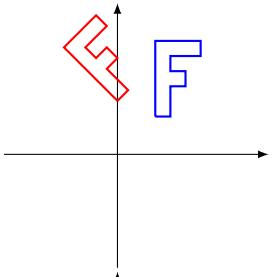
(ii) Spiegelung an $\mathbb{R}e_2$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}$$



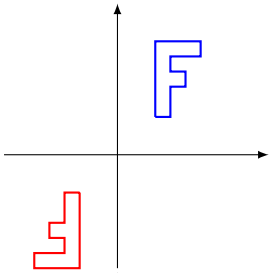
(iii) Drehung um $\frac{\pi}{4}$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \frac{\sqrt{2}}{2} \begin{pmatrix} x_1 - x_2 \\ x_1 + x_2 \end{pmatrix}$$



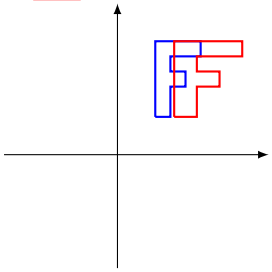
(iv) Spiegelung an 0 / Drehung um π:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}$$



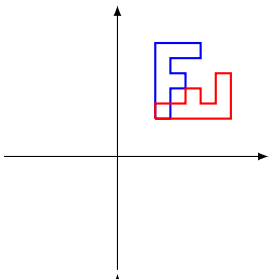
(v) Horizontale Skalierung um $\frac{3}{2}$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} \frac{3}{2}x_1 \\ x_2 \end{pmatrix}$$



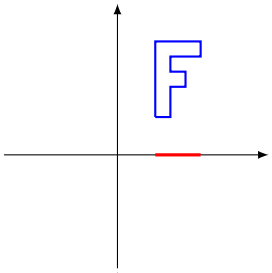
(vi) Koordinatenvertauschung / Spiegelung an $\mathbb{R}(e_1 + e_2)$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$



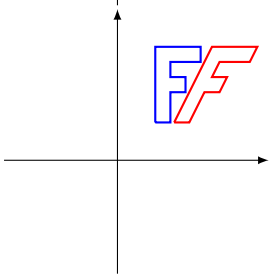
(vii) Orthogonale Projektion auf $\mathbb{R}e_1$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$



(viii) Horizontale Scherung um den Scherungsfaktor $\frac{1}{2}$:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + \frac{1}{2}x_2 \\ x_2 \end{pmatrix}$$



Beispiel 4.1.8 (Auswertung)**.** Sei X eine beliebige Menge, V ein K-Vektorraum und $\mathrm{Abb}(X, V)$ der K-Vektorraum aller Abbildungen von X nach V (siehe Beispiel 3.2.5). Zu jedem $x \in X$ gibt es eine *Auswertungsabbildung*

$$\begin{aligned} \mathrm{ev}_x \colon \mathrm{Abb}(X, V) &\to V, \\ f &\mapsto f(x). \end{aligned}$$

Aus der Definition der Addition und der Skalarmultiplikation auf $\mathrm{Abb}(X, V)$ folgt unmittelbar, dass ev_x eine K-lineare Abbildung ist. Die Abbildung ev_x ist auch die kanonische Projektion π_x aus Definition 1.3.17, wenn wir die Menge $\mathrm{Abb}(X, V)$ als das Produkt $\prod_{x \in X} V$ betrachten.

Beispiel 4.1.9 (analytische Beispiele)**.** Zwei der wichtigsten Operationen in der Analysis, Differentiation und Integration, sind Beispiele von linearen Abbildungen (siehe Beispiel 3.2.14).

(i) Sei $\mathrm{Diff}(\mathbb{R}, \mathbb{R}) \subset \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ der Untervektorraum aller differenzierbaren Funktionen auf $\mathbb{R}$. Dann ist die Abbildung

$$\begin{aligned} \mathrm{Diff}(\mathbb{R}, \mathbb{R}) &\to \mathrm{Abb}(\mathbb{R}, \mathbb{R}), \\ f &\mapsto f', \end{aligned}$$

$\mathbb{R}$-linear.

(ii) Seien $a < b$ reelle Zahlen und sei $\mathrm{Riem}([a, b], \mathbb{R}) \subset \mathrm{Abb}([a, b], \mathbb{R})$ der Untervektorraum aller Riemann-integrierbaren Funktionen auf $[a, b]$. Integration definiert eine $\mathbb{R}$-lineare Abbildung

$$\begin{aligned} \mathrm{Riem}([a, b], \mathbb{R}) &\to \mathbb{R}, \\ f &\mapsto \int_a^b f(x)\, dx. \end{aligned}$$

Andere analytische Beispiele stammen aus Folgen und Reihen:

(iii) Sei $\mathrm{Konv}(\mathbb{R}) \subset \mathbb{R}^{\mathbb{N}}$ der Untervektorraum aller konvergenten Folgen in $\mathbb{R}$ (siehe Beispiel 3.2.13). Die Grenzwertabbildung

$$\begin{aligned} \lim\colon \mathrm{Konv}(\mathbb{R}) &\to \mathbb{R}, \\ (a_n)_{n\in\mathbb{N}} &\mapsto \lim_{n\to\infty} a_n, \end{aligned}$$

ist dann $\mathbb{R}$-linear.

(iv) Sei $\mathrm{Konv}^{\Sigma}(\mathbb{R}) \subset \mathbb{R}^{\mathbb{N}}$ der Untervektorraum aller Folgen $(a_n)_{n\in\mathbb{N}}$, deren Reihe $\sum_{n=0}^{\infty} a_n$ konvergiert. Dann ist die Abbildung

$$\begin{aligned} \mathrm{Konv}^{\Sigma}(\mathbb{R}) &\to \mathbb{R}, \\ (a_n)_{n\in\mathbb{N}} &\mapsto \sum_{n=0}^{\infty} a_n, \end{aligned}$$

$\mathbb{R}$-linear.

Proposition 4.1.10 (Rechnen mit linearen Abbildungen)**.** *Seien V, W Vektorräume über K und sei $f\colon V \to W$ eine K-lineare Abbildung.*

(i) *Es gilt $f(0) = 0$.*

(ii) *Für alle $v \in V$ gilt $f(-v) = -f(v)$.*

(iii) *Sei I eine endliche Menge, $(v_i)_{i\in I}$ eine Familie von Vektoren aus V und $(\lambda_i)_{i\in I}$ eine Familie von Skalaren. Dann gilt*

$$f\left(\sum_{i\in I} \lambda_i \cdot v_i\right) = \sum_{i\in I} \lambda_i \cdot f(v_i).$$

Beweis. Zu (i). Da f linear ist, gilt

$$f(0) + f(0) = f(0 + 0) = f(0).$$

Wenn wir $f(0)$ von beiden Seiten subtrahieren, erhalten wir $f(0) = 0$.

Zu (ii). Nach der Linearität von f und (i) gilt

$$f(v) + f(-v) = f(v + (-v)) = f(0) = 0.$$

Da das inverse Element eindeutig ist, folgt $f(-v) = -f(v)$.

Zu (iii). Dies wird durch Induktion über die Mächtigkeit von I bewiesen. Falls $|I| = 0$ ist die Aussage dieselbe wie (i). Zum Induktionsschritt verwendet man die Gleichung $f(\lambda \cdot v + w) = \lambda \cdot f(v) + f(w)$, die direkt aus der Definition der Linearität folgt. □

Bemerkung 4.1.11. In den Beweisen von Aussagen (i) und (ii) haben wir nur Axiom (i) der Definition 4.1.1 benutzt. Insbesondere gelten diese Aussagen auch für Gruppenhomomorphismen. Es ist aber auch möglich, beide Aussagen nur durch Axiom (ii) nachzuprüfen.

Proposition 4.1.12. *Seien U, V, W Vektorräume über K.*

(i) *Sind $f\colon U \to V$ und $g\colon V \to W$ lineare Abbildungen, so ist $g \circ f\colon U \to W$ linear.*

(ii) *Sind $f, g\colon V \to W$ lineare Abbildungen, so ist ihre Summe*

$$\begin{aligned} f + g\colon V &\to W, \\ v &\mapsto f(v) + g(v), \end{aligned}$$

linear.

(iii) *Ist $f\colon V \to W$ eine lineare Abbildung und ist $\lambda \in K$, so ist*

$$\begin{aligned} \lambda \cdot f\colon V &\to W, \\ v &\mapsto \lambda \cdot f(v), \end{aligned}$$

linear.

Beweis. Jede Aussage ist eine direkte Berechnung. Wir beweisen stellvertretend (iii). Seien $v, w \in V$ und $\mu \in K$. Man berechnet:

$$\begin{aligned} (\lambda \cdot f)(v+w) &= \lambda \cdot f(v+w) \\ &= \lambda(f(v)+f(w)) \\ &= \lambda \cdot f(v) + \lambda \cdot f(w) \\ &= (\lambda \cdot f)(v) + (\lambda \cdot f)(w), \\ (\lambda \cdot f)(\mu \cdot v) &= \lambda \cdot f(\mu \cdot v) \\ &= \lambda \cdot (\mu \cdot f(v)) \\ &= \mu \cdot (\lambda \cdot f(v)) \\ &= \mu \cdot (\lambda \cdot f)(v). \end{aligned}$$

Man beachte dabei, dass zusätzlich zu den Vektorraumaxiomen auch die Kommutativität der Multiplikation auf K in der vorletzten Gleichung benutzt wurde. □

Bemerkung 4.1.13 (lineare Abbildungen und Körpererweiterungen)**.** Sei $K \subset L$ eine Körpererweiterung (Beispiel 3.2.4), seien V, W Vektorräume über L, und sei $f\colon V \to W$ eine L-lineare Abbildung. Dann ist f auch K-linear, wenn wir V und W als K-Vektorräume betrachten (siehe Bemerkung 3.2.15). Die Umkehrung gilt nicht: Beispielsweise ist die komplexe Konjugation $\mathbb{C} \to \mathbb{C}$, $a + bi \mapsto a - bi$, $\mathbb{R}$-linear aber nicht $\mathbb{C}$-linear.

Definition 4.1.14 (Isomorphismus, isomorph)**.** Seien V, W Vektorräume über K.

- Eine K-lineare Abbildung $f\colon V \to W$ heißt *Isomorphismus*, wenn eine K-lineare Abbildung $g\colon W \to V$ existiert, so dass $g \circ f = \mathrm{id}_V$ und $f \circ g = \mathrm{id}_W$. Man schreibt manchmal $f\colon V \xrightarrow{\sim} W$, wenn f ein Isomorphismus ist.
- V ist *isomorph* zu W, in Zeichen $V \cong W$, wenn ein Isomorphismus von V nach W existiert.

Die Idee hinter dieser Definition ist, dass isomorphe K-Vektorräume V und W genau dieselben Vektorraumeigenschaften haben sollen. Genauer, wenn ein Isomorphismus $f\colon V \to W$ gegeben ist, können wir jede Aussage über V, die nur die Vektorraumstruktur von V benutzt, auf eine Aussage über W durch f übertragen. Was damit gemeint ist wird nach und nach deutlich werden, aber hier sind ein paar Beispiele:

- Isomorphe Vektorräume haben dieselbe Dimension.
- Das Bild einer Basis unter einem Isomorphismus ist wieder eine Basis.

Bemerkung 4.1.15. Das Wort „Isomorphismus“ wird in vielen verschiedenen Zusammenhängen verwendet, zum Beispiel auch bei Gruppen, Körpern, partiell geordneten Mengen, usw. (siehe Bemerkung 4.1.2). Es bedeutet immer dasselbe: Ein Isomorphismus ist eine strukturerhaltende Abbildung, die ein strukturerhaltende Umkehrabbildung besitzt.

Bemerkung 4.1.16. Isomorphie ist eine Äquivalenzrelation zwischen K-Vektorräumen:

- V ist isomorph zu sich selbst, da id_V ein Isomorphismus ist.
- Ist $f\colon V \to W$ ein Isomorphismus, so ist seine Umkehrabbildung $f^{-1}\colon W \to V$ auch ein Isomorphismus.

- Die Komposition zweier Isomorphismen ist wieder ein Isomorphismus.

Folgende Proposition ist eine lineare Variante von Satz 1.3.23:

Proposition 4.1.17. *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Die folgenden Aussagen sind äquivalent:*

(i) *f ist ein Isomorphismus.*

(ii) *f ist bijektiv.*

Beweis. Die Implikation (i) $\Rightarrow$ (ii) folgt aus dem Satz 1.3.23. Sei umgekehrt $f\colon V \to W$ bijektiv. Dann besitzt f eine (eindeutige) Umkehrabbildung $g\colon W \to V$, und es bleibt zu zeigen, dass g auch K-linear ist. Seien $w, w' \in W$ und $\lambda \in K$, und seien $v = g(w)$ und $v' = g(w')$. Da $f \circ g = \mathrm{id}_W$ gilt $f(v) = w$ und $f(v') = w'$. Wir berechnen:

$$g(w + w') = g(f(v) + f(v')) = g(f(v + v')) = v + v' = g(w) + g(w'),$$
$$g(\lambda \cdot w) = g(\lambda \cdot f(v)) = g(f(\lambda \cdot v)) = \lambda \cdot v = \lambda \cdot g(w).$$

Dabei haben wir die Linearität von f und die Gleichung $g \circ f = \mathrm{id}_V$ verwendet. □

Definition 4.1.18 (Endomorphismus, Automorphismus)**.** Sei V ein K-Vektorraum.

- Ein *Endomorphismus* von V ist eine lineare Abbildung von V nach V. Die Menge aller Endomorphismen von V wird mit $\mathrm{End}_K(V)$ bezeichnet.
- Ein Endomorphismus von V, der auch ein Isomorphismus ist, heißt *Automorphismus* von V. Die Menge aller Automorphismen von V wird mit $\mathrm{Aut}_K(V)$ bezeichnet.

Bemerkung 4.1.19. Nach Propositionen 4.1.12(i) und 4.1.17 ist die Menge $\mathrm{Aut}_K(V)$ eine Gruppe bezüglich Komposition.

4.1.1 Lineare Abbildungen und Basen

Proposition 4.1.20. *Seien V, W Vektorräume über K und $f\colon V \to W$ eine lineare Abbildung.*

(i) *Für jede Teilmenge $E \subset V$ gilt*

$$f(\mathrm{Span}_K(E)) = \mathrm{Span}_K(f(E)).$$

Insbesondere, ist f surjektiv und ist E ein Erzeugendensystem von V, so ist $f(E)$ ein Erzeugendensystem von W.

(ii) *Sei $(v_i)_{i\in I}$ eine Familie von Vektoren aus V. Ist die Familie $(f(v_i))_{i\in I}$ in W linear unabhängig, so ist $(v_i)_{i\in I}$ linear unabhängig. Die Umkehrung gilt, wenn f injektiv ist.*

Beweis. Die erste Aussage folgt aus Propositionen 3.2.18 und 4.1.10(iii). Sei $(f(v_i))_{i\in I}$ linear unabhängig, und sei $(\lambda_i)_{i\in I} \in K^{(I)}$ mit $\sum_{i\in I} \lambda_i \cdot v_i = 0$. Nach Proposition 4.1.10(i,iii) gilt dann $\sum_{i\in I} \lambda_i \cdot f(v_i) = 0$. Aus der vorausgesetzten linearen Unabhängigkeit folgt jetzt $\lambda_i = 0$ für alle $i \in I$. Also ist $(v_i)_{i\in I}$ linear unabhängig.

Sei umgekehrt $(v_i)_{i\in I}$ linear unabhängig und f injektiv, und sei $(\lambda_i)_{i\in I} \in K^{(I)}$ mit $\sum_{i\in I} \lambda_i \cdot f(v_i) = 0$. Nach Proposition 4.1.10(i,iii) gilt dann $f\left(\sum_{i\in I} \lambda_i \cdot v_i\right) = f(0)$, und damit $\sum_{i\in I} \lambda_i \cdot v_i = 0$ nach der Injektivität von f. Aus der vorausgesetzten linearen Unabhängigkeit folgt jetzt $\lambda_i = 0$ für alle $i \in I$. Also ist $(f(v_i))_{i\in I}$ linear unabhängig. □

Korollar 4.1.21. *Sei $f\colon V \to W$ ein Isomorphismus von K-Vektorräumen. Ist $(v_i)_{i\in I}$ eine Basis von V, so ist $(f(v_i))_{i\in I}$ eine Basis von W. Insbesondere gilt*

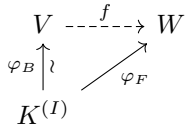
$$\dim_K(V) = \dim_K(W).$$

Beweis. Die Familie $(f(v_i))_{i\in I}$ ist erzeugend bzw. linear unabhängig nach Proposition 4.1.20 (i) bzw. (ii). □

Satz 4.1.22 (universelle Eigenschaft von Basen)**.** *Sei V ein K-Vektorraum mit einer Basis $(v_i)_{i\in I}$. Zu jedem K-Vektorraum W und jeder Familie $(w_i)_{i\in I}$ in W mit Indexmenge I, gibt es* genau eine *K-lineare Abbildung $f\colon V \to W$ mit $f(v_i) = w_i$ für alle $i \in I$. Außerdem:*

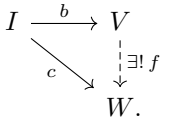
- *f ist genau dann surjektiv, wenn $(w_i)_{i\in I}$ erzeugend ist.*
- *f ist genau dann injektiv, wenn $(w_i)_{i\in I}$ linear unabhängig ist.*
- *f ist genau dann bijektiv, wenn $(w_i)_{i\in I}$ eine Basis ist.*

Beweis. Sei $B = (v_i)_{i\in I}$ and $F = (w_i)_{i\in I}$. Da B erzeugend ist, ist jedes $v \in V$ Linearkombination der Vektoren v_i. Die Eindeutigkeit von f folgt dann aus Proposition 4.1.10(iii). Da B eine Basis von V ist, ist die lineare Abbildung $\varphi_B\colon K^{(I)} \to V$ aus Konstruktion 3.3.8 bijektiv (siehe Bemerkung 3.3.18). Die lineare Abbildung $f = \varphi_F \circ \varphi_B^{-1}\colon V \to W$ hat dann die gewünschte Eigenschaft, da $\varphi_F(\varphi_B^{-1}(v_i)) = \varphi_F(e_i) = w_i$.



Die drei zusätzlichen Aussagen folgen aus den entsprechenden Aussagen für φ_F (siehe Bemerkung 3.3.18). □

Bemerkung 4.1.23. Wenn man die Basis $(v_i)_{i\in I}$ von V als eine Abbildung $b\colon I \to V$ betrachtet, kann man die universelle Eigenschaft so formulieren: Zu jeder Abbildung $c\colon I \to W$ gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $f \circ b = c$:



Korollar 4.1.24 (Klassifikation von Vektorräumen bis auf Isomorphie)**.**

(i) *Jeder K-Vektorraum V ist zu $K^{(I)}$ isomorph, wobei I die Indexmenge einer Basis von V ist.*

(ii) *Zwei K-Vektorräume V und W sind genau dann isomorph, wenn sie dieselbe Dimension haben (d.h., wenn ihre jeweiligen Basen gleichmächtig sind).*

Beweis. Zu (i). Wir wenden den Satz 4.1.22 mit der Basis $(e_i)_{i\in I}$ von $K^{(I)}$ und einer beliebigen I-indizierten Basis von V an, um einen Isomorphismus $f\colon K^{(I)} \xrightarrow{\sim} V$ zu erhalten.

Zu (ii). Eine der beiden Implikationen folgt bereits aus Korollar 4.1.21. Seien umgekehrt $(v_i)_{i\in I}$ und $(w_j)_{j\in J}$ Basen von V und W, so dass I und J gleichmächtig sind. Es existiert also eine bijektive Abbildung $a\colon I \xrightarrow{\sim} J$. Da a bijektiv ist, ist $(w_{a(i)})_{i\in I}$ wieder eine Basis von W. Nach Satz 4.1.22 ist die eindeutige lineare Abbildung $f\colon V \to W$ mit $f(v_i) = w_{a(i)}$ ein Isomorphismus. □

Diese Klassifikation kann man auf folgende Weise zusammenfassen: Es gibt eine Bijektion

$$\{\text{Vektorräume über } K\}/\text{Isomorphie} \leftrightarrow \{\text{Mengen}\}/\text{Gleichmächtigkeit},$$
$$K^{(I)} \leftarrow\!\shortmid I.$$

Das ist aber etwas schlampig, da K-Vektorräume bzw. Mengen keine Menge bilden.

Beispiel 4.1.25. Ist V endlich-dimensional der Dimension n, so gilt $V \cong K^n$.

Bemerkung 4.1.26. Das Korollar 4.1.24 impliziert insbesondere, dass der K-Vektorraum $K^{\mathbb{N}}$ aller Folgen in K zu einem K-Vektorraum der Gestalt $K^{(I)}$ isomorph ist. Aber die Mächtigkeit von I hängt von der des Körpers K ab. Wenn $|K| \leq |\mathbb{R}|$, zum Beispiel wenn K ein Teilkörper von $\mathbb{C}$ oder ein endlicher Körper ist, dann gilt $K^{\mathbb{N}} \cong K^{(\mathbb{R})}$.

4.1.2 Kern und Bild linearer Abbildungen

Definition 4.1.27 (Kern, Bild)**.** Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen.

- Der *Kern* (oder der *Nullraum*) von f ist
$$\ker f := f^{-1}(\{0\}) = \{v \in V \mid f(v) = 0\}.$$
- Das *Bild* von f ist
$$\operatorname{im} f := f(V) = \{w \in W \mid \text{es existiert } v \in V \text{ mit } f(v) = w\}.$$

Proposition 4.1.28 (Kern und Bild sind Untervektorräume)**.** *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen.*

(i) *Ist $U \subset W$ ein Untervektorraum, so ist $f^{-1}(U) \subset V$ ein Untervektorraum. Insbesondere ist der Kern von f ist ein Untervektorraum von V.*

(ii) *Ist $U \subset V$ ein Untervektorraum, so ist $f(U) \subset W$ ein Untervektorraum. Insbesondere ist das Bild von f ist ein Untervektorraum von W.*

Beweis. Wir verwenden das Kriterium 3.2.8.

Zu (i). $f^{-1}(U)$ enthält den Nullvektor (nach Proposition 4.1.10(i)). Sind $v, v' \in f^{-1}(U)$ und $\lambda \in K$, so gilt
$$f(v + v') = f(v) + f(v') \in U \quad \text{und} \quad f(\lambda \cdot v) = \lambda \cdot f(v) \in U,$$
und damit liegen $v + v'$ und $\lambda \cdot v$ auch in $f^{-1}(U)$.

Zu (ii). $f(U)$ enthält den Nullvektor (nach Proposition 4.1.10(i)). Seien $w, w' \in f(U)$ und $\lambda \in K$. Nach Definition von $f(U)$ existieren $v, v' \in U$ mit $f(v) = w$ und $f(v') = w'$. Dann gilt
$$f(v + v') = f(v) + f(v') = w + w' \quad \text{und} \quad f(\lambda \cdot v) = \lambda \cdot f(v) = \lambda \cdot w,$$
und damit liegen $w + w'$ und $\lambda \cdot w$ auch in $f(U)$. □

Beispiel 4.1.29. Wir berechnen den Kern und das Bild der Abbildungen aus Beispiel 4.1.4.

(i) Für die Identität $\mathrm{id}_V\colon V \to V$ gilt $\ker \mathrm{id}_V = \{0\}$ und $\operatorname{im} \mathrm{id}_V = V$.

(ii) Für die Nullabbildung $0\colon V \to W$ gilt $\ker 0 = V$ und $\operatorname{im} 0 = \{0\}$.

(iii) Für die Inklusionsabbildung $i\colon U \hookrightarrow V$ gilt $\ker i = \{0\}$ und $\operatorname{im} i = U$.

(iv) Für die Quotientenabbildung $q\colon V \twoheadrightarrow V/U$ gilt $\ker q = U$ und $\operatorname{im} q = V/U$.

(v) Für die kanonische Abbildung $\iota_1\colon V \hookrightarrow V \oplus W$ gilt $\ker \iota_1 = \{0\}$ und $\operatorname{im} \iota_1 = V \times \{0\}$.

(vi) Für die kanonische Abbildung $\pi_1\colon V \oplus W \twoheadrightarrow V$ gilt $\ker \pi_1 = \{0\} \times W$ und $\operatorname{im} \pi_1 = V$.

Beispiel 4.1.30. Sei $\lim\colon \mathrm{Konv}(\mathbb{R}) \to \mathbb{R}$ die Grenzwertabbildung aus Beispiel 4.1.9(iii). Der Kern von lim besteht aus allen Folgen, die gegen 0 konvergieren. Der $\mathbb{R}$-Vektorraum $\mathrm{Konv}^{\Sigma}(\mathbb{R})$ aus Beispiel 4.1.9(iv) ist ein Untervektorraum von $\ker(\lim)$.

Beispiel 4.1.31 (lineare Differentialgleichungen)**.** Sei $V = C^\infty(\mathbb{R}, \mathbb{R})$ der $\mathbb{R}$-Vektorraum aller glatten (d.h., beliebig oft differenzierbaren) Funktionen von $\mathbb{R}$ nach $\mathbb{R}$, und sei $D\colon V \to V$ die lineare Abbildung $f \mapsto f'$. Dann ist die Abbildung D surjektiv, und ihr Kern ist der Untervektorraum Konst$(\mathbb{R}, \mathbb{R})$ der konstanten Funktionen. Eine Abbildung $L\colon V \to V$ der Gestalt

$$L = D^n + f_{n-1} \cdot D^{n-1} + \cdots f_1 \cdot D + f_0 \cdot \mathrm{id}_V$$

mit $n \in \mathbb{N}$ und $f_i \in C^\infty(\mathbb{R}, \mathbb{R})$ heißt *linearer Differentialoperator n-ter Ordnung* (dabei werden die Potenzen von D bezüglich der Komposition genommen, d.h., $D^i(f)$ ist die i-te Ableitung von f). In der Analysis wird gezeigt, dass $\ker L$ immer ein n-dimensionaler Untervektorraum von V ist. Beispielsweise wird $\ker(D - \mathrm{id}_V)$ von der Exponentialfunktion exp erzeugt, und wird $\ker(D^2 + \mathrm{id}_V)$ von den Winkelfunktionen cos und sin erzeugt.

Proposition 4.1.32. *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen.*

(i) *f ist genau dann injektiv, wenn* $\ker f = \{0\}$.

(ii) *f ist genau dann surjektiv, wenn* $\operatorname{im} f = W$.

Beweis. Die zweite Aussage ist genau die Definition der Surjektivität. Sei $f\colon V \to W$ injektiv. Dann hat $0 \in W$ höchstens ein Urbild unter f. Da $0 \in V$ ein solches Urbild ist, ist $\ker f$ genau gleich $\{0\}$. Sei umgekehrt $\ker f = \{0\}$, und seien $v, v' \in V$ mit $f(v) = f(v')$. Dann

$$f(v - v') = f(v) - f(v') = 0,$$

und damit $v - v' \in \ker f = \{0\}$, d.h., $v = v'$. Also ist f injektiv. □

Die folgende Proposition ist eine lineare Variante der universallen Eigenschaft der Quotientenmenge (Satz 1.4.10).

Proposition 4.1.33 (universelle Eigenschaft des Quotientenvektorraums)**.** *Sei V ein K-Vektorraum, $U \subset V$ ein Untervektorraum, V/U der Quotientenvektorraum (siehe Definition 3.2.24) und $q\colon V \to V/U$ die Quotientenabbildung. Zu jedem K-Vektorraum W und jeder linearen Abbildung $f\colon V \to W$ mit $U \subset \ker f$ gibt es* genau eine *lineare Abbildung $\bar{f}\colon V/U \to W$ mit $\bar{f} \circ q = f$.*

Beweis. Zur Erinnerung ist V/U die Quotientenmenge $V/\!\sim_U$, wobei $x \sim_U y \Leftrightarrow x - y \in U$. Falls $x \sim_U y$, folgt aus der Voraussetzung $f(U) = \{0\}$, dass $f(x - y) = 0$, d.h., $f(x) = f(y)$. Nach der universellen Eigenschaft der Quotientenmenge gibt es genau eine Abbildung $\bar{f}\colon V/U \to W$ mit $\bar{f} \circ q = f$. Es bleibt zu zeigen, dass $\bar{f}$ linear ist. Dies folgt aus der Linearität von f und q:

$$\begin{aligned}
\bar{f}((v + U) + (v' + U)) &= \bar{f}((v + v') + U) = f(v + v') \\
&= f(v) + f(v') = \bar{f}(v + U) + \bar{f}(v' + U), \\
\bar{f}(\lambda \cdot (v + U)) &= \bar{f}(\lambda v + U) = f(\lambda v) = \lambda \cdot f(v) = \lambda \cdot \bar{f}(v + U).
\end{aligned}$$

□

Bemerkung 4.1.34. Es gibt eine ähnliche universelle Eigenschaft für Untervektorräume, aber sie ist offensichtlicher: Ist $i\colon U \hookrightarrow V$ die Inklusionsabbildung eines Untervektorraums und ist $f\colon W \to V$ eine lineare Abbildung mit $\operatorname{im} f \subset U$, so gibt es genau eine lineare Abbildung $\bar{f}\colon W \to U$ mit $i \circ \bar{f} = f$.

Satz 4.1.35 (Homomorphiesatz für Vektorräume)**.** *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Dann ist die von f induzierte Abbildung*

$$\begin{aligned}
\bar{f}\colon V/\ker f &\to \operatorname{im} f, \\
v + \ker f &\mapsto f(v),
\end{aligned}$$

ein Isomorphismus von K-Vektorräumen.

Beweis. Die Abbildung $\bar{f}$ ist wohldefiniert und linear nach Proposition 4.1.33, und sie ist surjektiv nach Definition von $\operatorname{im} f$. Zur Injektivität berechnen wir den Kern von $\bar{f}$. Es gilt:

$$\begin{aligned}\ker \bar{f} &= \{v + \ker f \mid f(v) = 0\} \\ &= \{v + \ker f \mid v \in \ker f\} \\ &= \{0 + \ker f\}.\end{aligned}$$

Also ist $\bar{f}$ injektiv nach Proposition 4.1.32(i). □

Bemerkung 4.1.36. Der Homomorphiesatz impliziert folgenden Zerlegungssatz für lineare Abbildungen: Jede lineare Abbildung $f\colon V \to W$ lässt sich kanonisch als Komposition einer Quotientenabbildung, eines Isomorphismus und einer Inklusionsabbildung zerlegen:

$$V \overset{q}{\twoheadrightarrow} V/\ker f \overset{\bar{f}}{\to} \operatorname{im} f \overset{i}{\hookrightarrow} W.$$

Beispiel 4.1.37. Sei $C^0(\mathbb{R},\mathbb{R})$ der $\mathbb{R}$-Vektorraum aller stetigen reellen Funktionen auf $\mathbb{R}$, und sei $C^1(\mathbb{R},\mathbb{R}) \subset C^0(\mathbb{R},\mathbb{R})$ der Untervektorraum aller stetig differenzierbaren Funktionen (d.h., Funktionen f, deren Ableitung f' existiert und stetig ist). Dann haben wir die $\mathbb{R}$-lineare Differentiationsabbildung

$$\begin{aligned}D\colon C^1(\mathbb{R},\mathbb{R}) &\to C^0(\mathbb{R},\mathbb{R}), \\ f &\mapsto f'\end{aligned}$$

(siehe Beispiel 4.1.9). In der Analysis wird gezeigt, dass jede stetige Abbildung $g \in C^0(\mathbb{R},\mathbb{R})$ eine Stammfunktion besitzt, d.h., eine Funktion f mit $f' = g$; zum Beispiel,

$$f(x) = \int_0^x g(t)\,dt.$$

Anders gesagt ist die obige Abbildung D surjektiv. In der Analysis wird auch gezeigt, dass f' genau dann gleich Null ist, wenn f eine konstante Funktion ist. Der Kern von D ist also der Untervektorraum $\operatorname{Konst}(\mathbb{R},\mathbb{R}) \subset C^1(\mathbb{R},\mathbb{R})$ aller konstanten Funktionen, der offensichtlich zu $\mathbb{R}$ isomorph ist. Der Homomorphiesatz impliziert, dass D einen Isomorphismus

$$\bar{D}\colon C^1(\mathbb{R},\mathbb{R})/\operatorname{Konst}(\mathbb{R},\mathbb{R}) \xrightarrow{\sim} C^0(\mathbb{R},\mathbb{R})$$

induziert.

Korollar 4.1.38 (Dimensionsformel für lineare Abbildungen)**.** *Seien V, W Vektorräume über K und sei $f\colon V \to W$ eine K-lineare Abbildung. Ist V endlich-dimensional, so gilt*

$$\dim_K(V) = \dim_K(\ker f) + \dim_K(\operatorname{im} f).$$

Beweis. Nach dem Homomorphiesatz 4.1.35 gilt $\dim_K(\operatorname{im} f) = \dim_K(V/\ker f)$. Die gewünschte Gleichung folgt jetzt aus der Dimensionsformel für Quotientenvektorräume (Proposition 3.3.46). □

Korollar 4.1.39 (Charakterisierung von Isomorphismen)**.** *Seien V und W K-Vektorräume derselben* endlichen *Dimension und sei $f\colon V \to W$ eine lineare Abbildung. Die folgenden Aussagen sind dann äquivalent:*

(i) *f ist bijektiv, d.h., ein Isomorphismus.*

(ii) *f ist injektiv, d.h.,* $\ker f = \{0\}$.

(iii) *f ist surjektiv, d.h.,* $\operatorname{im} f = W$.

Beweis. Sei $\ker f = \{0\}$. Nach Korollar 4.1.38 gilt $\dim_K(\operatorname{im} f) = \dim_K(V) = \dim_K(W)$. Aus Proposition 3.3.35 folgt, dass $\operatorname{im} f = W$. Sei umgekehrt $\operatorname{im} f = W$. Aus Korollar 4.1.38 folgt dann, dass $\dim(\ker f) = 0$, d.h., dass $\ker f = \{0\}$. □

Bemerkung 4.1.40. Das Korollar 4.1.39 gilt nicht bei unendlich-dimensionalen Vektorräumen, selbst wenn wir die Dimension im Sinne der Bemerkung 3.3.31 verstehen. Die lineare Abbildung

$$K^{\mathbb{N}} \to K^{\mathbb{N}}, \quad (x_0, x_1, x_2, \dots) \mapsto (0, x_0, x_1, \dots),$$

ist injektiv aber nicht bijektiv, und die lineare Abbildung

$$K^{\mathbb{N}} \to K^{\mathbb{N}}, \quad (x_0, x_1, x_2, \dots) \mapsto (x_1, x_2, x_3, \dots),$$

ist surjektiv aber nicht bijektiv. Ein anderes Gegenbeispiel ist die Differentiationsabbildung $D\colon C^\infty(\mathbb{R}, \mathbb{R}) \to C^\infty(\mathbb{R}, \mathbb{R})$ aus Beispiel 4.1.31, die surjektiv aber nicht injektiv ist (insbesondere ist $C^\infty(\mathbb{R}, \mathbb{R})$ unendlich-dimensional).

Definition 4.1.41 (Rang)**.** Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Man nennt die Dimension des Bildes von f den *Rang* von f:

$$\operatorname{rg} f := \dim_K(\operatorname{im} f).$$

Man beachte dabei, dass $\operatorname{rg} f \leq \dim_K(V)$ (nach Satz 4.1.35) und $\operatorname{rg} f \leq \dim_K(W)$ (da $\operatorname{im} f \subset W$). Der Rang ist ein wichtiger Begriff, denn er spielt eine ähnliche Rolle bei linearen Abbildungen wie die Dimension bei Vektorräumen: Bis auf Isomorphie werden Vektorräume durch ihre Dimension bestimmt (Korollar 4.1.24), und man kann den folgenden Satz so verstehen, dass lineare Abbildungen bis auf Isomorphie durch ihren Rang bestimmt werden:

Satz 4.1.42 (Klassifikation von linearen Abbildungen bis auf Isomorphie)**.** *Seien V und W Vektorräume über K und $f\colon V \to W$ eine lineare Abbildung. Dann existieren Mengen I, J und $L \subset I \cap J$ mit $|L| = \operatorname{rg} f$, und Isomorphismen $\varphi\colon K^{(I)} \xrightarrow{\sim} V$ und $\psi\colon K^{(J)} \xrightarrow{\sim} W$, so dass*

$$\psi^{-1} \circ f \circ \varphi = \iota_L \circ \pi_L\colon K^{(I)} \to K^{(J)},$$

wobei $\pi_L\colon K^{(I)} \twoheadrightarrow K^{(L)}$ und $\iota_L\colon K^{(L)} \hookrightarrow K^{(J)}$ die kanonischen Abbildungen sind, d.h., die Einschränkung auf L und die Nullfortsetzung auf J.

Beweis. Sei $(v_i)_{i\in M}$ eine Basis von $\ker f$. Nach dem Basisergänzungssatz können wir sie zu einer Basis $B = (v_i)_{i\in I}$ von V ergänzen; sei $L = I \setminus M$ und $U = \operatorname{Span}_K\{v_i \mid i \in L\}$. Da U und $\ker f$ komplementär sind, ist die Einschränkung $f|_U$ injektiv. Nach Proposition 4.1.20(ii) ist die Familie $(f(v_i))_{i\in L}$ in W linear unabhängig, und damit eine Basis von $\operatorname{im} f$ (insbesondere gilt $|L| = \operatorname{rg} f$). Nach dem Basisergänzungssatz gibt es eine Basis $C = (w_j)_{j\in J}$ von W mit $L \subset J$ und $w_i = f(v_i)$ für alle $i \in L$. Dann gilt $\varphi_C^{-1} \circ f \circ \varphi_B = \iota_L \circ \pi_L$ nach Konstruktion. □

4.1.3 Homomorphismenräume

Seien V und W Vektorräume über K. Zur Erinnerung bezeichnen wir mit $\operatorname{Hom}_K(V, W)$ die Menge aller K-linearen Abbildungen von V nach W. Sie ist auf kanonische Weise ein K-Vektorraum:

Proposition 4.1.43. *Seien V, W Vektorräume über K. Wenn wir die Menge* $\operatorname{Abb}(V, W)$ *als K-Vektorraum bezüglich der punktweisen Addition bzw. Skalarmultiplikation betrachten (Beispiel 3.2.5), dann ist* $\operatorname{Hom}_K(V, W)$ *ein Untervektorraum von* $\operatorname{Abb}(V, W)$.

Beweis. Dies folgt aus Proposition 4.1.12(ii),(iii) und dem Kriterium 3.2.8 ($\operatorname{Hom}_K(V, W)$ ist nicht leer, da die Nullabbildung K-linear ist). □

Bemerkung 4.1.44. Insbesondere ist $\mathrm{End}_K(V)$ ein K-Vektorraum. Falls $V \neq \{0\}$ ist die Teilmenge $\mathrm{Aut}_K(V)$ von $\mathrm{End}_K(V)$ *kein* Untervektorraum, da die Nullabbildung kein Automorphismus ist.

Beispiel 4.1.45 (lineare Abbildungen von K nach K)**.** Sei $f\colon K \to K$ eine K-linear Abbildung. Für alle $x \in K$ gilt dann

$$f(x) = f(x \cdot 1) = x \cdot f(1) = f(1) \cdot x.$$

Das heißt, f ist gleich der Multiplikation mit $f(1) \in K$. Ist umgekehrt $\lambda \in K$, so ist $x \mapsto \lambda \cdot x$ eine K-lineare Abbildung $K \to K$ nach Beispiel 4.1.3(i), die 1 auf λ abbildet. Es gibt also zueinander inverse Bijektionen

$$\begin{aligned} K &\rightleftarrows \mathrm{Hom}_K(K, K), \\ \lambda &\mapsto (x \mapsto \lambda \cdot x), \\ f(1) &\mapsfrom f. \end{aligned}$$

Außerdem kann man leicht nachprüfen, dass beide Abbildungen linear sind. Insbesondere gilt $\mathrm{Hom}_K(K, K) \cong K$.

Proposition 4.1.46 (Funktorialität der Homomorphismenräume)**.** *Seien V, V', W, W' Vektorräume über K und $f\colon V \to V'$ und $g\colon W \to W'$ lineare Abbildungen. Dann ist die Abbildung*

$$\begin{aligned} \mathrm{Hom}_K(f, g)\colon \mathrm{Hom}_K(V', W) &\to \mathrm{Hom}_K(V, W'), \\ h &\mapsto g \circ h \circ f, \end{aligned}$$

K-linear. Insbesondere sind die Abbildungen

$$\begin{aligned} \mathrm{Hom}_K(\mathrm{id}_V, g)\colon \mathrm{Hom}_K(V, W) &\to \mathrm{Hom}_K(V, W'), \\ h &\mapsto g \circ h, \\ \mathrm{Hom}_K(f, \mathrm{id}_W)\colon \mathrm{Hom}_K(V', W) &\to \mathrm{Hom}_K(V, W), \\ h &\mapsto h \circ f \end{aligned}$$

K-linear.

Beweis. Dies folgt durch Nachrechnen aus der Linearität von g. □

Definition 4.1.47 (Linearform, Dualraum, duale Abbildung)**.**

- Sei V ein Vektorraum über K. Eine *Linearform* auf V ist eine lineare Abbildung $V \to K$. Der *Dualraum* V^* von V ist der K-Vektorraum aller Linearformen auf V:

$$V^* = \mathrm{Hom}_K(V, K).$$

- Sei $f\colon V \to W$ eine K-lineare Abbildung. Die *duale Abbildung* f^* zu f ist die K-lineare Abbildung

$$\begin{aligned} f^* = \mathrm{Hom}_K(f, \mathrm{id}_K)\colon W^* &\to V^*, \\ \alpha &\mapsto \alpha \circ f. \end{aligned}$$

Beispiel 4.1.48. Sei $n \in \mathbb{N}$. Die n kanonischen Projektionen $\pi_i\colon K^n \to K$ sind Linearformen auf K^n.

Beispiel 4.1.49. Seien $a < b$ reelle Zahlen und sei $\mathrm{Riem}([a, b], \mathbb{R})$ der $\mathbb{R}$-Vektorraum aller Riemman-integrierbaren Funktionen auf $[a, b]$. Das Riemannsche Integral $\int_a^b$ ist eine Linearform auf $\mathrm{Riem}([a, b], \mathbb{R})$.

Bemerkung 4.1.50. Sind $f\colon V \to W$ und $g\colon W \to U$ lineare Abbildungen, so gilt

$$(g \circ f)^* = f^* \circ g^*\colon U^* \to V^*.$$

Man kann auch leicht nachprüfen, dass die Abbildung

$$\begin{aligned} \mathrm{Hom}_K(V, W) &\to \mathrm{Hom}_K(W^*, V^*), \\ f &\mapsto f^*, \end{aligned}$$

linear ist.

Lemma 4.1.51. *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen.*

(i) *Ist f injektiv, so ist f^* surjektiv.*

(ii) *Ist f surjektiv, so ist f^* injektiv.*

Beweis. Sei $(v_i)_{i\in I}$ eine Basis von V. Ist f injektiv, so ist die Familie $(f(v_i))_{i\in I}$ linear unabhängig (Proposition 4.1.20(ii)) und kann zu einer Basis von W ergänzt werden (Satz 3.3.20(iii)). Nach Satz 4.1.22 kann dann jede lineare Abbildung $V \to K$ zu einer linearen Abbildung $W \to K$ fortgesetzt werden, d.h., f^* ist surjektiv. Ist f surjektiv und ist $\alpha \circ f = 0$, so muss α null sein, d.h., f^* ist injektiv. □

Proposition 4.1.52 (Kern und Bild dualer Abbildungen)**.** *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen.*

(i) *Die duale Abbildung der Quotientenabbildung $q\colon W \twoheadrightarrow W/\operatorname{im} f$ induziert einen Isomorphismus*

$$(W/\operatorname{im} f)^* \xrightarrow{\sim} \ker(f^*).$$

(ii) *Die duale Abbildung der Inklusionsabbildung $i\colon \ker f \hookrightarrow V$ induziert einen Isomorphismus*

$$V^*/\operatorname{im}(f^*) \xrightarrow{\sim} (\ker f)^*.$$

(iii) *Die duale Abbildung der von f induzierten Abbildung $\bar{f}\colon V \twoheadrightarrow \operatorname{im} f$ induziert einen Isomorphismus*

$$(\operatorname{im} f)^* \xrightarrow{\sim} \operatorname{im}(f^*).$$

Beweis. Zu (i). Nach Lemma 4.1.51(ii) ist die Abbildung $q^*\colon (W/\operatorname{im} f)^* \to W^*$ injektiv. Es bleibt zu zeigen, dass $\operatorname{im}(q^*) = \ker(f^*)$. Aus $q \circ f = 0$ folgt $f^* \circ q^* = (q \circ f)^* = 0$ und somit $\operatorname{im}(q^*) \subset \ker f^*$. Sei umgekehrt $\alpha \in \ker(f^*) \subset W^*$, d.h., $\alpha \circ f = 0$. Nach der universellen Eigenschaft der Quotientenvektorraum existiert $\bar{\alpha}\colon W/\operatorname{im} f \to K$ so dass $\bar{\alpha} \circ q = \alpha$, d.h., $q^*(\bar{\alpha}) = \alpha$. Insbesondere ist α im Bild von q^*.

Zu (ii). Nach Lemma 4.1.51(i) ist die Abbildung $i^*\colon V^* \to (\ker f)^*$ surjektiv. Nach Satz 4.1.35 bleibt es zu zeigen, dass $\ker(i^*) = \operatorname{im}(f^*)$. Aus $f \circ i = 0$ folgt $i^* \circ f^* = 0$ und somit $\operatorname{im}(f^*) \subset \ker(i^*)$. Sei umgekehrt $\alpha \in \ker(i^*) \subset V^*$, d.h., $\alpha \circ i = 0$. Nach der universellen Eigenschaft der Quotientenvektorraum ist α im Bild von r^*, wobei $r\colon V \to V/\ker f$ die Quotientenabbildung ist. Nach Satz 4.1.35 ist $f = j \circ r$ mit einer injektiven linearen Abbildung $j\colon V/\ker f \hookrightarrow W$. Nach Lemma 4.1.51(i) ist j^* surjektiv, und damit ist α im Bild von f^*.

Zu (iii). Nach Lemma 4.1.51(ii) ist die Abbildung $\bar{f}^*\colon (\operatorname{im} f)^* \to V^*$ injektiv. Es bleibt zu zeigen, dass $\operatorname{im}(\bar{f}^*) = \operatorname{im}(f^*)$. Es gilt $f = u \circ \bar{f}$, wobei $u\colon \operatorname{im} f \hookrightarrow W$ die Inklusionsabbildung ist. Daraus folgt $f^* = \bar{f}^* \circ u^*$. Nach Lemma 4.1.51(i) ist u^* surjektiv, und damit ist $\operatorname{im}(f^*) = \operatorname{im}(\bar{f}^*)$, wie gewünscht. □

Bemerkung 4.1.53. Sei $f\colon V \to W$ eine lineare Abbildung. Der Quotientenvektorraum $W/\operatorname{im} f$ heißt der *Kokern* von f und wird mit $\operatorname{coker} f$ bezeichnet. Aussagen (i) und (ii) der Proposition 4.1.52 sagen insbesondere, dass $\ker(f^*) \cong (\operatorname{coker} f)^*$ und $\operatorname{coker}(f^*) \cong (\ker f)^*$. Das heißt, die Dualität vertauscht den Kern und den Kokern, aber sie erhält das Bild.

Konstruktion 4.1.54 (duale Basis)**.** Sei V ein K-Vektorraum und $B = (v_i)_{i\in I}$ eine Basis von V. Nach der universellen Eigenschaft von Basen (Satz 4.1.22) gibt es zu jedem $i \in I$ genau eine Linearform $v_i^* \in V^*$, so dass

$$v_i^*(v_j) = \begin{cases} 1, & \text{falls } j = i, \\ 0, & \text{andernfalls.} \end{cases}$$

(Trotz der Notation hängt die Linearform v_i^* nicht nur von v_i ab, sondern von der ganzen Basis B!) Nach demselben Satz gibt es dann genau eine lineare Abbildung

$$\begin{aligned} \varepsilon_B \colon V &\to V^*, \\ v_i &\mapsto v_i^*. \end{aligned}$$

Die nächste Proposition zeigt, dass die Familie $B^* = (v_i^*)_{i\in I}$ in V^* immer linear unabhängig ist. Sie ist sogar eine Basis von V^*, wenn V endlich-dimensional ist. In diesem Fall heißt die Basis B^* von V^* die *duale Basis* zu B.

Beispiel 4.1.55. Sei $B = (e_1, \dots, e_n)$ die Standardbasis von K^n. Die duale Basis von $(K^n)^*$ ist $B^* = (\pi_1, \dots, \pi_n)$.

Proposition 4.1.56. *Sei V ein K-Vektorraum und $B = (v_i)_{i\in I}$ eine Basis von V.*

(i) *Die lineare Abbildung $\varepsilon_B \colon V \to V^*$ ist injektiv.*

(ii) *Ist V endlich-dimensional, so ist $\varepsilon_B \colon V \to V^*$ ein Isomorphismus.*

Beweis. Zu (i). Nach Satz 4.1.22 genügt es zu zeigen, dass die Familie $(v_i^*)_{i\in I}$ linear unabhängig ist. Sei $\sum_{i\in I} \lambda_i \cdot v_i^* = 0$ mit $(\lambda_i)_{i\in I} \in K^{(I)}$. Nach Definition von v_i^* ist der Wert von $\sum_{i\in I} \lambda_i \cdot v_i^*$ in v_j gleich λ_j, und damit sind alle λ_i gleich 0.

Zu (ii). In diesem Fall ist I endlich (Satz 3.3.27(i)). Sei $\alpha \in V^*$ beliebig. Die lineare Abbildungen α und $\sum_{i\in I} \alpha(v_i) \cdot v_i^*$ haben denselben Wert in v_j für alle $j \in I$. Nach Satz 4.1.22 stimmen sie deshalb auf ganz V überein. Dies zeigt, dass α im Bild von ε_B liegt, und damit dass ε_B surjektiv ist. □

Beispiel 4.1.57. Sei I eine Menge. Aus dem Satz 4.1.22 folgt, dass die Abbildung

$$\begin{aligned} (K^{(I)})^* &\to K^I, \\ \alpha &\mapsto (\alpha(e_i))_{i\in I}, \end{aligned}$$

ein Isomorphismus ist. Für die Basis $B = (e_i)_{i\in I}$ von $K^{(I)}$ ist dann die Komposition von $\varepsilon_B \colon K^{(I)} \hookrightarrow (K^{(I)})^*$ mit diesem Isomorphismus die Inklusionsabbildung $K^{(I)} \hookrightarrow K^I$.

Korollar 4.1.58 (Rang der dualen Abbildung)**.** *Sei $f \colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Ist* $\operatorname{im} f$ *endlich-dimensonal (z.B. ist V oder W endlich-dimensional), so gilt*

$$\operatorname{rg} f = \operatorname{rg} f^*.$$

Beweis. Nach Proposition 4.1.52 gilt

$$\dim_K \operatorname{im}(f^*) = \dim_K (\operatorname{im} f)^*.$$

Wenn V oder W endlich-dimensional ist, ist auch $\operatorname{im} f$ endlich-dimensional nach Korollar 4.1.38 oder Proposition 3.3.35. Nach Proposition 4.1.56(ii) gilt dann

$$\dim_K (\operatorname{im} f)^* = \dim_K \operatorname{im} f.$$

Aus diesen zwei Formeln folgt, dass $\operatorname{rg} f = \operatorname{rg} f^*$. □

Zur Erinnering gibt es zu jeder Menge X und jedem Element $x \in X$ eine Auswertungsabbildung

$$\begin{aligned} \mathrm{ev}_x \colon \mathrm{Abb}(X, K) &\to K, \\ f &\mapsto f(x), \end{aligned}$$

die K-linear ist (Beispiel 4.1.8). Wenn $X = V$ ein K-Vektorraum ist, ist insbesondere die Einschränkung von ev_x auf V^* eine Linearform auf V^*, also ein Element des Dualraums $(V^*)^*$, das wir auch mit ev_v bezeichnen.

Proposition 4.1.59 (Doppeldual)**.** *Sei V ein K-Vektorraum. Die Abbildung*

$$\begin{aligned} \mathrm{ev} \colon V &\to (V^*)^*, \\ v &\mapsto \mathrm{ev}_v \end{aligned}$$

ist K-linear und injektiv. Falls V endlich-dimensional ist, ist ev *ein Isomorphismus.*

Beweis. Zur Linearität haben wir zu zeigen:

$$\mathrm{ev}_{v+w} = \mathrm{ev}_v + \mathrm{ev}_w \quad \text{und} \quad \mathrm{ev}_{\lambda \cdot v} = \lambda \cdot \mathrm{ev}_v \,.$$

Dies folgt unmittelbar aus den Definitionen. Zum Beispiel gilt für alle $\alpha \in V^*$:

$$\mathrm{ev}_{\lambda \cdot v}(\alpha) = \alpha(\lambda \cdot v) = \lambda \cdot \alpha(v) = \lambda \cdot \mathrm{ev}_v(\alpha) = (\lambda \cdot \mathrm{ev}_v)(\alpha).$$

Zur Injektivität: Sei $v \in V \setminus \{0\}$. Nach dem Basisergänzungssatz 3.3.20 können wir v zu einer Basis B von V ergänzen. Für die Abbildung $\varepsilon_B \colon V \to V^*$ gilt dann nach Konstruktion $\varepsilon_B(v)(v) = 1 \neq 0$. Insbondere existiert $\alpha \in V^*$, so dass $\mathrm{ev}_v(\alpha) = \alpha(v) \neq 0$, und damit ist $\mathrm{ev}_v \neq 0$. Also ist ev injektiv.

Wenn V endlich-dimensional ist, dann gilt $\dim_K(V) = \dim_K(V^*) = \dim_K((V^*)^*)$ nach Proposition 4.1.56(ii). Die letzte Aussage folgt daraus, da jede injektive lineare Abbildung zwischen endlich-dimensionalen Vektorräumen derselben Dimension ein Isomorphismus ist (Korollar 4.1.39). □

Bemerkung 4.1.60. Sei V ein endlich-dimensional K-Vektorraum. Man beachte, dass der Isomorphismus $\varepsilon_B \colon V \to V^*$ von der Wahl der Basis B abhängt. Da im Allgemeinen V keine bevorzugte Basis besitzt, gibt es keinen bevorzugten Isomorphismus zwischen V und seinem Dualraum. Im Gegensatz dazu ist der Isomorphismus $\mathrm{ev} \colon V \to (V^*)^*$ unabhängig von irgendwelchen Wahlen. Deswegen wird es oft gesagt, dass ein endlich-dimensionaler Vektorraum zu seinem Doppeldual *kanonisch* isomorph ist, aber zu seinem Dual nur unkanonisch.

Bemerkung 4.1.61. Falls V unendlich-dimensional ist, dann sind weder $\varepsilon_B \colon V \to V^*$ noch $\mathrm{ev} \colon V \to (V^*)^*$ surjektiv. Man kann sogar zeigen, dass die Mächtigkeit einer Basis von V^* wirklich größer als die einer Basis von V ist. Insbesondere ist die Bijektivität von ev, oder die Existenz eines Isomorphismus zwischen V und V^*, eine *Charakterisierung* der Endlichdimensionalität eines Vektorraums V.

4.2 Matrizen

Das Ziel dieses Abschnitts ist, lineare Abbildungen zwischen Vektorräumen der Gestalt K^n ganz konkret zu verstehen. Dazu führen wir den Beriff der *Matrix* ein.

Definition 4.2.1 (Matrix)**.** Sei K ein Körper und seien $m, n \in \mathbb{N}$. Eine *$m \times n$-Matrix über K* (oder *mit Koeffizienten in K*) ist eine Familie

$$A = (a_{ij})_{(i,j) \in \{1,\dots,m\} \times \{1,\dots,n\}}$$

mit $a_{ij} \in K$. Wenn m und n festgelegt sind, schreibt man oft $(a_{ij})_{i,j}$ oder sogar (a_{ij}) für eine solche Familie. Der Skalar a_{ij} ist der (i,j)-te *Koeffizient* oder *Eintrag* der Matrix A, und wird auch mit A_{ij} bezeichnet.

Eine $m \times n$-Matrix $A = (a_{ij})_{i,j}$ wird üblicherweise als Rechteck mit m Zeilen und n Spalten dargestellt:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Inbesondere betrachten wir $m \times 1$-Matrizen als Spaltenvektoren und $1 \times n$-Matrizen als Zeilenvektoren.

Notation 4.2.2. Die Menge aller $m \times n$-Matrizen über K wird mit $M_{m\times n}(K)$ bezeichnet. Da wir schon Elemente von K^m als Spaltenvektoren betrachten, werden wir üblicherweise $M_{m\times 1}(K)$ mit K^m identifizieren. Wir können außerdem $M_{1\times 1}(K)$ mit K identifizieren. Wenn $m = n$ schreiben wir auch $M_n(K)$ anstelle von $M_{n\times n}(K)$. Matrizen in $M_n(K)$ heißen *quadratische* Matrizen.

Notation 4.2.3. Seien $A = (a_{ij})_{i,j} \in M_{m\times n}(K)$, $i \in \{1, \dots, m\}$ und $j \in \{1, \dots, n\}$. Wir bezeichnen mit

$$A_{i*} = \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$$

die i-te Zeile von A und mit

$$A_{*j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

die j-te Spalte von A.

Matrizen fester Größe bilden einen Vektorraum über K, indem wir Addition und Skalarmultiplikation koeffizientenweise definieren (vgl. Definition 3.1.2):

Definition 4.2.4 (Addition und Skalarmultiplikation von Matrizen)**.** Seien $m, n \in \mathbb{N}$.

- Seien $A = (a_{ij})_{i,j}$ und $B = (b_{ij})_{i,j}$ $m \times n$-Matrizen über K. Ihre Summe ist die Matrix

$$A + B := (a_{ij} + b_{ij})_{i,j}.$$

- Sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K und $\lambda \in K$ ein Skalar. Dann setzen wir

$$\lambda \cdot A := (\lambda \cdot a_{ij})_{i,j}.$$

Mit dieser Addition und Skalarmultiplikation ist es klar, dass $M_{m\times n}(K)$ ein Vektorraum über K ist (siehe Proposition 3.1.5). Der Nullvektor ist die *Nullmatrix* $0 = 0_{m,n} = (0)_{i,j}$, und das Inverse einer Matrix $A = (a_{ij})_{i,j}$ bzgl. $+$ ist $-A = (-a_{ij})_{i,j}$. Außerdem hat $M_{m\times n}(K)$ eine Basis $(E_{rs})_{(r,s)\in\{1,\dots,m\}\times\{1,\dots,n\}}$, wobei E_{rs} die Matrix ist, deren Einträge alle null sind, außer dem Eintrag in der r-ten Zeile und s-ten Spalte, welcher 1 ist. Insbesondere gilt:

$$\dim_K M_{m\times n}(K) = m \cdot n.$$

Um die Definition von E_{rs} als Formel schreiben zu können, ist folgende Notation hilfreich:

Notation 4.2.5 (das Kronecker-Delta)**.**

$$\delta_{ij} := \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{falls } i \neq j. \end{cases}$$

Mithilfe des Kronecker-Deltas kann man schreiben: $E_{rs} = (\delta_{ir}\delta_{js})_{i,j}$.

Bemerkung 4.2.6. Die Addition von Matrizen unterschiedlicher Größe ist *nicht* definiert.

Definition 4.2.7 (transponierte Matrix)**.** Seien $m, n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K. Die *transponierte Matrix* zu A ist die $n \times m$-Matrix

$$A^\intercal := (a_{ij})_{j,i} \in M_{n\times m}(K).$$

Beispiel 4.2.8. Für die 3×2-Matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

gilt

$$A^\intercal = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Durch das Transponieren werden also Zeilen zu Spalten und umgekehrt. Die Abbildung

$$\begin{aligned} M_{m\times n}(K) &\to M_{n\times m}(K), \\ A &\mapsto A^\intercal, \end{aligned}$$

ist offensichtlich K-linear, und es gilt $(A^\intercal)^\intercal = A$.

Definition 4.2.9 (Zeilenraum, Spaltenraum)**.** Seien $m, n \in \mathbb{N}$ und sei $A \in M_{m\times n}(K)$.

- Der *Zeilenraum* $\mathrm{ZR}(A)$ von A ist der von den Zeilen von A erzeugte Untervektorraum von K^n.
- Der *Spaltenraum* $\mathrm{SR}(A)$ von A ist der von den Spalten von A erzeugte Untervektorraum von K^m.

Bemerkung 4.2.10. Es gilt $\mathrm{ZR}(A) = \mathrm{SR}(A^\intercal)$ und $\mathrm{SR}(A) = \mathrm{ZR}(A^\intercal)$.

Definition 4.2.11 (Diagonalmatrix, obere/untere Dreiecksmatrix)**.** Sei $n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j}$ eine $n \times n$-Matrix über K.

- A heißt *Diagonalmatrix*, wenn alle Koeffizienten außerhalb der Hauptdiagonale null sind, d.h., wenn $a_{ij} = 0$ für alle $i \neq j$.
- A heißt *obere Dreiecksmatrix*, wenn $a_{ij} = 0$ für alle $i > j$.
- A heißt *untere Dreicksmatrix*, wenn $a_{ij} = 0$ für alle $i < j$.
- A heißt *Dreiecksmatrix*, wenn A eine obere oder untere Dreiecksmatrix ist.

Notation 4.2.12. Man schreibt $\mathrm{diag}(d_1, \ldots, d_n)$ für die $n \times n$-Diagonalmatrix $(a_{ij})_{i,j}$ mit $a_{ii} = d_i$.

4.2.1 Multiplikation von Matrizen

Definition 4.2.13 (Matrixmultiplikation)**.** Seien $m, n, p \in \mathbb{N}$. Sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K und $B = (b_{jk})_{j,k}$ eine $n \times p$-Matrix über K. Das *Produkt* von A und B ist die $m \times p$-Matrix

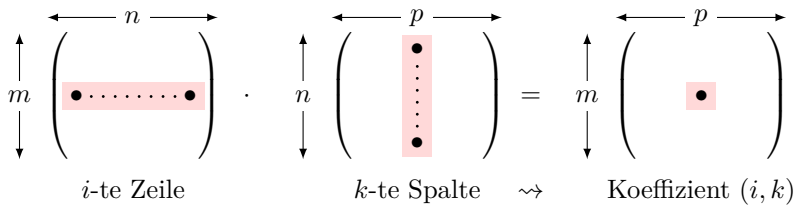
$$A \cdot B := \left(\sum_{j=1}^{n} a_{ij} \cdot b_{jk} \right)_{i,k}.$$

Um diese Definition zu verstehen ist es hilfreich, den Spezialfall $m = p = 1$ zu betrachten. In diesem Fall ist $A \in M_{1\times n}(K)$ ein Zeilenvektor und $B \in M_{n\times 1}(K)$ ein Spaltenvektor, und ihr Produkt liegt in $M_{1\times 1}(K)$, also ist ein einziger Skalar:

$$A \cdot B = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{j=1}^{n} a_j b_j \in K.$$

Im allgemeinen Fall mit $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$ ist der (i,k)-te Eintrag in dem Produkt $A \cdot B$ das Produkt der i-te Zeile von A mit der k-te Spalte von B wie im Spezialfall:

$$A \cdot B = (A_{i*} \cdot B_{*k})_{i,k}.$$



Beispiel 4.2.14. Es gilt

$$\begin{pmatrix} 1 & -1 & 2 \\ 5 & 0 & -3 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ -2 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ -3 & -1 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 1 \\ -2 & 4 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 2 \\ 5 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 5 & 0 & -3 \\ 18 & 2 & -16 \\ 11 & -1 & -4 \end{pmatrix}.$$

Definition 4.2.15 (Einheitsmatrix)**.** Sei $n \in \mathbb{N}$. Die $n \times n$-*Einheitsmatrix* über K ist die $n \times n$-Matrix

$$I_n := (\delta_{ij})_{i,j} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in M_n(K).$$

Dabei ist δ_{ij} das Kronecker-Delta, siehe Notation 4.2.5.

Proposition 4.2.16 (Eigenschaften der Matrixmultiplikation)**.**

(i) *Matrixmultiplikation ist assoziativ. Das heißt, für alle $m, n, p, q \in \mathbb{N}$, $A \in M_{m\times n}(K)$, $B \in M_{n\times p}(K)$ und $C \in M_{p\times q}(K)$ gilt*

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

(ii) *Die Einheitsmatrix ist ein neutrales Element bzgl. Matrixmultiplikation. Das heißt, für alle $m, n \in \mathbb{N}$ und $A \in M_{m\times n}(K)$ gilt*

$$I_m \cdot A = A \quad \text{und} \quad A \cdot I_n = A.$$

(iii) *Matrixmultiplikation ist distributiv über Addition. Das heißt, für alle $m, n, p \in \mathbb{N}$, $A, B \in M_{m\times n}(K)$ und $C, D \in M_{n\times p}(K)$ gilt*

$$A \cdot (C + D) = A \cdot C + A \cdot D \quad \text{und} \quad (A + B) \cdot C = A \cdot C + B \cdot C.$$

(iv) *Matrixmultiplikation ist kompatibel mit Skalarmultiplikation im folgenden Sinne: Für alle* $m, n, p \in \mathbb{N}$, $A \in M_{m\times n}(K)$, $B \in M_{n\times p}(K)$ *und* $\lambda \in K$ *gilt*

$$(\lambda \cdot A) \cdot B = \lambda \cdot (A \cdot B) \quad \textit{und} \quad A \cdot (\lambda \cdot B) = \lambda \cdot (A \cdot B).$$

(v) *Matrixmultiplikation ist kompatibel mit Transponieren im folgenden Sinne: Für alle* $m, n, p \in \mathbb{N}$, $A \in M_{m\times n}(K)$ *und* $B \in M_{n\times p}(K)$ *gilt*

$$(A \cdot B)^\intercal = B^\intercal \cdot A^\intercal.$$

Beweis. Zu (i). Nach Definition der Matrixmultiplikaton gilt

$$(A \cdot B)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk} \quad \text{und} \quad (B \cdot C)_{jl} = \sum_{k=1}^{p} B_{jk}C_{kl},$$

und daher

$$\begin{aligned}
((A \cdot B) \cdot C)_{il} &= \sum_{k=1}^{p} (A \cdot B)_{ik}C_{kl} \\
&= \sum_{k=1}^{p} \left(\sum_{j=1}^{n} A_{ij}B_{jk} \right) C_{kl} \\
&= \sum_{k=1}^{p} \sum_{j=1}^{n} A_{ij}B_{jk}C_{kl} \\
&= \sum_{j=1}^{n} \sum_{k=1}^{p} A_{ij}B_{jk}C_{kl} \\
&= \sum_{j=1}^{n} A_{ij} \left(\sum_{k=1}^{p} B_{jk}C_{kl} \right) \\
&= \sum_{j=1}^{n} A_{ij}(B \cdot C)_{jl} \\
&= (A \cdot (B \cdot C))_{il}.
\end{aligned}$$

Dabei haben wir mehrere Eigenschaften von $+$ und $\cdot$ im Körper K verwendet: das verallgemeinerte Disbtributivgesetz (Bemerkung 2.3.3), die Assoziativität und Kommutativität von $+$ und die Assoziativität von $\cdot$. Diese Berechnung gilt für alle $i \in \{1, \ldots, m\}$ und $l \in \{1, \ldots, q\}$, und damit ist die gewünschte Matrixgleichung bewiesen.

Zu (ii). Für alle i, j gilt

$$(I_m \cdot A)_{ij} = \sum_{e=1}^{m} \delta_{ie}A_{ej} = A_{ij},$$

und somit $I_m \cdot A = A$. Der Beweis der Gleichung $A \cdot I_n = A$ ist ähnlich.

Zu (iii). Wir überprüfen nur die erste Gleichung:

$$\begin{aligned}
(A \cdot (C+D))_{ik} &= \sum_{j=1}^{n} A_{ij}(C+D)_{jk} && \text{(Definition der Matrixmultiplikation)}\\
&= \sum_{j=1}^{n} A_{ij}(C_{jk}+D_{jk}) && \text{(Definition der Matrixaddition)}\\
&= \sum_{j=1}^{n} (A_{ij}C_{jk} + A_{ij}D_{jk}) && \text{(Distributivgesetz in } K)\\
&= \sum_{j=1}^{n} A_{ij}C_{jk} + \sum_{j=1}^{n} A_{ij}D_{jk} && \text{(Assoz. \& Komm. der Addition in } K)\\
&= (A \cdot C)_{ik} + (A \cdot D)_{ik} && \text{(Definition der Matrixmultiplikation)}\\
&= ((A \cdot C) + (A \cdot D))_{ik}. && \text{(Definition der Matrixaddition)}
\end{aligned}$$

Zu (iv). Wir überprüfen nur die zweite Gleichung:

$$\begin{aligned}
(A \cdot (\lambda \cdot B))_{ik} &= \sum_{j=1}^{n} A_{ij}(\lambda \cdot B)_{jk} && \text{(Definition der Matrixmultiplikation)}\\
&= \sum_{j=1}^{n} A_{ij}(\lambda B_{jk}) && \text{(Def. der Skalarmultiplaktion von Matrizen)}\\
&= \sum_{j=1}^{n} \lambda(A_{ij}B_{jk}) && \text{(Assoz. \& Komm. der Multiplikation in } K)\\
&= \lambda \sum_{j=1}^{n} A_{ij}B_{ij} && \text{(Distributivgesetz in } K)\\
&= \lambda(A \cdot B)_{ik} && \text{(Definition der Matrixmultiplikation)}\\
&= (\lambda \cdot (A \cdot B))_{ik}. && \text{(Def. der Skalarmultiplaktion von Matrizen)}
\end{aligned}$$

Zu (v). Für alle $i \in \{1, \ldots, m\}$ und $k \in \{1, \ldots, p\}$ gilt

$$((A \cdot B)^{\intercal})_{ki} = (A \cdot B)_{ik} = \sum_{j=1}^{n} A_{ij} \cdot B_{jk} = \sum_{j=1}^{n} (B^{\intercal})_{kj} \cdot (A^{\intercal})_{ji} = (B^{\intercal} \cdot A^{\intercal})_{ki}. \qquad \square$$

Bemerkung 4.2.17. Sei $n \in \mathbb{N}$. Die Matrixmultiplikation definiert eine Verknüpfung

$$\cdot \colon M_n(K) \times M_n(K) \to M_n(K)$$

auf quadratischen $n \times n$-Matrizen. Diese Verknüpfung ist assoziativ nach Proposition 4.2.16(i) und distributiv über die Addition nach Proposition 4.2.16(iii). Wenn $n \geq 2$ ist sie aber nicht kommutativ. Zum Beispiel:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Das Tripel $(M_n(K), +, \cdot)$ ist ein Beispiel eines *nicht-kommutativen Ringes* (siehe Bemerkung 2.3.5).

Notation 4.2.18 (Matrixpotenzen)**.** Sei $n \in \mathbb{N}$ und $A \in M_n(K)$ eine quadratische Matrix. Man definiert die Potenzen A^k mit $k \in \mathbb{N}$ rekursiv wie folgt:

$$\begin{aligned}
&A^0 = I_n,\\
&A^n = A \cdot A^{n-1} \quad \text{(für alle } n \geq 1).
\end{aligned}$$

Wegen der Assoziativität der Matrixmultiplikation (und die Neutralität von I_n) gilt dann $A^1 = A$ und

$$A^{k+l} = A^k \cdot A^l \quad \text{und} \quad (A^l)^k = A^{kl}$$

für alle $k, l \in \mathbb{N}$ (beide Formeln lassen sich leicht durch Induktion über k nachprüfen, vgl. Proposition 2.1.7).

Beispiel 4.2.19 (Fibonacci-Zahlen)**.** Die Folge $(F_n)_{n\in\mathbb{N}}$ der *Fibonacci-Zahlen* wird durch folgende Gleichungen rekursiv definiert:

$$\begin{aligned} F_0 &= 0, \\ F_1 &= 1, \\ F_{n+1} &= F_n + F_{n-1} \quad \text{(für alle } n \geq 1\text{)}. \end{aligned}$$

Die ersten Fibonacci-Zahlen sind also

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

Diese rekursive Definition kann mithilfe der Matrixmultiplikation ausgedrückt werden:

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \begin{pmatrix} F_n + F_{n-1} \\ F_n \end{pmatrix} = A \cdot \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}, \quad \text{wobei } A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \in M_{2\times 2}(\mathbb{Q}).$$

Daraus folgt die folgende nicht-rekursive Formel für die Fibonacci-Zahlen:

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = A^n \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Damit stellt sich noch die Frage, ob man die Matrixpotenz A^n effektiv berechnen kann. Das werden wir später schaffen, und damit eine explizite Formel für F_n erhalten (siehe Beispiel 6.2.33).

Definition 4.2.20 (invertierbare Matrix, inverse Matrix)**.** Sei $n \in \mathbb{N}$. Eine quadratische Matrix $A \in M_n(K)$ heißt *invertierbar* oder *regulär*, wenn sie ein inverses Element bezüglich Matrixmultiplikation besitzt, d.h., wenn eine Matrix $B \in M_n(K)$ existiert, so dass

$$A \cdot B = I_n \quad \text{und} \quad B \cdot A = I_n.$$

Die Matrix B ist dann eindeutig bestimmt (siehe Proposition 2.1.3); sie heißt die *inverse Matrix* zu A und wird mit A^{-1} bezeichnet.

Wir werden später untersuchen, wie man die Invertierbarkeit einer Matrix bestimmen kann und wie man die inverse Matrix berechnen kann (siehe Abschnitt 5.2.1).

Beispiel 4.2.21.

(i) Eine 1×1-Matrix über K entspricht einem einzigen Skalar $a \in K$. Sie ist genau dann invertierbar, wenn $a \in K^\times$.

(ii) Für die Matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

gilt $A^2 = I_2$. Also ist A invertierbar mit $A^{-1} = A$.

(iii) Für $a, b \in K$ gilt

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}.$$

Daraus folgt, dass

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix}.$$

Bemerkung 4.2.22. Sind $A, B \in M_n(K)$ invertierbar, so ist $A \cdot B$ invertierbar und es gilt

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}.$$

Definition 4.2.23 (allgemeine lineare Gruppe)**.** Sei $n \in \mathbb{N}$. Die *allgemeine lineare Gruppe* über K ist die Menge

$$\mathrm{GL}_n(K) = \{A \in M_n(K) \mid A \text{ ist invertierbar}\}$$

versehen mit der Matrixmultiplikation.

Bemerkung 4.2.24. Wenn $n \geq 2$ ist die allgemeine lineare Gruppe $\mathrm{GL}_n(K)$ nicht abelsch (nach Bemerkung 4.2.17).

4.2.2 Lineare Abbildungen aus Matrizen

Sei A eine $m \times n$-Matrix über K. Man definiert eine Abbildung $L_A \colon K^n \to K^m$ wie folgt:

$$\begin{aligned} L_A \colon K^n &\to K^m, \\ v &\mapsto A \cdot v. \end{aligned}$$

Dabei betrachten wir v als Spaltenvektor, d.h., $v \in M_{n\times 1}(K)$, und $A \cdot v \in M_{m\times 1}$ ist das Matrixprodukt von A mit v. Konkreter, ist $A = (a_{ij})_{i,j}$, so ist

$$L_A(v) = A \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + a_{m3}v_3 + \cdots + a_{mn}v_n \end{pmatrix}.$$

Bemerkung 4.2.25 (Bild der Standardeinheitsvektoren)**.** Seien $e_1, \ldots, e_n \in K^n$ die Standardeinheitsvektoren und sei $A \in M_{m\times n}(K)$. Für alle $j \in \{1, \ldots, n\}$ gilt

$$L_A(e_j) = A \cdot e_j = A_{*j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Das heißt: *Das Bild von e_j unter L_A ist die j-te Spalte von A.*

Proposition 4.2.26. *Seien $m, n, p \in \mathbb{N}$.*

(i) *Für alle $A \in M_{m\times n}(K)$ ist die Abbildung $L_A \colon K^n \to K^m$ K-linear.*

(ii) *Für alle $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$ gilt $L_{A\cdot B} = L_A \circ L_B$.*

(iii) *Es gilt $L_{I_n} = \mathrm{id}_{K^n}$.*

(iv) *Für alle $A, B \in M_{m\times n}(K)$ gilt $L_{A+B} = L_A + L_B$.*

(v) *Für alle $A \in M_{m\times n}(K)$ und $\lambda \in K$ gilt $L_{\lambda\cdot A} = \lambda \cdot L_A$.*

Beweis. Zur ersten Aussage ist zu zeigen: Für alle $v, v' \in K^n$ und $\lambda \in K$ gelten

$$L_A(v + v') = L_A(v) + L_A(v') \quad \text{und} \quad L_A(\lambda \cdot v) = \lambda \cdot L_A(v).$$

Dies folgt aus der Definition von L_A und Proposition 4.2.16(iii,iv). Aussagen (ii) bis (v) folgen direkt aus Aussagen (i) bis (iv) der Proposition 4.2.16. □

Bemerkung 4.2.27. Nach Bemerkung 4.2.25 und Proposition 4.1.20(i) ist der Spaltenraum von $A \in M_{m\times n}(K)$ genau das Bild von L_A:

$$\mathrm{SR}(A) = \operatorname{im} L_A \subset K^m.$$

Der Zusammenhang zwischen L_A und dem Zeilenraum von A ist nicht so offensichtlich (eigentlich ist $\mathrm{ZR}(A)$ das orthogonale Komplement von $\ker L_A$ in K^n).

Bemerkung 4.2.28. Aussagen (i), (iv) und (v) der Proposition 4.2.26 können wie folgt zusammengefasst werden: Es gibt eine K-lineare Abbildung

$$\begin{aligned} L\colon M_{m\times n}(K) &\to \mathrm{Hom}_K(K^n, K^m), \\ A &\mapsto L_A. \end{aligned}$$

Dabei ist $\mathrm{Hom}_K(K^n, K^m)$ der K-Vektorraum aller K-linearen Abbildungen von K^n nach K^m (siehe Proposition 4.1.43).

Satz 4.2.29. *Seien $m, n \in \mathbb{N}$ und sei $f\colon K^n \to K^m$ eine K-lineare Abbildung. Dann existiert genau eine $m \times n$-Matrix A über K, so dass $f = L_A$. Anders gesagt ist die Abbildung*

$$L\colon M_{m\times n}(K) \to \mathrm{Hom}_K(K^n, K^m)$$

ein Isomorphismus.

Beweis. Dass höchstens eine solche Matrix A existiert folgt aus Bemerkung 4.2.25: Wenn $L_A = L_B$, dann ist inbesondere

$$A_{*j} = L_A(e_j) = L_B(e_j) = B_{*j}$$

für alle $j \in \{1, \dots, n\}$, und daher ist $A = B$. Es bleibt zu zeigen, dass A existiert. Sei A die $m \times n$-Matrix, deren j-te Spalte gleich $f(e_j) \in K^m$ ist. Wir behaupten, dass $L_A = f$. Beide Abbildungen L_A und f sind K-linear (die erste nach Proposition 4.2.26(i)). Nach Bemerkung 4.2.25 gilt $L_A(e_j) = f(e_j)$ für alle $j \in \{1, \dots, n\}$. Da $(e_1, \dots, e_n)$ eine Basis von K^n ist, folgt aus dem Satz 4.1.22, dass $L_A = f$. □

Notation 4.2.30. Die Umkehrabbildung von L bezeichnen wir mit

$$M\colon \mathrm{Hom}_K(K^n, K^m) \xrightarrow{\sim} M_{m\times n}(K).$$

Das heißt, ist $f\colon K^n \to K^m$ eine linear Abbildung, so ist $M(f)$ die $m\times n$-Matrix mit Spalten $f(e_1)$, …, $f(e_n)$.

Bemerkung 4.2.31. Nach Proposition 4.2.26(ii) ist der Isomorphismus M auch mit den multiplikativen Verknüpfungen verträglich: Für alle $m, n, p \in \mathbb{N}$ ist folgendes Diagramm kommutativ:

$$\begin{array}{ccc} \mathrm{Hom}_K(K^n, K^m) \times \mathrm{Hom}_K(K^p, K^n) & \xrightarrow{\circ} & \mathrm{Hom}_K(K^p, K^m) \\ \downarrow{\scriptstyle M\times M} & & \downarrow{\scriptstyle M} \\ M_{m\times n}(K) \times M_{n\times p}(K) & \xrightarrow{\cdot} & M_{m\times p}(K). \end{array}$$

Beispiel 4.2.32. Wir listen die Matrizen $A \in M_2(\mathbb{R})$ auf, deren zugeordnete $\mathbb{R}$-lineare Abbildungen $L_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ die im Beispiel 4.1.7 sind:

(i) Skalierung um $\frac{1}{2}$: $\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$.

(ii) Spiegelung an $\mathbb{R}e_2$: $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

(iii) Drehung um $\frac{\pi}{4}$: $\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$.

(iv) Spiegelung an 0: $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.

(v) Horizontale Skalierung um $\frac{3}{2}$: $\begin{pmatrix} \frac{3}{2} & 0 \\ 0 & 1 \end{pmatrix}$.

(vi) Koordinatenvertauschung: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(vii) Orthogonale Projektion auf $\mathbb{R}e_1$: $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

(viii) Horizontale Scherung um $\frac{1}{2}$: $\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix}$.

Beispiel 4.2.33 (Drehmatrizen)**.** Sei $\alpha \in \mathbb{R}$ eine reelle Zahl und sei

$$D(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \in M_2(\mathbb{R}).$$

Die entsprechende lineare Abbildung $L_{D(\alpha)} \colon \mathbb{R}^2 \to \mathbb{R}^2$ ist eine Drehung um den Winkel α um den Nullpunkt. Insbesondere gilt $L_{D(\alpha+\beta)} = L_{D(\alpha)} \circ L_{D(\beta)}$. Aus Proposition 4.2.26(ii) und der Injektivität von L folgt:

$$D(\alpha + \beta) = D(\alpha) \cdot D(\beta).$$

Diese Gleichung liefert die gewöhnlichen Formeln für $\cos(\alpha + \beta)$ und $\sin(\alpha + \beta)$.

Bemerkung 4.2.34. Man kann den Satz 4.2.29 auf Vektorräume der Gestalt $K^{(I)}$ verallgemeinern, aber die Aussage wird komplizierter. Man nennt eine $I \times J$-indizierte Familie in K eine $I \times J$-*Matrix* über K. Aus dem Satz 4.1.22 erhalten wir einen Isomorphismus

$$L \colon M_{I\times J}(K) \xrightarrow{\sim} \mathrm{Hom}_K(K^{(J)}, K^I).$$

Ist $A \in M_{I\times J}(K)$, so landet die lineare Abbildung $L_A \colon K^{(J)} \to K^I$ in $K^{(I)}$ genau dann, wenn jede Spalte von A in $K^{(I)}$ liegt, d.h., wenn jede Spalte von A nur endlich viele Koeffizienten enthält, die nicht null sind. Eine solche Matrix heißt *spaltenendlich*. Dementsprechend schränkt sich der obige Isomorphismus zu einem Isomorphismus zwischen $\mathrm{Hom}_K(K^{(J)}, K^{(I)})$ und dem Vektorraum der spaltenendlichen $I \times J$-Matrizen ein.

Definition 4.2.35 (Rang einer Matrix)**.** Seien $m, n \in \mathbb{N}$ und sei $A \in M_{m\times n}(K)$. Der *Rang* von A ist der Rang von L_A (siehe Definition 4.1.41):

$$\mathrm{rg}\, A := \mathrm{rg}\, L_A = \dim_K(\mathrm{im}\, L_A).$$

Bemerkung 4.2.36. Da das Bild von L_A der Spaltenraum von A ist (Bemerkung 4.2.27), gilt

$$\mathrm{rg}\, A = \dim_K \mathrm{SR}(A).$$

Wir werden später beweisen, dass auch $\mathrm{rg}\, A = \dim_K \mathrm{ZR}(A)$ (siehe Korollar 4.2.49).

Proposition 4.2.37 (Charakterisierung der Invertierbarkeit)**.** *Sei $n \in \mathbb{N}$ und $A \in M_n(K)$. Die folgenden Aussagen sind äquivalent:*

(i) *A ist invertierbar.*

(ii) *A ist von links invertierbar, d.h., es existiert $B \in M_n(K)$ mit $B \cdot A = I_n$.*

(iii) *A ist von rechts invertierbar, d.h., es existiert $B \in M_n(K)$ mit $A \cdot B = I_n$.*

(iv) *Die lineare Abbildung $L_A \colon K^n \to K^n$ ist bijektiv.*

(v) *Die lineare Abbildung $L_A \colon K^n \to K^n$ ist injektiv.*

(vi) *Die lineare Abbildung $L_A \colon K^n \to K^n$ ist surjektiv.*

(vii) *Es gilt* $\operatorname{rg} A = n$.

Beweis. Die Implikationen (i) $\Rightarrow$ (ii) und (i) $\Rightarrow$ (iii) sind klar. Die Äquivalenz von (iv), (v) und (vi) folgt aus Korollar 4.1.39, und die Äquivalenz von (vi) und (vii) folgt aus Proposition 3.3.35: $L_A \colon K^n \to K^n$ ist genau dann surjektiv, wenn $\dim_K(\operatorname{im} L_A) = n$. Die Implikation (ii) $\Rightarrow$ (v) folgt aus Proposition 4.2.26(ii,iii), denn $B \cdot A = I_n$ impliziert $L_B \circ L_A = \mathrm{id}_{K^n}$, und damit ist L_A injektiv (nach Proposition 1.3.25(iii)). Die Implikation (iii) $\Rightarrow$ (vi) folgt auf ähnliche Weise.

Zum Schluss beweisen wir (iv) $\Rightarrow$ (i). Sei g die Umkehrabbildung von L_A und sei $B = M(g)$ die zugehörige Matrix. Nach Bemerkung 4.2.31 gilt

$$B \cdot A = M(g) \cdot M(L_A) = M(g \circ L_A) = M(\mathrm{id}_{K^n}) = I_n,$$

und ebenso $A \cdot B = I_n$. □

4.2.3 Darstellung von linearen Abbildungen

Nach Satz 4.2.29 können wir jede K-lineare Abbildung $f \colon K^n \to K^m$ als eine $m \times n$-Matrix A auffassen. Diese Darstellung von linearen Abbildungen als Matrizen ist sehr nützlich für Berechnungen. Zum Beispiel entspricht die Komposition von linearen Abbildungen der Matrixmultiplikation.

In diesem Abschnitt wollen wir diese Matrixdarstellung auf lineare Abbildungen $f \colon V \to W$ zwischen beliebigen endlich-dimensionalen K-Vektorräumen verallgemeinern. Wenn $B = (v_1, \dots, v_n)$ eine Basis von V ist, gibt es bekanntlich einen Isomorphismus

$$\varphi_B \colon K^n \xrightarrow{\sim} V$$

mit $\varphi_B(e_i) = v_i$ (siehe Bemerkung 3.3.18). Ist $v \in V$, so heißt der Spaltenvektor

$$[v]_B := \varphi_B^{-1}(v) \in K^n$$

der *Koordinatenvektor* von v bzgl. B (Definition 3.3.17). Sind V und W Vektorräume über K der Dimension n und m mit Basen B und C, so erhalten wir einen Isomorphismus

$$\begin{aligned} \mathrm{Hom}_K(V, W) &\xrightarrow{\sim} M_{m\times n}(K), \\ f &\mapsto [f]_C^B, \end{aligned}$$

als die Komposition der Isomorphismen

$$\mathrm{Hom}_K(V, W) \xrightarrow{\mathrm{Hom}_K(\varphi_B, \varphi_C^{-1})} \mathrm{Hom}_K(K^n, K^m) \xrightarrow{M} M_{m\times n}(K).$$

Seine Umkehrabbildung ist die Komposition

$$M_{m\times n}(K) \xrightarrow{L} \mathrm{Hom}_K(K^n, K^m) \xrightarrow{\mathrm{Hom}_K(\varphi_B^{-1}, \varphi_C)} \mathrm{Hom}_K(V, W).$$

Definition 4.2.38 (Darstellungsmatrix)**.** Seien V und W endlich-dimensionale K-Vektorräume mit Basen B und C und sei $f \colon V \to W$ eine lineare Abbildung. Die Matrix $[f]_C^B$ heißt die *Darstellungsmatrix* oder *Abbildungsmatrix* von f bzgl. der Basen B und C.

Seien $B = (v_1, \ldots, v_n)$ und $C = (w_1, \ldots, w_m)$. Nach Definition ist $[f]_C^B$ die $m \times n$-Matrix entsprechend der linearen Abbildung

$$\varphi_C^{-1} \circ f \circ \varphi_B \colon K^n \to K^m,$$

d.h., $[f]_C^B = M(\varphi_C^{-1} \circ f \circ \varphi_B)$. Die j-te Spalte von $[f]_C^B$ ist also

$$(\varphi_C^{-1} \circ f \circ \varphi_B)(e_j) = \varphi_C^{-1}(f(v_j)) = [f(v_j)]_C.$$

Anders gesagt, ist $[f]_C^B = (a_{ij})_{i,j}$, so gilt

$$f(v_j) = \sum_{i=1}^{m} a_{ij} \cdot w_i$$

für alle $j \in \{1, \ldots, n\}$, und die Matrix $[f]_C^B$ ist dadurch eindeutig bestimmt.

Proposition 4.2.39. *Seien V, W und U endlich-dimensionale K-Vektorräume mit Basen B, C und D.*

(i) *Für jede lineare Abbildung $f \colon V \to W$ und jeden Vektor $v \in V$ gilt*

$$[f(v)]_C = [f]_C^B \cdot [v]_B.$$

(ii) *Für lineare Abbildungen $f \colon V \to W$ und $g \colon W \to U$ gilt*

$$[g \circ f]_D^B = [g]_D^C \cdot [f]_C^B.$$

(iii) *Es gilt*

$$[\mathrm{id}_V]_B^B = I_n,$$

wobei $n = \dim_K(V)$.

(iv) *Für jeden Isomorphismus $f \colon V \xrightarrow{\sim} W$ ist die Matrix $[f]_C^B$ invertierbar, und es gilt*

$$[f^{-1}]_B^C = ([f]_C^B)^{-1}.$$

Beweis. Zu (i). Es gilt

$$[f]_C^B \cdot [v]_B = (\varphi_C^{-1} \circ f \circ \varphi_B)(\varphi_B^{-1}(v)) = \varphi_C^{-1}(f(v)) = [f(v)]_C.$$

Zu (ii). Dies folgt aus der Gleichung

$$(\varphi_D^{-1} \circ g \circ \varphi_C) \circ (\varphi_C^{-1} \circ f \circ \varphi_B) = \varphi_D^{-1} \circ (g \circ f) \circ \varphi_B,$$

indem wir M auf beide Seiten anwenden.

Zu (iii). $[\mathrm{id}_V]_B^B = M(\varphi_B^{-1} \circ \mathrm{id}_V \circ \varphi_B) = M(\mathrm{id}_{K^n}) = I_n$.

Zu (iv). Dies folgt aus (ii) and (iii). □

Bemerkung 4.2.40 (Rang einer Darstellungsmatrix)**.** Der Rang der Darstellungsmatrix $[f]_C^B$ ist gleich dem Rang von f. Denn es gilt

$$\mathrm{rg}\, [f]_C^B = \mathrm{rg}(\varphi_C^{-1} \circ f \circ \varphi_B) = \dim_K \mathrm{im}(\varphi_C^{-1} \circ f \circ \varphi_B)$$

und

$$\mathrm{im}(\varphi_C^{-1} \circ f \circ \varphi_B) = \mathrm{im}(\varphi_C^{-1} \circ f) = \varphi_C^{-1}(\mathrm{im}\, f) \cong \mathrm{im}\, f,$$

da φ_B surjektiv ist und φ_C^{-1} injektiv ist.

Definition 4.2.41 (Basiswechselmatrix)**.** Seien B und B' zwei Basen eines n-dimensionalen K-Vektorraums V. Die *Basiswechselmatrix* $T_{B'}^{B}$ von B nach B' ist die Matrix

$$T_{B'}^{B} := [\mathrm{id}_V]_{B'}^{B} = M(\varphi_{B'}^{-1} \circ \varphi_B) \in M_n(K).$$

Nach Proposition 4.2.39(i) gilt also

$$T_{B'}^{B} \cdot [v]_B = [v]_{B'}$$

für alle Vektoren $v \in V$. Nach Proposition 4.2.39(iv) ist die Basiswechselmatrix $T_{B'}^{B}$ invertierbar, mit

$$(T_{B'}^{B})^{-1} = T_{B}^{B'}.$$

Beispiel 4.2.42. Sei $B = (v_1, \ldots, v_n)$ eine Basis von K^n. Die Basiswechselmatrix von B nach der Standardbasis ist die Matrix $\begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}$.

Proposition 4.2.43 (Basiswechselformel)**.** *Seien V und W endlich-dimensionale Vektorräume über K. Seien B, B' Basen von V und C, C' Basen von W. Für jede lineare Abbildung $f\colon V \to W$ gilt*

$$[f]_{C'}^{B'} = T_{C'}^{C} \cdot [f]_{C}^{B} \cdot T_{B}^{B'}.$$

Insbesondere: Für jeden Endomorphismus $f\colon V \to V$ gilt

$$[f]_{B'}^{B'} = T_{B'}^{B} \cdot [f]_{B}^{B} \cdot T_{B}^{B'}.$$

Beweis. Dies folgt unmittelbar aus Proposition 4.2.39(ii):

$$T_{C'}^{C} \cdot [f]_{C}^{B} \cdot T_{B}^{B'} = [\mathrm{id}_W]_{C'}^{C} \cdot [f]_{C}^{B} \cdot [\mathrm{id}_V]_{B}^{B'} = [\mathrm{id}_W \circ f \circ \mathrm{id}_V]_{C'}^{B'} = [f]_{C'}^{B'}. \qquad \square$$

Beispiel 4.2.44. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die lineare Abbildung $f = L_A$ mit

$$A = \begin{pmatrix} 3 & -1 \\ 2 & 0 \end{pmatrix}.$$

Wir betrachten die Basis $B = (v_1, v_2)$ von $\mathbb{R}^2$ mit

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Die Basiswechselmatrix von B nach der Standardbasis E ist also

$$T_{E}^{B} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

und man kenn leicht durch Multiplizieren nachprüfen, dass

$$T_{B}^{E} = (T_{E}^{B})^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Nach Proposition 4.2.43 gilt

$$[f]_{B}^{B} = T_{B}^{E} \cdot A \cdot T_{E}^{B} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Diese Darstellungsmatrix zeigt, dass die lineare Abbildung f folgende einfache geometrische Beschreibung besitzt: Sie skaliert um den Faktor 2 in Richtung von v_1, und sie lässt die Vektoren in Richtung von v_2 fest. Die lineare Abbildung f ist ein Beispiel einer *diagonalisierbaren* Abbildung (siehe Abschnitt 6.2.1).

Dieses Beispiel zeigt folgendes: Selbst wenn wir nur an Vektorräumen der Gestalt K^n Interesse haben, ist es oft hilfreich, die Darstellungsmatrix einer linearen Abbildung bezüglich beliebiger Basen zu betrachten. Denn mit einer geeigneten Wahl von Basis kann die Darstellungsmatrix besonders einfach sein, was uns hilft, die Abbildung besser zu verstehen. Man kann insbesondere die folgenden Fragen stellen:

Frage 4.2.45. Seien V und W endlich-dimensionale Vektorräume über K.

(i) Für eine lineare Abbildung $f\colon V \to W$, wie kann man Basen B von V und C von W finden, so dass die Matrix $[f]_C^B$ so einfach wie möglich ist?

(ii) Für einen Endomorphismus $g\colon V \to V$, wie kann man eine Basis B von V finden, so dass die Matrix $[g]_B^B$ so einfach wie möglich ist?

Die möglichst einfache Darstellungsmatrix $[f]_C^B$ bzw. $[g]_B^B$ heißt die *Smith-Normalform* der linearen Abbildung f bzw. die *Jordansche Normalform* des Endomorphismus g. Die Existenz der Smith-Normalform können wir bereits beweisen:

Satz 4.2.46 (Smith-Normalform linearer Abbildungen)**.** *Seien V und W endlich-dimensionale K-Vektorräume der Dimension n und m, und sei $f\colon V \to W$ eine lineare Abbildung. Dann existieren Basen B von V und C von W, so dass*

$$[f]_C^B = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix},$$

wobei $r = \operatorname{rg} f$.

Beweis. Das ist der endlich-dimensionale Fall des Satzes 4.1.42, aber wir geben wieder einen vollständigen Beweis. Nach der Dimensionsformel für lineare Abbildungen hat der Kern von f die Dimension $n - r$. Nach dem Basisergänzungssatz existiert eine Basis $B = (v_1, \dots, v_n)$ von V, so dass $(v_{r+1}, \dots, v_n)$ eine Basis von $\ker f$ ist. Sei $U = \operatorname{Span}_K\{v_1, \dots, v_r\}$. Da U und $\ker f$ komplementär sind, ist die Einschränkung $f|_U$ injektiv, und damit ist die Familie $(f(v_1), \dots, f(v_r))$ in W linear unabhängig (Proposition 4.1.20(ii)). Nach dem Basisergänzungssatz gibt es weitere Vektoren $w_{r+1}, \dots, w_m$, so dass die Familie

$$C = (f(v_1), \dots, f(v_r), w_{r+1}, \dots, w_m)$$

eine Basis von W ist. Für $i \le r$ gilt dann $[f(v_i)]_C = e_i$ und für $i > r$ gilt $f(v_i) = 0$. Also hat die Matrix $[f]_C^B$ die gewünschte Form. □

Es stellt sich noch die Frage, wie man solche Basen B und C im konkreten Fall $V = K^n$ und $W = K^m$ finden kann. Eine effektive Methode zu solchen Berechnungen werden wir im Kapitel 5 erklären, und damit wird die Frage 4.2.45(i) vollständig beantwortet werden. Die Frage 4.2.45(ii) ist wesentlich schwieriger. Wir werden sie im Kapitel 6 nur teilweise beantworten; eine vollständige Antwort ist ein Ziel der Vorlesung *Lineare Algebra II*.

Zum Abschluss dieses Kapitels erklären wir den Zusammenhang zwischen der dualen Abbildung (Definition 4.1.47) und der transponierten Matrix (Definition 4.2.7):

Proposition 4.2.47 (Darstellungsmatrix der dualen Abbildung)**.** *Seien V und W endlich-dimensionale K-Vektorräume mit Basen B und C, und seien B^* und C^* die dualen Basen von V^* und W^*. Für jede lineare Abbildung $f\colon V \to W$ gilt*

$$[f^*]_{B^*}^{C^*} = ([f]_C^B)^\intercal.$$

Beweis. Seien

$$B = (v_1, \dots, v_n) \quad \text{und} \quad C = (w_1, \dots, w_m)$$

die gegebenen Basen und seien

$$B^* = (v_1^*, \dots, v_n^*) \quad \text{und} \quad C^* = (w_1^*, \dots, w_m^*),$$

die zugehörigen dualen Basen. Sei $[f]_C^B = (a_{ij})_{i,j}$. Nach Definition der Matrix $[f]_C^B$ ist ihre j-te Spalte der Koordinatenvektor $[f(v_j)]_C$, das heißt:

$$f(v_j) = \sum_{i=1}^{m} a_{ij} \cdot w_i.$$

Unser Ziel ist, die Koordinatenvektoren $[f^*(w_i^*)]_{B^*}$ zu berechnen. Es gilt

$$f^*(w_i^*)(v_j) = w_i^*(f(v_j)) = \sum_{k=1}^{m} a_{kj} \cdot w_i^*(w_k) = a_{ij},$$

da $w_i^*(w_k) = \delta_{ik}$. Daraus folgt, dass

$$f^*(w_i^*) = \sum_{j=1}^{n} a_{ij} \cdot v_j^*,$$

da beide linearen Abbildungen $V \to K$ dieselbe Werte auf der Basis B von V nehmen. Diese Gleichung bedeutet, dass der Koordinatenvektor $[f^*(w_i^*)]_{B^*}$ gleich der i-ten Zeile von $[f]_C^B$ ist, d.h., dass $[f^*]_{B^*}^{C^*}$ die transponierte Matrix zu $[f]_C^B$ ist. □

Bemerkung 4.2.48. Sei $A \in M_{m\times n}(K)$. Die Standardbasis von K^n induziert einen Isomorphismus $\varepsilon\colon K^n \xrightarrow{\sim} (K^n)^*$ mit $\varepsilon(e_i) = \pi_i$ (siehe Konstruktion 4.1.54). Proposition 4.2.47, angewendet auf die lineare Abbildung L_A und die Standardbasen, liefert ein kommutatives Quadrat

$$\begin{array}{ccc} K^m & \xrightarrow{L_{A^\intercal}} & K^n \\ {\scriptstyle\varepsilon}\downarrow\wr & & {\scriptstyle\varepsilon}\downarrow\wr \\ (K^m)^* & \xrightarrow{L_A^*} & (K^n)^*. \end{array}$$

Korollar 4.2.49 (Rang der transponierten Matrix)**.** *Seien $m, n \in \mathbb{N}$ und sei $A \in M_{m\times n}(K)$. Dann gilt*

$$\operatorname{rg} A^\intercal = \operatorname{rg} A,$$

das heißt, nach Bemerkung 4.2.36,

$$\dim_K \operatorname{ZR}(A) = \dim_K \operatorname{SR}(A).$$

Beweis. Wir wenden Proposition 4.2.47 mit der Abbildung $L_A\colon K^n \to K^m$ an. Die Matrix $A^\intercal$ ist also die Darstellungsmatrix der dualen Abbildung L_A^* bzgl. geeigneter Basen. Aus der Bemerkung 4.2.40 und dem Korollar 4.1.58 folgt

$$\operatorname{rg} A^\intercal = \operatorname{rg} L_A^* = \operatorname{rg} L_A = \operatorname{rg} A.$$ □

Kapitel 5

Lineare Gleichungen

Es gibt viele Arten von Gleichungen in der Mathematik: Polynomgleichungen, Diophantische Gleichungen, gewöhnliche und partielle Differentialgleichungen, usw. Ganz allgemein gesagt ist eine Gleichung ein Ausdruck der Gestalt $f(x) = g(x)$, wobei $f, g\colon X \to Y$ zwei Abbildungen mit derselben Definitions- und Zielmenge sind. Eine solche Gleichung zu *lösen* bedeutet, alle Elemente $x \in X$ zu finden, für die $f(x) = g(x)$ gilt. In den meisten Fällen ist g eine konstante Abbildung, d.h., die Gleichung hat die Form $f(x) = b$ mit einem $b \in Y$, und dann suchen wir alle Urbilder von b unter f (wenn Y eine Gruppe ist, ist das keine Beschränkung der Allgemeinheit, da jede Gleichung $f(x) = g(x)$ als $f(x)g(x)^{-1} = e$ umgeschrieben werden kann). Eine *lineare* Gleichung ist eine Gleichung $f(x) = b$, wobei f eine lineare Abbildung ist. Beispiele davon sind lineare Differentialgleichungen (siehe Beispiel 4.1.31).

In diesem Kapitel betrachten wir den konkreten Spezialfall, in dem f eine lineare Abbildung von K^n nach K^m ist, mit K einem beliebigen Körper. Das heißt, wir betrachten Gleichungen der Gestalt $A \cdot x = b$, wobei A eine $m \times n$-Matrix über K ist. Man kann solche Gleichungen als Systeme von m Gleichungen mit n Unbekannten $x_1, \dots, x_n \in K$ auffassen. Wir werden insbesondere einen Algorithmus lernen, das *Gaußsche Eliminationsverfahren*, mit dem man solche Gleichungssysteme systematisch und vollständig lösen kann. Dieser Algorithmus hat viele weitere Anwendungen in der linearen Algebra und auch außerhalb der Mathematik, da viele praktische Probleme durch lineare Gleichungssysteme modelliert werden können. Mit ihm kann man auch effektiv bestimmen, ob eine quadratische Matrix invertierbar ist, und falls ja, ihre inverse Matrix berechnen.

5.1 Lineare Gleichungssysteme

Definition 5.1.1 (lineares Gleichungssystem, Lösungsmenge)**.** Seien $m, n \in \mathbb{N}$. Ein *lineares Gleichungssystem* über K mit m Gleichungen und n Unbekannten (oder Variablen) ist ein Paar (A, b) bestehend aus einer Matrix $A \in M_{m\times n}(K)$ und einem Spaltenvektor $b \in K^m$. Das Gleichungssystem (A, b) heißt *homogen*, wenn $b = 0$.

Eine *Lösung* von (A, b) ist ein Spaltenvektor $x \in K^n$, so dass $A \cdot x = b$. Die *Lösungsmenge* von (A, b) ist die Menge aller Lösungen von (A, b); sie wird mit $\mathcal{L}(A, b)$ bezeichnet:

$$\mathcal{L}(A, b) = \{x \in K^n \mid A \cdot x = b\} \subset K^n.$$

Eine Lösung von (A, b) besteht konkreter aus n Skalaren $x_1, \dots, x_n \in K$, die gleichzeitig

die folgenden m Gleichungen erfüllen:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Bemerkung 5.1.2. Es gilt $\mathcal{L}(A,b) = L_A^{-1}(\{b\})$, wobei $L_A\colon K^n \to K^m$ die der Matrix A zugeordnete lineare Abbildung ist. Insbesondere ist die homogene Lösungsmenge

$$\mathcal{L}(A,0) = \ker L_A$$

ein Untervektorraum von K^n (nach Proposition 4.1.28), und es gilt

$$\mathcal{L}(A,b) \neq \varnothing \iff b \in \operatorname{im} L_A.$$

Proposition 5.1.3 (Struktur der Lösungsmenge eines linearen Gleichungssystems)**.** *Seien $m,n \in \mathbb{N}$, $A \in M_{m\times n}(K)$ und $b \in K^m$. Dann ist $\mathcal{L}(A,b)$ entweder leer oder eine Verschiebung des Untervektorraums $\mathcal{L}(A,0)$. Genauer: Für jede Lösung $x \in \mathcal{L}(A,b)$ gilt*

$$\mathcal{L}(A,b) = x + \mathcal{L}(A,0).$$

Außerdem gilt $\dim_K \mathcal{L}(A,0) = n - r$, wobei $r = \operatorname{rg} A$.

Beweis. Zu $\supset$. Sei $v \in \mathcal{L}(A,0)$. Dann gilt

$$A \cdot (x+v) = A \cdot x + A \cdot v = b + 0 = b,$$

und damit ist $x + v \in \mathcal{L}(A,b)$.

Zu $\subset$. Sei $v \in \mathcal{L}(A,b)$. Dann ist $v = x + (v - x)$, und es bleibt zu zeigen, dass $v - x \in \mathcal{L}(A,0)$:

$$A \cdot (x - v) = A \cdot x - A \cdot v = b - b = 0. \qquad \square$$

Nach Definition ist $\operatorname{rg} A = \operatorname{rg} L_A = \dim_K(\operatorname{im} L_A)$. Da $\mathcal{L}(A,0) = \ker L_A$ folgt die letzte Aussage aus der Dimensionsformel für die lineare Abbildung L_A (Korollar 4.1.38).

Um ein lineares Gleichungssystem (A,b) explizit zu lösen, soll man zuerst bestimmen, ob eine Lösung existiert. Wenn ja, genügt es nach Proposition 5.1.3 eine besondere Lösung $v_0 \in \mathcal{L}(A,b)$ und eine Basis $(v_1,\ldots,v_{n-r})$ von $\mathcal{L}(A,0)$ zu finden. Die allgemeine Lösung von $A \cdot x = b$ ist dann

$$x = v_0 + \sum_{i=1}^{n-r} \lambda_i v_i$$

mit beliebigen Skalaren $\lambda_i \in K$.

5.1.1 Zeilenstufenform

Es gibt besondere Matrizen A, die in sogenannter *Zeilenstufenform*, für die man die Lösungsmengen aller Gleichungssysteme (A,b) einfach bestimmen kann.

Definition 5.1.4 (Zeilenstufenform, Pivotelemente, Pivotspalten, reduzierte Zeilenstufenform)**.** Seien $m,n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K. Man sagt, dass A in *Zeilenstufenform* ist, wenn es Indizes $r \in \{0,\ldots,m\}$ und $1 \leq k_1 < \ldots < k_r \leq n$ gibt, so dass:

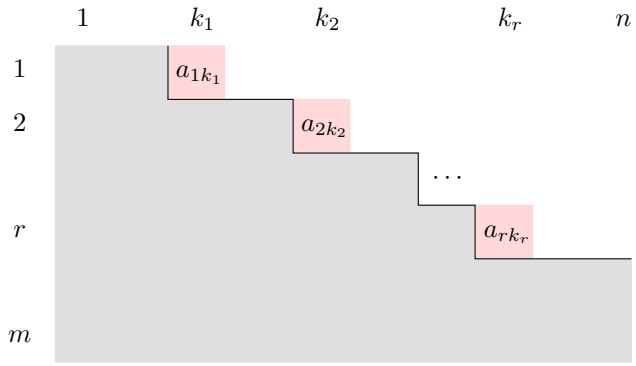
- $a_{ij} = 0$ für alle $i \leq r$ und alle $j < k_i$.

- $a_{ik_i} \neq 0$ für alle $i \leq r$.
- $a_{ij} = 0$ für alle $i > r$ und alle j.

Die r Koeffizienten $a_{1k_1}, \ldots, a_{rk_r}$ heißen die *Pivotelemente* von A, und ihre Spalten $A_{*k_1}, \ldots, A_{*k_r}$ sind die *Pivotspalten* von A.

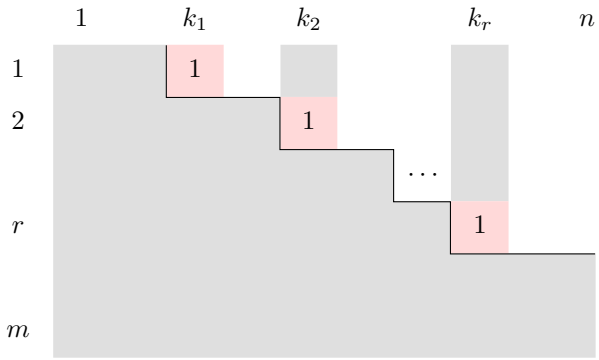
Man sagt, dass A in *reduzierter Zeilenstufenform* ist, wenn außerdem $A_{*k_i} = e_i$ für alle $i \leq r$. In einer Matrix in reduzierter Zeilenstufenform sind insbesondere alle Pivotelemente gleich 1.

Eine $m \times n$-Matrix A in Zeilenstufenform sieht wie folgt aus:



Das graue Gebiet besteht nur aus Nullen, die rote Einträge sind die Pivotelemente und das weisse Gebiet enthält beliebige Elemente aus K. Alle Spalten A_{*j} mit $j < k_1$, sowie alle Zeilen A_{i*} mit $i > r$, sind null (es kann aber sein, dass $k_1 = 1$ oder $r = m$).

Wenn A in *reduzierter* Zeilenstufenform ist, dann ist die i-te Pivotspalte A_{*k_i} der Standardeinheitsvektor e_i, d.h., alle Pivotelemente sind gleich 1 und alle darüberliegenden Koeffizienten sind null:



Proposition 5.1.5 (Zeilen- und Spaltenraum einer Matrix in Zeilenstufenform)**.** *Sei A eine $m \times n$-Matrix über K in Zeilenstufenform, und seien $k_1 < \cdots < k_r$ die Indizes der Pivotspalten mit $r \in \{0, \ldots, m\}$. Dann:*

(i) *Die ersten r Zeilen von A bilden eine Basis von* $\mathrm{ZR}(A)$.

(ii) *Es gilt* $\mathrm{SR}(A) = K^r \times \{0\} \subset K^m$ *und die Pivotspalten von A bilden eine Basis von* $\mathrm{SR}(A)$. *Insbesondere gilt* $\operatorname{rg} A = r$.

Beweis. Zu (i). Der Zeilenraum ist von den ersten r Zeilen erzeugt, weil alle anderen Zeilen null sind. Wegen der Zeilenstufenform ist es klar, dass die ersten r Zeilen linear unabhängig sind. Sie bilden also eine Basis von $\mathrm{ZR}(A)$.

Zu (ii). Da jede Spalte von A im Untervektorraum $K^r \times \{0\}$ liegt, gilt $\mathrm{SR}(A) \subset K^r \times \{0\}$. Wegen der Zeilenstufenform ist es aber klar, dass die r Pivotspalten linear unabhängig sind, so dass $\dim \mathrm{SR}(A) \geq r$. Aus Proposition 3.3.35 folgt, dass $\mathrm{SR}(A) = K^r \times \{0\}$. □

Rezept 5.1.6 (Lösung eines linearen Gleichungssystems in Zeilenstufenform)**.** Sei (A, b) ein lineares Gleichungssystem über K mit m Gleichungen und n Unbekannten, wobei A in Zeilenstufenform ist. Seien $k_1 < \cdots < k_r$ die Indizes der Pivotspalten von A. Dann kann die Lösungsmenge $\mathcal{L}(A, b)$ wie folgt bestimmt werden:

- Falls es einen Index $i > r$ mit $b_i \neq 0$ gibt, dann ist $\mathcal{L}(A, b) = \varnothing$. Denn die i-te Gleichung des Systems ist $0 \cdot x = b_i$ und ist nicht mit $x \in K^n$ lösbar.
- Andernfalls ist $\mathcal{L}(A, b)$ nicht leer, und man kann die allgemeine Lösung $x \in K^n$ wie folgt bestimmen:
  - Für die Unbekannten x_{k_i} mit $i \in \{1, \ldots, r\}$ gilt
$$x_{k_i} = \frac{1}{a_{ik_i}} \left(b_{k_i} - \sum_{j=k_i+1}^{n} a_{ij} x_j \right).$$
  - Es gibt keine weitere Bedingungen, d.h., die Unbekannten x_j mit $j \notin \{k_1, \ldots, k_r\}$ sind beliebig. Diese Unbekannten heißen die *freien Variablen*.
  - Durch *Rückwärtssubstitution* kann man die x_{k_i} nur durch die freien Variablen ausdrücken (dieser Schritt ist nicht nötig, wenn die Matrix A in *reduzierter* Zeilenstufenform ist).

  Dann erhalten wir die allgemeine Lösung in der Form
$$x = v_0 + \sum_{j \notin \{k_1, \ldots, k_r\}} x_j v_j, \quad x_j \text{ beliebig},$$
  mit geeigneten Vektoren $v_0, v_j \in K^n$. Dabei ist also v_0 eine besondere Lösung von (A, b) und ist $(v_j)_{j \notin \{k_1, \ldots, k_r\}}$ eine Basis von $\mathcal{L}(A, 0)$.

Beispiel 5.1.7.

(i) Seien
$$A = \begin{pmatrix} 2 & -1 & 4 & 0 \\ 0 & 3 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}.$$
Wenn $\mathrm{char}(K) \notin \{2, 3\}$ ist die Matrix A in Zeilenstufenform mit drei Pivotspalten. Für das System $A \cdot x = b$ erhalten wir
$$\begin{aligned} x_1 &= \frac{1}{2}(x_2 - 4x_3) \\ x_2 &= \frac{1}{3}(3 - x_3 + 3x_4) \\ x_4 &= 1, \end{aligned}$$
und die Variable x_3 ist frei. Durch Rückwärtssubstitution erhalten wir
$$\begin{aligned} x_4 &= 1, \\ x_2 &= \frac{1}{3}(3 - x_3 + 3) = 2 - \frac{1}{3}x_3, \\ x_1 &= \frac{1}{2}\left(2 - \frac{1}{3}x_3 - 4x_3\right) = 1 - \frac{13}{6}x_3, \end{aligned}$$
was die allgemeine Lösung des Gleichungssystems ergibt:
$$x = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} -\frac{13}{6} \\ -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix}, \quad x_3 \text{ beliebig}.$$

(ii) Seien

$$A = \begin{pmatrix} 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad c = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Die Matrix A ist in reduzierter Zeilenstufenform mit zwei Pivotspalten. Das System $A \cdot x = b$ hat keine Lösung, weil $b_3 \neq 0$. Für das System $A \cdot x = c$ erhalten wir

$$\begin{aligned} x_2 &= 2 - 3x_3, \\ x_4 &= 1, \end{aligned}$$

und die Variablen x_1 und x_3 sind frei. Da A in reduzierter Zeilenstufenform war, braucht man hier keine Rüchwärtssubstitution durchzuführen. Die allgemeine Lösung ist also

$$x = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} + x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \quad x_1, x_3 \text{ beliebig.}$$

5.2 Das Gaußsche Eliminationsverfahren

Das Gaußsche Eliminationsverfahren ist ein Algorithmus, der beliebige lineare Gleichungssysteme löst. Dieser Algorithmus ist sowohl von theoretischer als auch von praktischer Bedeutung: Er ist sehr effizient und wird tatsächlich von Computern zur Lösung großer linearer Gleichungssysteme verwendet. Andererseits werden wir diesen Algorithmus verwenden, um weitere Sätze zu beweisen.

Die Strategie zur Lösung eines beliebigen linearen Gleichungsystems ist die folgende:

- Wie bereits im Abschnitt 5.1.1 erklärt, gibt es besondere Matrizen A, die in sogenannter *Zeilenstufenform*, für die man die Lösungsmengen aller Gleichungssysteme (A, b) einfach bestimmen kann.
- Zu jedem linearen Gleichungssystem (A, b) kann man ein lineares Gleichungssystem (A', b') mit derselben Lösungsmenge effektiv finden, durch sogenannte *elementare Zeilenumformungen*, wobei A' in Zeilenstufenform ist.

Definition 5.2.1 (Zeilenumformung)**.** Seien $m, n \in \mathbb{N}$. Eine *Zeilenumformung* oder *Zeilenoperation* auf $m \times n$-Matrizen ist eine Abbildung der Gestalt

$$\begin{aligned} M_{m\times n}(K) &\to M_{m\times n}(K), \\ A &\mapsto Z \cdot A, \end{aligned}$$

wobei Z eine invertierbare $m \times m$-Matrix ist.

Nach Definition der Matrixmultiplikation ergibt sich die umgeformte Matrix $Z \cdot A$ aus A, indem man jede Zeile durch eine Linearkombination der Zeilen ersetzt: Für alle $k \in \{1, \ldots, m\}$ gilt

$$(Z \cdot A)_{k*} = \sum_{i=1}^{m} Z_{ki} \cdot A_{i*}.$$

Da Z invertierbar ist, ist außerdem die entsprechende Zeilenumformung $A \mapsto Z \cdot A$ bijektiv, mit Umkehrabbildung $A \mapsto Z^{-1} \cdot A$.

Proposition 5.2.2 (Zeilen- und Spaltenraum bei Zeilenumformungen)**.** *Sei A eine $m \times n$-Matrix über K und sei $Z \in \mathrm{GL}_m(K)$.*

(i) *Es gilt* $\mathrm{ZR}(Z \cdot A) = \mathrm{ZR}(A)$.

(ii) *Es gilt* $\mathrm{SR}(Z \cdot A) = L_Z(\mathrm{SR}(A))$ *und insbesondere* $\mathrm{SR}(Z \cdot A) \cong \mathrm{SR}(A)$.

Beweis. Zu (i). Jede Zeile von $Z \cdot A$ ist eine Linearkombination der Zeilen von A und damit liegt im Zeilenraum von A. Dies zeigt, dass $\mathrm{ZR}(Z \cdot A) \subset \mathrm{ZR}(A)$. Umgekehrt gilt $\mathrm{ZR}(A) = \mathrm{ZR}(Z^{-1} \cdot Z \cdot A) \subset \mathrm{ZR}(Z \cdot A)$.

Zu (ii). Mit der Bemerkung 4.2.27 und der Proposition 4.2.26(ii) berechnen wir

$$\mathrm{SR}(Z \cdot A) = \operatorname{im} L_{Z \cdot A} = \operatorname{im}(L_Z \circ L_A) = L_Z(\operatorname{im} L_A) = L_Z(\mathrm{SR}(A)).$$

Da Z invertierbar ist, ist L_Z nach Proposition 4.2.37 ein Isomorphismus und damit gilt $L_Z(\mathrm{SR}(A)) \cong \mathrm{SR}(A)$. □

Proposition 5.2.3 (Invarianz der Lösungsmenge bei Zeilenumformungen)**.** *Sei* (A, b) *ein lineares Gleichungssystem über* K *mit* m *Gleichungen und* n *Unbekannten, und sei* $Z \in \mathrm{GL}_m(K)$*. Dann gilt*

$$\mathcal{L}(A, b) = \mathcal{L}(Z \cdot A, Z \cdot b).$$

Beweis. Sei $x \in \mathcal{L}(A, b)$, d.h., $A \cdot x = b$. Daraus folgt, dass $Z \cdot A \cdot x = Z \cdot b$, d.h., $x \in \mathcal{L}(Z \cdot A, Z \cdot b)$. Sei umgekehrt $x \in \mathcal{L}(Z \cdot A, Z \cdot b)$, d.h., $Z \cdot A \cdot x = Z \cdot b$. Dann gilt

$$A \cdot x = Z^{-1} \cdot Z \cdot A \cdot x = Z^{-1} \cdot Z \cdot b = b,$$

d.h., $x \in \mathcal{L}(A, b)$. □

Zur Erinnerung ist E_{rs} die Matrix mit $(E_{rs})_{ij} = \delta_{ir}\delta_{js}$.

Definition 5.2.4 (Elementarmatrizen, elementare Zeilenumformung)**.** Sei $m \in \mathbb{N}$. Die folgenden $m \times m$-Matrizen über K heißen *Elementarmatrizen*:

- Seien $i, j \in \{1, \ldots, m\}$ mit $i < j$. Die Matrix $V_{ij} \in M_m(K)$ ist
$$V_{ij} := I_m - E_{ii} - E_{jj} + E_{ij} + E_{ji}.$$
- Seien $i \in \{1, \ldots, m\}$ und $\lambda \in K^\times$. Die Matrix $M_i(\lambda) \in M_m(K)$ ist
$$M_i(\lambda) := I_m - E_{ii} + \lambda \cdot E_{ii}.$$
- Seien $i, j \in \{1, \ldots, m\}$ mit $i \neq j$ und sei $\alpha \in K$. Die Matrix $A_{ij}(\alpha) \in M_m(K)$ ist
$$A_{ij}(\alpha) := I_m + \alpha \cdot E_{ij}.$$

Eine Zeilenumformung $A \mapsto Z \cdot A$ mit Z einer Elementarmatrix heißt *elementare Zeilenumformung.*

Bei dieser Definition soll man nachprüfen, dass Elementarmatrizen invertierbar sind. Ist Z eine Elementarmatrix, so gilt folgendes für die Zeilen von $Z \cdot A$:

$$(V_{ij} \cdot A)_{k*} = \begin{cases} A_{j*}, & \text{falls } k = i, \\ A_{i*}, & \text{falls } k = j, \\ A_{k*}, & \text{andernfalls,} \end{cases}$$

$$(M_i(\lambda) \cdot A)_{k*} = \begin{cases} \lambda \cdot A_{i*}, & \text{falls } k = i, \\ A_{k*}, & \text{andernfalls,} \end{cases}$$

$$(A_{ij}(\alpha) \cdot A)_{k*} = \begin{cases} A_{i*} + \alpha \cdot A_{j*}, & \text{falls } k = i, \\ A_{k*}, & \text{andernfalls.} \end{cases}$$

Es folgt daraus (und aus der Proposition 4.2.37 (i) $\Leftrightarrow$ (iii)), dass Elementarmatrizen invertierbar sind, mit Inversen

$$V_{ij}^{-1} = V_{ij}, \quad M_i(\lambda)^{-1} = M_i(\lambda^{-1}), \quad A_{ij}(\alpha)^{-1} = A_{ij}(-\alpha).$$

Diese Beobachtungen sind in folgender Tabelle zusammengefasst:

Matrix Z	Zeilenumformung $A \mapsto Z \cdot A$	inverse Matrix Z^{-1}
V_{ij}	Vertauschung der i-ten und j-ten Zeilen	V_{ij}
$M_i(\lambda)$	Multiplikation der i-ten Zeile mit λ	$M_i(\lambda^{-1})$
$A_{ij}(\alpha)$	Addition des α-fachen der j-ten Zeile zur i-ten Zeile	$A_{ij}(-\alpha)$

Bemerkung 5.2.5. Multiplikation mit einer invertierbaren Matrix von *rechts* ergibt eine *Spaltenumformung*, indem jede Spalte durch eine Linearkombination der Spalten ersetzt wird. Elementarmatrizen liefern dabei *elementare Spaltenumformungen*. Da $A \cdot Z = (Z^\intercal \cdot A^\intercal)^\intercal$ und die transponierte Matrix zu einer Elementarmatrix wieder eine Elementarmatrix ist, sind elementare Spaltenumformungen die offensichtlichen Entsprechungen von elementaren Zeilenumformungen.

Satz 5.2.6 (Gaußsches Eliminationsverfahren)**.** *Seien $m, n \in \mathbb{N}$. Jede $m \times n$-Matrix A über K kann durch elementare Zeilenumformungen auf reduzierte Zeilenstufenform gebracht werden. Das heißt, es existiert Elementarmatrizen $Z_1, \ldots, Z_k$, so dass die Matrix $Z_k \cdot \ldots \cdot Z_1 \cdot A$ in reduzierter Zeilenstufenform ist.*

Das Gaußsche Eliminationsverfahren ist sogar ein Algorithmus, das jede Matrix auf reduzierte Zeilenstufenform bringt. Um diesen Algorithmus zu verstehen, ist es hilfreich, ein Bild einer Matrix in Zeilenstufenform parat zu haben. Wir erklären zunächst, wie man eine Matrix durch elementare Zeilenumformungen auf Zeilenstufenform bringen kann:

Algorithmus 5.2.7 (Gaußsches Eliminationsverfahren I: Eine Matrix auf Zeilenstufenform bringen)**.** Sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K. Man zeigt durch Induktion über $j \in \{0, \ldots, n\}$, wie man die ersten j Spalten von A auf Zeilenstufenform bringen kann. Wenn $j = 0$ gibt es nichts zu tun. Sei also $j \in \{1, \ldots, n\}$. Angenommen sind die ersten $j-1$ Spalten von A schon in Zeilenstufenform. Sei $s - 1$ die Anzahl der bisherigen Pivotspalten ($s \geq 1$). Falls $a_{ij} = 0$ für alle $i \geq s$, dann sind die ersten j Spalten von A schon in Zeilenstufenform. Andernfalls entspricht j einer neuen Pivotspalte.

(i) Man wähle $i \geq s$ mit $a_{ij} \neq 0$ und man vertausche die i-te une s-te Zeilen, so dass $a_{sj} \neq 0$. Das Element a_{sj} ist das s-te Pivotelement.

(ii) Für alle $i > s$ verwende man die elementare Zeilenumformung $A_{is}(-a_{ij}/a_{sj})$, um a_{ij} zu Null zu machen. Die ersten j Spalten von A sind jetzt in Zeilenstufenform.

Wenn eine Matrix in Zeilenstufenform ist, kann man sie weiter folgendermaßen auf reduzierte Zeilenstufenform bringen:

Algorithmus 5.2.8 (Gaußsches Eliminationsverfahren II: Eine Matrix in Zeilenstufenform auf reduzierte Zeilenstufenform bringen)**.** Sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K in Zeilenstufenform. Seien $1 \leq k_1 < \cdots < k_r \leq n$ die Indizes der Pivotspalten von A. Für $i = r, r-1, \ldots, 1$ tue man nacheinander folgendes:

(i) Man multipliziere die i-te Zeile mit $1/a_{ik_i}$, um das Pivotelement a_{ik_i} zu 1 zu machen.

(ii) Für alle $e < i$ verwende man die elementare Zeilenumformung $A_{ei}(-a_{ek_i})$, um a_{ek_i} zu Null zu machen.

Rezept 5.2.9 (Lösung eines beliebigen linearen Gleichungssystems)**.** Sei (A, b) ein lineares Gleichungssystem über K. Um seine Lösungsmenge zu bestimmen, betrachten wir die *erweiterte Matrix* $(A \,|\, b)$, indem wir den Spaltenvektor b am Ende der Matrix A hinzufügen. Mit dem Gaußschen Eliminationsverfahren können wir die Matrix $(A \,|\, b)$ durch elementare Zeilenumformungen in eine Matrix $(A' \,|\, b')$ bringen, wobei A' in (reduzierter) Zeilenstufenform ist. Das heißt, es gilt $A' = Z_k \cdot \ldots \cdot Z_1 \cdot A$ und $b' = Z_k \cdot \ldots \cdot Z_1 \cdot b$ mit geeigneten Elementarmatrizen $Z_1, \ldots, Z_k$. Nach Proposition 5.2.3 gilt dann $\mathcal{L}(A, b) = \mathcal{L}(A', b')$, und mit dem Rezept 5.1.6 können wir diese Lösungsmenge völlig bestimmen.

Beispiel 5.2.10. Seien

$$A = \begin{pmatrix} 1 & 6 & 2 & -5 & -2 \\ -2 & -12 & -2 & 2 & 3 \\ 3 & 18 & 4 & -7 & -3 \end{pmatrix}, \quad b = \begin{pmatrix} -4 \\ 11 \\ -9 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

Wir lösen die Gleichungssysteme $A \cdot x = b$ und $A \cdot x = c$ gleichzeitig durch das Gaußsche Eliminationsverfahren:

$$\left(\begin{array}{ccccc|c|c} 1 & 6 & 2 & -5 & -2 & -4 & 1 \\ -2 & -12 & -2 & 2 & 3 & 11 & 2 \\ 3 & 18 & 4 & -7 & -3 & -9 & 0 \end{array}\right) \xrightarrow[A_{31}(-3)]{A_{21}(2)} \left(\begin{array}{ccccc|c|c} 1 & 6 & 2 & -5 & -2 & -4 & 1 \\ 0 & 0 & 2 & -8 & -1 & 3 & 4 \\ 0 & 0 & -2 & 8 & 3 & 3 & -3 \end{array}\right)$$

$$\xrightarrow{A_{32}(1)} \left(\begin{array}{ccccc|c|c} 1 & 6 & 2 & -5 & -2 & -4 & 1 \\ 0 & 0 & 2 & -8 & -1 & 3 & 4 \\ 0 & 0 & 0 & 0 & 2 & 6 & 1 \end{array}\right).$$

Wenn $\mathrm{char}(K) \neq 2$ ist das System jetzt in Zeilenstufenform mit drei Pivotspalten, und die freien Variablen sind x_2 und x_4. Für das System $A \cdot x = b$ erhalten wir durch Rückwärtssubstitution

$$\begin{aligned} x_5 &= 3, \\ x_3 &= \frac{1}{2}(3 + 8x_4 + x_5) = 3 + 4x_4, \\ x_1 &= -4 - 6x_2 - 2x_3 + 5x_4 + 2x_5 = -4 - 6x_2 - 3x_4. \end{aligned}$$

Die allgemeine Lösung von (A, b) ist also

$$x = \begin{pmatrix} -4 \\ 0 \\ 3 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} -6 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{pmatrix},$$

mit $x_2, x_4 \in K$ beliebig. Die zwei letzten Vektoren bilden also eine Basis von $\mathcal{L}(A, 0)$. Für das System $A \cdot x = c$ erhalten wir durch Rückwärtssubstitution

$$\begin{aligned} x_5 &= \frac{1}{2}, \\ x_3 &= \frac{1}{2}(4 + 8x_4 + x_5) = \frac{9}{4} + 4x_4, \\ x_1 &= 1 - 6x_2 - 2x_3 + 5x_4 + 2x_5 = -\frac{5}{2} - 6x_2 - 3x_4. \end{aligned}$$

Die allgemeine Lösung von (A, c) ist also

$$x = \frac{1}{4} \begin{pmatrix} -10 \\ 0 \\ 9 \\ 0 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} -6 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{pmatrix},$$

mit $x_2, x_4 \in K$ beliebig.

Wenn $\mathrm{char}(K) = 2$ haben wir das System

$$\left(\begin{array}{ccccc|c|c} 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right),$$

das bereits in reduzierter Zeilenstufenform ist. Jetzt gibt es zwei Pivotspalten und die freien Variablen sind x_2, x_3, x_4. Für das System (A, b) gilt $x_5 = 1$ und $x_1 = -x_4 = x_4$. Aber für das System (A, c) gilt $\mathcal{L}(A, c) = \varnothing$ wegen der dritten Gleichung $0 = 1$.

Das Gaußsche Eliminationsverfahren liefert einen alternativen, konkreteren Beweis der Tatsache, dass der Zeilenraum und der Spaltenraum einer beliebigen $m \times n$-Matrix immer die gleiche Dimension haben (Korollar 4.2.49):

Alternativer Beweis des Korollars 4.2.49. Sei A' eine Zeilenstufenform von A, d.h., A' ist in Zeilenstufenform und wird aus A durch elementare Zeilenumformungen erhalten. Nach Proposition 5.2.2 gelten

$$\dim_K \mathrm{ZR}(A) = \dim_K \mathrm{ZR}(A'),$$
$$\dim_K \mathrm{SR}(A) = \dim_K \mathrm{SR}(A').$$

Nach Proposition 5.1.5 gilt

$$\dim_K \mathrm{ZR}(A') = \dim_K \mathrm{SR}(A').$$

Aus diesen drei Gleichungen folgt, dass $\dim_K \mathrm{ZR}(A) = \dim_K \mathrm{SR}(A)$. □

Korollar 5.2.11. *Seien $n \in \mathbb{N}$ und $A \in M_n(K)$. Die folgenden Aussagen sind äquivalent:*

(i) *A ist invertierbar.*

(ii) *A ist ein Produkt von Elementarmatrizen.*

Insbesondere ist jede Zeilenumformung eine Komposition elementarer Zeilenumformungen.

Beweis. Die Implikation (ii) ⇒ (i) folgt daraus, dass Elementarmatrizen invertierbar sind. Sei umgekehrt A invertierbar, d.h., $\mathrm{rg}\, A = n$ (Proposition 4.2.37). Nach Satz 5.2.6 existieren Elementarmatrizen $Z_1, \ldots, Z_k$, so dass $A' = Z_k \cdot \ldots \cdot Z_1 \cdot A$ in reduzierter Zeilenstudenform ist. Nach Proposition 5.2.2(ii) ist $\mathrm{rg}\, A' = \mathrm{rg}\, A = n$, und damit hat A' n Pivotspalten (Proposition 5.1.5). Aber die einzige $n \times n$-Matrix in reduzierter Zeilenstufenform mit n Pivotspalten ist die Einheitsmatrix I_n. Daraus folgt, dass $A = Z_1^{-1} \cdot \ldots \cdot Z_k^{-1}$. □

Ein theoretischer Vorteil der *reduzierten* Zeilenstufenform gegenüber der Zeilenstufenform ist die folgende Eindeutigkeitsaussage:

Proposition 5.2.12 (Eindeutigkeit der reduzierten Zeilenstufenform)**.** *Seien $m, n \in \mathbb{N}$ und $A \in M_{m\times n}(K)$. Seien $Z, Z' \in \mathrm{GL}_m(K)$, so dass die Matrizen $B = Z \cdot A$ und $B' = Z' \cdot A$ in reduzierter Zeilenstufenform sind. Dann ist $B = B'$.*

**Beweis.* Nach Proposition 5.2.2(ii) haben B und B' denselben Rang. Es genügt also die folgende Aussage zu beweisen: Sind B und B' $m\times n$-Matrizen in reduzierter Zeilenstufenform mit demselben Rang r, so dass jede Zeile von B' eine Linearkombination der Zeilen von B ist, so gilt $B = B'$. Seien $k_1 < \cdots < k_r$ bzw. $k'_1 < \cdots < k'_r$ die Indizes der Pivotspalten von B bzw. B'. Wir beweisen die Aussage durch Induktion über r. Wenn $r = 0$ sind beide B und B' null. Sei $r \geq 1$ und seien C und C' die $(m-1) \times n$-Matrizen, die aus B und B' entstehen, wenn man die ersten Zeilen streicht. Die Matrizen C und C' sind wieder in reduzierter Zeilenstufenform und haben Rang $r-1$. Da die Zeile B'_{1*} eine Linearkombination der Zeilen B_{i*} ist und die Spalten B_{*j} für alle $j < k_1$ null sind, muss $k'_1 \geq k_1$ sein. Für jedes $i \in \{2, \ldots, m\}$ gilt insbesondere $B'_{ik_1} = 0$. In einer Linearkombination der Zeilen B_{1*}, …, B_{m*}, die B'_{i*} ergibt, muss deswegen B_{1*} mit dem Koeffizient 0 vorkommen. Das heißt, jede Zeile von C' ist eine Linearkombination der Zeilen von C. Nach der Induktionsvoraussetzung ist $C = C'$; insbesondere ist $k_e = k'_e$ für alle $e \in \{2, \ldots, r\}$. Es bleibt zu zeigen, dass $B_{1*} = B'_{1*}$. Sei $B'_{1*} = \sum_{e=1}^{r} \lambda_e B_{e*}$. Für alle $e \geq 2$ gilt $B'_{1k_e} = 0$ aber $B_{ek_e} = 1$, und deswegen muss $\lambda_e = 0$ sein. Das heißt, es gilt $B'_{1*} = \lambda_1 B_{1*}$, und die einzige Möglichkeit ist dann $\lambda_1 = 1$. □

Bemerkung 5.2.13. Seien $A, B \in M_{m\times n}(K)$. Man sagt, dass A *zeilenäquivalent* zu B ist, wenn eine Matrix $Z \in \mathrm{GL}_m(K)$ mit $Z \cdot A = B$ existiert. Es ist klar, dass Zeilenäquivalenz eine Äquivalenzrelation auf $M_{m\times n}(K)$ ist. Das Gaußsche Eliminationsverfahren und die Proposition 5.2.12 implizieren, dass jede Äquivalenzklasse *genau eine* Matrix in reduzierter Zeilenstufenform enthält. Insbesondere können wir durch das Gaußsche Eliminationsverfahren effektiv bestimmen, ob zwei Matrizen zeilenäquivalent sind, indem wir beide Matrizen in reduzierte Zeilenstufenform bringen.

5.2.1 Rezepte

Mit dem Gaußschen Eliminationsverfahren können wir jetzt beliebige lineare Gleichungssysteme effektiv lösen (siehe Rezept 5.2.9). Aber dieser Algorithmus hilft auch bei vielen anderen Problemen in der linearen Algebra. In diesem Abschnitt erklären wir einige solcher weiteren Anwendungen.

Alle nachfolgenden Rezepte werden mit Vektorräumen der Gestalt K^n und Matrizen formuliert. Aber diese Rezepte können auch auf beliebige endlich-dimensionale Vektorräume und lineare Abbildungen zwischen denen angewendet werden, sofern Basen dieser Vektorräume bekannt sind: Dann kann das Problem auf Vektorräume der Gestalt K^n übertragen werden.

Rezept 5.2.14 (Berechnung des Ranges)**.** Gegeben sei eine Matrix $A \in M_{m\times n}(K)$. Gesucht ist der Rang von A. Man überführt A mit dem Gaußschen Eliminationsverfahren in Zeilenstufenform. Nach Propositionen 5.2.2(ii) und 5.1.5(ii) ist die Anzahl der Pivotspalten gleich dem Rang von A

Rezept 5.2.15 (Berechnung des Kerns)**.** Gegeben sei eine Matrix $A \in M_{m\times n}(K)$. Gesucht ist eine Basis von $\ker L_A$. Da $\ker L_A = \mathcal{L}(A, 0)$, ist das ein Sonderfall vom Rezept 5.2.9.

Rezept 5.2.16 (Berechnung des Bildes)**.** Gegeben sei eine Matrix $A \in M_{m\times n}(K)$. Gesucht ist eine Basis von $\operatorname{im} L_A$. Man bringt A auf Zeilenstufenform A' mit dem Gaußschen Eliminationsverfahren. Sind $k_1 < \cdots < k_r$ die Indizes der Pivotspalten von A', so ist $(A_{*k_1}, \ldots, A_{*k_r})$ eine Basis von $\operatorname{im} L_A$. Denn $A' = Z \cdot A$ mit einem $Z \in \mathrm{GL}_n(K)$, und die Pivotspalten $Z \cdot A_{*k_i}$ bilden eine Basis von $\mathrm{SR}(Z \cdot A)$ (Proposition 5.1.5(ii)). Nach Proposition 5.2.2(ii) bilden dann die Spalten A_{*k_i} eine Basis von $\mathrm{SR}(A) = \operatorname{im} L_A$.

Alternativ dazu führt man das Gaußsche Eliminationsverfahren mit der transponierten Matrix $A^\intercal$ durch, bis sie in Zeilenstufenform ist. Nach Propositionen 5.2.2(i) und 5.1.5(i) bilden jetzt die Nicht-Null-Zeilen eine Basis von $\mathrm{ZR}(A^\intercal) = \mathrm{SR}(A) = \operatorname{im} L_A$.

Rezept 5.2.17 (Test auf Invertierbarkeit)**.** Gegeben sei eine Matrix $A \in M_n(K)$. Zu bestimmen ist, ob A invertierbar ist. Man berechnet den Rang von A mit dem Rezept 5.2.14. Nach Proposition 4.2.37 ist A genau dann invertierbar, wenn $\operatorname{rg} A = n$.

Rezept 5.2.18 (Berechnung der inversen Matrix)**.** Gegeben sei eine Matrix $A \in M_n(K)$. Gesucht ist die inverse Matrix A^{-1}, wenn sie existiert. Man führt das Gaußsche Eliminationsverfahren mit der erweiterten Matrix $(A \mid I_n)$ durch, bis A in reduzierter Zeilenstufenform ist (sofern die Zeilenstufenform n Pivotspalten besitzt, sonst ist A nicht invertierbar). Die erweiterte Matrix hat jetzt die Form $(I_n \mid B)$, und es gilt $A^{-1} = B$. Denn es gibt ein $Z \in \mathrm{GL}_n(K)$ mit $Z \cdot A = I_n$ und $Z \cdot I_n = B$, und damit ist $B = Z = A^{-1}$.

Beispiel 5.2.19. Sei

$$A = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 1 & 7 \\ 1 & 6 & 0 \end{pmatrix} \in M_3(K).$$

Wir führen das Gaußsche Eliminationsverfahren durch:

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 6 & 1 & 0 & 0 \\ 0 & 1 & 7 & 0 & 1 & 0 \\ 1 & 6 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{V_{13}} \left(\begin{array}{ccc|ccc} 1 & 6 & 0 & 0 & 0 & 1 \\ 0 & 1 & 7 & 0 & 1 & 0 \\ 0 & 1 & 6 & 1 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{A_{32}(-1)} \left(\begin{array}{ccc|ccc} 1 & 6 & 0 & 0 & 0 & 1 \\ 0 & 1 & 7 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & -1 & 0 \end{array}\right).$$

Die linke Seite ist jetzt in Zeilenstufenform und hat Rang 3, so dass A invertierbar ist. Wir führen das Eliminationsverfahren weiter, bis die linke Seite die Einheitsmatrix wird:

$$\xrightarrow{M_3(-1)} \left(\begin{array}{ccc|ccc} 1 & 6 & 0 & 0 & 0 & 1 \\ 0 & 1 & 7 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right) \xrightarrow[A_{12}(-6)]{A_{23}(-7)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -42 & 36 & 1 \\ 0 & 1 & 0 & 7 & -6 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right)$$

Nach Rezept 5.2.18 gilt also

$$A^{-1} = \begin{pmatrix} -42 & 36 & 1 \\ 7 & -6 & 0 \\ -1 & 1 & 0 \end{pmatrix}.$$

Man kann natürlich diese Antwort durch Multiplizieren mit A bestätigen.

Rezept 5.2.20 (Test auf Enthaltensein)**.** Gegeben seien eine Familie $(v_1, \dots, v_k)$ und ein Vektor v in K^n. Zu bestimmen ist, ob v in $\operatorname{Span}_K\{v_1, \dots, v_k\}$ enthalten ist. Das gilt genau dann, wenn das System (A, v) eine Lösung besitzt, wobei A die Matrix mit Spalten v_1, …, v_k ist. Man verwendet also das Rezept 5.2.9.

Rezept 5.2.21 (Test auf Erzeugendensystem)**.** Gegeben sei eine Familie $F = (v_1, \dots, v_k)$ in K^n. Zu bestimmen ist, ob F erzeugend ist. Sei A die $n \times k$-Matrix mit Spalten v_1, …, v_k. Man berechnet den Rang von A mit dem Rezept 5.2.14. Dann ist F genau dann erzeugend, wenn $\operatorname{rg} A = n$. Denn $\operatorname{Span}_K\{v_1, \dots, v_k\} = \operatorname{SR}(A)$ und $\dim_K \operatorname{SR}(A) = \operatorname{rg} A$.

Rezept 5.2.22 (Test auf lineare Unabhängigkeit)**.** Gegeben sei eine Familie $F = (v_1, \dots, v_k)$ in K^n. Zu bestimmen ist, ob F linear unabhängig ist. Sei A die $n \times k$-Matrix mit Spalten v_1, …, v_k. Man berechnet den Rang von A mit dem Rezept 5.2.14. Dann ist F genau dann linear unabhängig, wenn $\operatorname{rg} A = k$. Denn $\operatorname{rg} A = k - \dim_K \ker L_A$ nach der Dimensionsformel für die lineare Abbildung L_A, und $\ker L_A$ besteht aus allen k-Tupeln $(\lambda_1, \dots, \lambda_k)$ mit $\sum_{i=1}^k \lambda_i v_i = 0$.

Rezept 5.2.23 (Einschränkung zu einer Basis)**.** Gegeben sei eine Familie $(v_1, \dots, v_k)$ in K^n. Gesucht ist eine Teilfamilie, die eine Basis von $\operatorname{Span}_K\{v_1, \dots, v_k\}$ ist. Sei A die $n \times k$-Matrix mit Spalten v_1, …, v_k. Seien $k_1 < \dots < k_r$ die Indizes der Pivotspalten einer Zeilenstufenform von A. Dann ist $(v_{k_1}, \dots, v_{k_r})$ eine Basis von $\operatorname{Span}_K\{v_1, \dots, v_k\}$. Denn eine Zeilenstufenform von A hat die Form $Z \cdot A$ mit einem $Z \in \operatorname{GL}_n(K)$, und ihre Pivotspalten $Z \cdot v_{k_i}$ bilden eine Basis von $\operatorname{SR}(Z \cdot A)$ (Proposition 5.1.5(ii)). Nach Proposition 5.2.2(ii) bilden dann die Spalten v_{k_i} eine Basis von $\operatorname{SR}(A)$.

Dieses Rezept hat die folgende zusätzliche Eigenschaft: War die Teilfamilie $(v_1, \dots, v_l)$ bereits linear unabhängig, so gilt $k_i = i$ für alle $i \le l$. Denn die Matrix $Z \cdot \begin{pmatrix} v_1 & \dots & v_l \end{pmatrix}$ ist in Zeilenstufenform und hat Rang l, und damit sind die ersten l Spalten von $Z \cdot A$ Pivotspalten.

Beispiel 5.2.24. Wir betrachten die Vektoren

$$v_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \quad v_5 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

in $\mathbb{F}_3^4$, und wir verwenden das Rezept 5.2.23, um eine Teilfamilie von $(v_1, \dots, v_5)$ zu finden, die eine Basis von $U = \mathrm{Span}_K\{v_1, \dots, v_5\}$ ist:

$$\begin{pmatrix} 2 & 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 \end{pmatrix} \xrightarrow{A_{21}(1)} \begin{pmatrix} 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 \end{pmatrix}$$

$$\xrightarrow[V_{23}]{A_{43}(-1)} \begin{pmatrix} 2 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \xrightarrow{A_{43}(-1)} \begin{pmatrix} 2 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Die Indizes der Pivotspalten sind 1, 2 und 4. Deshalb ist (v_1, v_2, v_4) eine Basis von U.

Rezept 5.2.25 (Ergänzung zu einer Basis)**.** Gegeben sei eine linear unabhängige Familie $(v_1, \dots, v_k)$ in K^n. Gesucht sind Vektoren v_{k+1}, …, v_n, so dass $(v_1, \dots, v_n)$ eine Basis von K^n ist. Dazu verwendet man das Rezept 5.2.23 mit der Familie $(v_1, \dots, v_k, e_1, \dots, e_n)$.

Rezept 5.2.26 (Durchschnitt von Untervektorräumen)**.** Gegeben seien Untervektorräume $U = \mathrm{Span}_K\{u_1, \dots, u_k\}$ und $W = \mathrm{Span}_K\{w_1, \dots, w_l\}$ von K^n. Gesucht ist eine Basis von $U \cap W$. Sei A die $n \times k$-Matrix mit Spalten u_1, …, u_k und sei B die $n \times l$-Matrix mit Spalten w_1, …, w_l. Man führt das Gaußsche Eliminationsverfahren mit der erweiterten Matrix $(A \,|\, B)$ durch, bis sie in Zeilenstufenform $(A' \,|\, B')$ ist. Sei s die Anzahl der Pivotspalten von A' und sei B'' die Matrix bestehend aus den Zeilen $B'_{(s+1)*}$, …, B'_{n*} von B'. Die Matrix B'' ist dann in Zeilenstufenform (und man kann sie weiter in reduzierte Zeilenstufenform bringen). Mithilfe von Rezept 5.1.6 erhält man eine Basis $(v_1, \dots, v_t)$ von $\mathcal{L}(B'', 0) \subset K^l$. Die Menge $\{B \cdot v_1, \dots, B \cdot v_t\}$ ist jetzt ein Erzeugendensystem von $U \cap W$ und man verwendet das Rezept 5.2.23, um eine Basis daraus auszuwählen. Denn $\mathcal{L}(B'', 0) \subset K^l$ ist das Bild von $\mathcal{L}(A' \,|\, B', 0) = \mathcal{L}(A \,|\, B, 0) \subset K^{k+l}$ unter der Projektion $K^{k+l} \to K^l$ auf die letzten l Koordinaten, und $\mathcal{L}(A \,|\, B, 0)$ besteht aus allen Vektoren $(\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l)$, so dass

$$\sum_{i=1}^{k} \lambda_i v_i + \sum_{j=1}^{l} \mu_j w_j = 0.$$

Damit besteht $U \cap W$ aus aller Linearkombinationen $\sum_{j=1}^{l} \mu_j w_j$, deren Koeffizienten die letzten l Koordinaten eines Vektors aus $\mathcal{L}(A \,|\, B, 0)$ sind, d.h., $U \cap W = \{B \cdot v \,|\, v \in \mathcal{L}(B'', 0)\}$.

Beispiel 5.2.27. Wir betrachten die folgenden Untervektorräume von $\mathbb{R}^3$:

$$U = \mathrm{Span}_{\mathbb{R}} \left\{ \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \right\}, \quad W = \mathrm{Span}_{\mathbb{R}} \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}.$$

Wir berechnen den Durchschnitt $U \cap W$ mit Rezept 5.2.26 :

$$(A \,|\, B) = \left(\begin{array}{cc|cc} 0 & 3 & 2 & 1 \\ 1 & 4 & 0 & -1 \\ 2 & 5 & 1 & 0 \end{array}\right) \xrightarrow[V_{12}]{A_{32}(-2)} \left(\begin{array}{cc|cc} 1 & 4 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & -3 & 1 & 2 \end{array}\right)$$

$$\xrightarrow{A_{32}(1)} \left(\begin{array}{cc|cc} 1 & 4 & 0 & -1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 3 & 3 \end{array}\right) = (A' \,|\, B').$$

Die Matrix A' hat zwei Pivotspalten, so dass $B'' = \begin{pmatrix} 3 & 3 \end{pmatrix}$. Der Nullraum $\mathcal{L}(B'', 0) \subset \mathbb{R}^2$ ist von $e_1 - e_2$ erzeugt. Daher bildet der Vektor

$$B \cdot (e_1 - e_2) = \begin{pmatrix} 2 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

ein Erzeugendensystem von $U \cap W$, und damit auch eine Basis.

Rezept 5.2.28 (Summe von Untervektorräumen)**.** Gegeben seien Untervektorräume $U = \mathrm{Span}_K\{u_1, \ldots, u_k\}$ und $W = \mathrm{Span}_K\{w_1, \ldots, w_l\}$ von K^n. Gesucht ist eine Basis von $U+W$. Dazu wendet man das Rezept 5.2.23 mit der Familie $(u_1, \ldots, u_k, w_1, \ldots, w_l)$ an.

Rezept 5.2.29 (komplementäre Untervektorräume)**.** Gegeben sei ein Untervektorraum $U = \mathrm{Span}_K\{u_1, \ldots, u_k\}$ von K^n. Gesucht ist eine Basis eines zu U komplementären Untervektorraums. Man verwendet das Rezept 5.2.25, um eine Basis von K^n der Gestalt $(u_{i_1}, \ldots, u_{i_r}, e_{j_1}, \ldots, e_{j_s})$ zu finden. Dann ist $(e_{j_1}, \ldots, e_{j_s})$ eine Basis eines zu U komplementären Untervektorraums.

Rezept 5.2.30 (Quotientenvektorräume)**.** Gegeben sei ein Untervektorraum $U = \mathrm{Span}_K\{u_1, \ldots, u_k\}$ von K^n. Gesucht ist eine Basis des Quotientenvektorraums K^n/U. Man verwendet das Rezept 5.2.29, um eine Basis $(w_1, \ldots, w_l)$ eines zu U komplementären Untervektorraums zu finden. Dann ist $(w_1 + U, \ldots, w_l + U)$ eine Basis von K^n/U.

Beispiel 5.2.31. Sei $K = \mathbb{R}$ und

$$U = \mathrm{Span}_{\mathbb{R}} \left\{ \begin{pmatrix} 0 \\ 1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^4.$$

Wir verwenden die Rezepte 5.2.29 und 5.2.30, um einen komplementären Untervektorraum zu U und eine Basis von $\mathbb{R}^4/U$ zu bestimmen:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[A_{42}(-3)]{\substack{A_{32}(1) \\ V_{12}}} \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 3 & 0 & -3 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[A_{42}(-3)]{A_{32}(1)} \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & -3 & -3 & 0 & 1 \end{pmatrix} \xrightarrow{A_{43}(3)} \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \end{pmatrix} = A'.$$

Die Spalten von A entsprechend den Pivotspalten von A' bilden eine Basis von $\mathbb{R}^4$. Damit ist (e_1, e_3) eine Basis eines zu U komplementären Untervektorraums, und ist $(e_1 + U, e_3 + U)$ eine Basis von $\mathbb{R}^4/U$.

Rezept 5.2.32 (Smith-Normalform)**.** Gegeben sei eine Matrix $A \in M_{m\times n}(K)$. Gesucht sind Basen B von K^n und C von K^m, so dass $[L_A]_C^B$ in Smith-Normalform ist (siehe Satz 4.2.46). Man verwendet das Rezept 5.2.15, um eine Basis $(v_1, \ldots, v_k)$ von $\ker L_A$ zu finden, und das Rezept 5.2.25, um die zu einer Basis $B = (v_1, \ldots, v_n)$ von K^n zu ergänzen. Dann verwendet man wieder das Rezept 5.2.25, um $(A \cdot v_{k+1}, \ldots, A \cdot v_n)$ zu einer Basis C von K^m zu ergänzen.

Alternativ dazu kann man beobachten, dass eine Matrix A genau dann in Smith-Normalform ist, wenn beide A und $A^\intercal$ in reduzierter Zeilenstufenform sind. Man kann erstens A auf reduzierte Zeilenstufenform A' durch elementare Zeilenumformungen bringen, und zweitens A' auf Smith-Normalform A'' durch elementare Spaltenumformungen bringen. Dabei findet man Elementarmatrizen $Z_1, \ldots, Z_k, W_1, \ldots, W_l$, so dass

$$A'' = Z_k \cdot \ldots \cdot Z_1 \cdot A \cdot W_1 \cdot \ldots \cdot W_l.$$

Nach der Basiswechselformel (Proposition 4.2.43) bilden die Spalten von $W_1 \cdot \ldots \cdot W_l$ und $Z_1^{-1} \cdot \ldots \cdot Z_k^{-1}$ Basen B und C mit der gewünschten Eigenschaft.

5.3 Die Determinante

Sei $n \in \mathbb{N}$. Die Determinante ist eine kanonische Abbildung $\det\colon M_n(K) \to K$ mit folgender wichtigen Eigenschaft: Eine $n \times n$-Matrix A ist genau dann invertierbar, wenn $\det(A)$ nicht null ist. Außerdem erlaubt uns die Determinante, direkte Formeln für die Koeffizienten von A^{-1} und für die Lösung eines linearen Gleichungssystems $A \cdot x = b$ zu schreiben (obwohl solche Formeln nur von theoretischer Bedeutung sind; praktisch ist das Gaußsche Eliminationsverfahren viel schneller).

5.3.1 Das Vorzeichen einer Permutation

Sei $n \in \mathbb{N}$. Zur Erinnerung ist die symmetrische Gruppe S_n die Menge aller Permutationen von $\{1, \dots, n\}$ mit der Verknüpfung $\circ$ (siehe Abschnitt 2.2.2). Falls $n \geq 3$ ist diese Gruppe nicht abelsch.

Definition 5.3.1 (Zyklus, Transposition)**.** Sei $n \in \mathbb{N}$ und sei $k \in \{1, \dots, n\}$. Eine Permutation $\sigma \in S_n$ heißt *Zyklus* der Länge k, wenn es paarweise verschiedene Elemente $a_1, \dots, a_k$ gibt, so dass $\sigma(a_i) = a_{i+1}$ für alle $i < k$, $\sigma(a_k) = a_1$, und $\sigma(b) = b$ für alle anderen Elemente b. Man schreibt dann

$$\sigma = (a_1\ a_2\ \dots\ a_k).$$

Ein Zyklus der Länge 2 heißt *Transposition.*

Beispiel 5.3.2. Mit der Zyklenschreibweise gilt:

$$S_2 = \{\mathrm{id}, (1\ 2)\},$$
$$S_3 = \{\mathrm{id}, (1\ 2), (1\ 3), (2\ 3), (1\ 2\ 3), (1\ 3\ 2)\}.$$

Bemerkung 5.3.3. Für einen Zyklus $(a_1\ a_2\ \dots\ a_k)$ gilt

$$(a_1\ a_2\ \dots\ a_k) = (a_1\ a_2) \circ \cdots \circ (a_{k-1}\ a_k).$$

Insbesondere ist jeder Zyklus der Länge k ein Produkt von $k-1$ Transpositionen.

Lemma 5.3.4. *Sei $n \in \mathbb{N}$. In der Gruppe S_n ist jedes Element ein Produkt von Transpositionen.*

Beweis. Wir verwenden Induktion über n. Falls $n = 0$ besteht S_n nur aus dem neutralen Element, das ein leeres Produkt ist. Angenommen gilt die Aussage für S_n, und sei $\sigma \in S_{n+1}$. Sei $\tau \in S_{n+1}$ die wie folgt definierte Permutation:

$$\tau = \begin{cases} \mathrm{id}, & \text{falls } \sigma(n+1) = n+1, \\ (n+1\ \sigma(n+1)), & \text{andernfalls.} \end{cases}$$

Dann gilt $(\tau \circ \sigma)(n+1) = n+1$. Das heißt, die Permutation $\tau \circ \sigma$ shränkt sich zu einer Permutation von $\{1, \dots, n\}$ ein. Nach der Induktionsvoraussetzung gibt es also Transpositionen $\tau_1, \dots, \tau_k$ in S_{n+1}, die $n+1$ festhalten, so dass $\tau \circ \sigma = \tau_k \circ \cdots \circ \tau_1$. Dann gilt $\sigma = \tau \circ \tau_k \circ \cdots \circ \tau_1$. □

Satz 5.3.5. *Sei $n \in \mathbb{N}$. Es gibt genau eine Abbildung*

$$\mathrm{sgn}\colon S_n \to \{1, -1\}$$

mit folgenden zwei Eigenschaften:

(i) sgn *ist ein* Gruppenhomomorphismus*, d.h.: Für alle $\sigma, \tau \in S_n$ gilt*

$$\mathrm{sgn}(\tau \circ \sigma) = \mathrm{sgn}(\tau) \cdot \mathrm{sgn}(\sigma).$$

(ii) *Für alle Transpositionen $\sigma \in S_n$ gilt* $\operatorname{sgn}(\sigma) = -1$.

Außerdem:

(iii) *Für alle Zyklen $\sigma \in S_n$ der Länge k gilt* $\operatorname{sgn}(\sigma) = (-1)^{k-1}$.

Beweis. Die Aussage (iii) folgt aus (i) und (ii), da jeder Zyklus der Länge k die Komposition von $k-1$ Transpositionen ist (siehe Bemerkung 5.3.3).

Zur Eindeutigkeit. Seien $s, t\colon S_n \to \{1, -1\}$ zwei Abbildungen, die (i) und (ii) erfüllen, und sei $\sigma \in S_n$. Nach Lemma 5.3.4 ist σ eine Komposition von Transpositionen $\tau_n \circ \cdots \circ \tau_1$. Nach (i) und (ii) gilt:

$$s(\sigma) = s(\tau_n) \cdot \ldots \cdot s(\tau_1) = t(\tau_n) \cdot \ldots \cdot t(\tau_1) = t(\sigma).$$

Also ist $s = t$.

Zur Existenz. Sei $\mathcal{Z}$ die Menge aller zweielementigen Teilmengen von $\{1, \ldots, n\}$:

$$\mathcal{Z} = \{I \subset \{1, \ldots, n\} \mid |I| = 2\}.$$

Sei $\sigma \in S_n$ und $I \in \mathcal{Z}$. Falls $I = \{i, j\}$ ist die rationale Zahl

$$e_I(\sigma) := \frac{\sigma(i) - \sigma(j)}{i - j} \in \mathbb{Q}^\times$$

unabhängig von der Reihenfolge von i und j. Für $\sigma, \tau \in S_n$ gilt:

$$e_I(\tau \circ \sigma) = \frac{\tau(\sigma(i)) - \tau(\sigma(j))}{i - j} = \frac{\tau(\sigma(i)) - \tau(\sigma(j))}{\sigma(i) - \sigma(j)} \cdot \frac{\sigma(i) - \sigma(j)}{i - j} = e_{\sigma(I)}(\tau) \cdot e_I(\sigma). \quad (5.3.6)$$

Man definiert:

$$\begin{aligned} \operatorname{sgn}\colon S_n &\to \mathbb{Q}^\times, \\ \sigma &\mapsto \prod_{I \in \mathcal{Z}} e_I(\sigma). \end{aligned}$$

Die Abbildung $I \mapsto \sigma(I)$ ist eine Permutation der endlichen Menge $\mathcal{Z}$, und damit gilt

$$\prod_{I \in \mathcal{Z}} e_{\sigma(I)}(\tau) = \prod_{I \in \mathcal{Z}} e_I(\tau). \quad (5.3.7)$$

Aus (5.3.6) und (5.3.7) folgt:

$$\operatorname{sgn}(\tau \circ \sigma) = \prod_{I \in \mathcal{Z}} e_{\sigma(I)}(\tau) \cdot \prod_{I \in \mathcal{Z}} e_I(\sigma) = \prod_{I \in \mathcal{Z}} e_I(\tau) \cdot \prod_{I \in \mathcal{Z}} e_I(\sigma) = \operatorname{sgn}(\tau) \cdot \operatorname{sgn}(\sigma).$$

Das heißt, $\operatorname{sgn}\colon S_n \to \mathbb{Q}^\times$ ist ein Gruppenhomomorphismus.

Sei σ die Transposition $(k\ l)$. Wir zeigen jetzt, dass $\operatorname{sgn}(\sigma) = -1$. Sei $I = \{i, j\} \in \mathcal{Z}$. Wenn $I \cap \{k, l\} = \varnothing$, dann ist $e_I(\sigma) = \frac{i-j}{i-j} = 1$, und damit können wir diese Faktoren in dem Produkt $\operatorname{sgn}(\sigma)$ ignorieren. Für jedes $m \notin \{k, l\}$ gilt $e_{\{k,m\}}(\sigma) = \frac{l-m}{k-m} = e_{\{l,m\}}(\sigma)^{-1}$, so dass diese Faktoren in dem Produkt $\operatorname{sgn}(\sigma)$ paarweise verschwinden. Es bleibt übrig die Teilmenge $I = \{k, l\}$ selbst, für die gilt $e_{\{k,l\}}(\sigma) = \frac{l-k}{k-l} = -1$. Also ist $\operatorname{sgn}(\sigma) = -1$, wie behauptet. Aus Lemma 5.3.4 folgt schließlich, dass $\operatorname{sgn}(S_n) \subset \{1, -1\}$, und damit erhalten wir einen Gruppenhomomorphismus $\operatorname{sgn}\colon S_n \to \{1, -1\}$ mit den gewünschten Eigenschaften. □

Definition 5.3.8 (Vorzeichen, gerade/ungerade Permutationen)**.** Sei $n \in \mathbb{N}$ und $\sigma \in S_n$. Die Zahl $\operatorname{sgn}(\sigma) \in \{1, -1\}$ heißt das *Vorzeichen* oder das *Signum* von σ. Permutationen σ mit $\operatorname{sgn}(\sigma) = 1$ heißen *gerade* und die mit $\operatorname{sgn}(\sigma) = -1$ heißen *ungerade*.

Bemerkung 5.3.9. Nach dem Lemma 5.3.4 und dem Satz 5.3.5 ist eine Permutation $\sigma \in S_n$ genau dann gerade bzw. ungerade, wenn sie die Komposition einer geraden bzw. ungeraden Anzahl von Transpositionen ist.

Bemerkung 5.3.10 (alternierende Gruppe)**.** Da $\operatorname{sgn}\colon S_n \to \{1, -1\}$ ein Gruppenhomomorphismus ist, ist die Komposition zweier geraden Permutationen sowie das Inverse einer geraden Permutation wieder gerade. Insbesondere ist die Teilmenge $A_n \subset S_n$ aller geraden Permutationen eine Gruppe bzgl. $\circ$. Sie heißt die *alternierende Gruppe* vom Grad n.

5.3.2 Determinantenfunktionen

Um die Determinante $\det\colon M_n(K) \to K$ zu verstehen, ist es hilfreich, der Vektorraum $M_n(K)$ mit $(K^n)^n$ zu identifizieren, indem wir eine Matrix A als das n-Tupel ihrer Spalten $(A_{*1}, \ldots, A_{*n})$ auffassen. Die Determinante $\det\colon (K^n)^n \to K$ ist dann *keine* lineare Abbildung, aber sie ist linear bezüglich jedes ihrer n Argumente im folgenden Sinne:

Definition 5.3.11 (multilineare Abbildung)**.** Seien $n \in \mathbb{N}$ und $V_1, \ldots, V_n, W$ Vektorräume über K. Eine Abbildung

$$f\colon V_1 \times \cdots \times V_n \to W$$

heißt *multilinear* oder *n-linear* (*bilinear* wenn $n = 2$), wenn sie bezüglich jedes ihrer n Argumente linear ist, d.h.: Für jedes $i \in \{1, \ldots, n\}$ und jedes

$$(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \in V_1 \times \cdots \times V_{i-1} \times V_{i+1} \times \cdots \times V_n$$

ist die folgende Abbildung linear:

$$\begin{aligned} V_i &\to W, \\ v &\mapsto f(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n). \end{aligned}$$

Definition 5.3.12 (symmetrisch, antisymmetrisch, alternierend)**.** Sei $n \in \mathbb{N}$, seien V, W Vektorräume über K und sei $f\colon V^n \to W$ eine n-lineare Abbildung.

- f heißt *symmetrisch*, wenn für alle $i, j \in \{1, \ldots, n\}$ mit $i < j$ und alle $(v_1, \ldots, v_n) \in V^n$ gilt
$$f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n).$$
- f heißt *antisymmetrisch*, wenn für alle $i, j \in \{1, \ldots, n\}$ mit $i < j$ und alle $(v_1, \ldots, v_n) \in V^n$ gilt
$$f(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -f(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n).$$
- f heißt *alternierend*, wenn für alle $i, j \in \{1, \ldots, n\}$ mit $i < j$ und alle $(v_1, \ldots, v_n) \in V^n$ mit $v_i = v_j$ gilt
$$f(v_1, \ldots, v_n) = 0.$$

Beispiel 5.3.13.

(i) Die Abbildung

$$K^n \times K^n \to K, \quad (x, y) \mapsto \sum_{i=1}^{n} x_i y_i,$$

ist eine symmetrische bilineare Abbildung.

(ii) Sei $K = \mathbb{R}$. Die Multiplikation von komplexen Zahlen $\mathbb{C} \times \mathbb{C} \to \mathbb{C}$ ist eine symmetrische bilineare Abbildung. Die Multiplikation von Quaternionen $\mathbb{H} \times \mathbb{H} \to \mathbb{H}$ ist eine bilineare Abbildung, die weder symmetrisch noch antisymmetrisch ist.

Bemerkung 5.3.14 (alternierend vs. antisymmetrisch)**.** Jede alternierende n-lineare Abbildung $f\colon V^n \to W$ ist antisymmetrisch. Ohne Beschränkung der Allgemeinheit (indem man alle Argumente von f bis auf zwei festhält), können wir $n = 2$ annehmen. Für alle $(v, v') \in V^2$ gilt dann:

$$0 = f(v + v', v + v') = f(v, v) + f(v, v') + f(v', v) + f(v', v') = f(v, v') + f(v', v),$$

und daher $f(v, v') = -f(v', v)$. Die Umkehrung gilt wenn die Charakteristik von K nicht 2 ist: In diesem Fall impliziert die Gleichung $f(v, v) = -f(v, v)$, dass $f(v, v) = 0$. Aber wenn die Charakteristik von K gleich 2 ist, dann sind „symmetrisch“ und „antisymmetrisch“ äquivalent, und „alternierend“ ist eine stärkere Bedingung.

Bemerkung 5.3.15. Multilineare Abbildungen $V_1 \times \cdots \times V_n \to W$ bilden einen Untervektorraum des Vektorraums $\mathrm{Abb}(V_1 \times \cdots \times V_n, W)$. Wenn $V_1 = \cdots = V_n = V$, dann bilden symmetrische, antisymmetrische und alternierende Abbildungen $V^n \to W$ weitere Untervektorräume davon.

Definition 5.3.16 (Determinantenfunktion)**.** Sei V ein K-Vektorraum der endlichen Dimension n. Eine *Determinantenfunktion* auf V ist eine alternierende n-lineare Abbildung $V^n \to K$. Der K-Vektorraum aller Determinantenfunktionen auf V wird mit $\mathrm{Det}(V)$ bezeichnet.

Proposition 5.3.17. *Sei $n \in \mathbb{N}$. Die Abbildung*

$$\begin{aligned} \Delta\colon (K^n)^n &\to K, \\ (v_1, \dots, v_n) &\mapsto \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} v_{i\sigma(i)}, \end{aligned}$$

ist eine Determinantenfunktion auf K^n mit $\Delta(e_1, \dots, e_n) = 1$.

Beweis. Es gilt

$$\Delta(e_1, \dots, e_n) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} \delta_{i\sigma(i)} = 1,$$

da das Produkt $\prod_{i=1}^{n} \delta_{i\sigma(i)}$ immer null ist, außer wenn $\sigma = \mathrm{id}$, in welchem Fall ist es gleich 1.

- Δ *ist multilinear.* Seien $i \in \{1, \dots, n\}$, $v_1, \dots, v_i, v_i', \dots, v_n \in K^n$ und $\lambda, \lambda' \in K$. Dann gilt

$$\begin{aligned} \Delta(v_1, \dots, \lambda v_i + \lambda' v_i', \dots, v_n) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot v_{1\sigma(1)} \cdots (\lambda v_{i\sigma(i)} + \lambda' v_{i\sigma(i)}') \cdots v_{n\sigma(n)} \\ &= \lambda \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot v_{1\sigma(1)} \cdots v_{i\sigma(i)} \cdots v_{n\sigma(n)} \\ &\quad + \lambda' \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot v_{1\sigma(1)} \cdots v_{i\sigma(i)}' \cdots v_{n\sigma(n)} \\ &= \lambda \Delta(v_1, \dots, v_i, \dots, v_n) + \lambda' \Delta(v_1, \dots, v_i', \dots, v_n). \end{aligned}$$

- Δ *ist alternierend.* Sei $v_i = v_j$ mit $i < j$. Man beachte, dass die Abbildung

$$A_n \to S_n \setminus A_n, \quad \sigma \mapsto \sigma \circ (i\ j),$$

bijektiv ist (mit Umkehrabbildung $\tau \mapsto \tau \circ (i\ j)$). Es gilt

$$\begin{aligned}\Delta(v_1, \dots, v_n) &= \sum_{\sigma \in A_n} v_{1\sigma(1)} \cdots v_{i\sigma(i)} \cdots v_{j\sigma(j)} \cdots v_{n\sigma(n)} \\ &\quad - \sum_{\tau \in S_n \setminus A_n} v_{1\tau(1)} \cdots v_{i\tau(i)} \cdots v_{j\tau(j)} \cdots v_{n\tau(n)} \\ &= \sum_{\sigma \in A_n} v_{1\sigma(1)} \cdots v_{i\sigma(i)} \cdots v_{j\sigma(j)} \cdots v_{n\sigma(n)} \\ &\quad - \sum_{\sigma \in A_n} v_{1\sigma(1)} \cdots v_{i\sigma(j)} \cdots v_{j\sigma(i)} \cdots v_{n\sigma(n)} \\ &= 0,\end{aligned}$$

da $v_{ik} = v_{jk}$ für alle $k \in \{1, \dots, n\}$. □

Die Formel für die Determinantenfunktion Δ sieht vielleicht eigenartig aus. Aber wir zeigen jetzt, dass $\{\Delta\}$ ein Erzeugendensystem von $\mathrm{Det}(K^n)$ ist, das heißt: Jede Determinantenfunktion auf K^n ist gleich $\lambda \cdot \Delta$ mit einem Skalar $\lambda \in K$.

Lemma 5.3.18. *Seien V, W Vektorräume über K, $n \in \mathbb{N}$ und $f\colon V^n \to W$ eine antisymmetrische n-lineare Abbildung. Für alle Vektoren $v_1, \dots, v_n \in V$ und alle Permutationen $\sigma \in S_n$ gilt*

$$f(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \mathrm{sgn}(\sigma) \cdot f(v_1, \dots, v_n).$$

Beweis. Wenn σ eine Transposition ist, folgt die Aussage aus der Antisymmetrie von f. Gilt die Aussage für σ und τ, so gilt sie auch für $\sigma \circ \tau$, da $\mathrm{sgn}(\sigma \circ \tau) = \mathrm{sgn}(\sigma) \cdot \mathrm{sgn}(\tau)$. Da jede Permutation eine Komposition von Transpositionen ist (Lemma 5.3.4), gilt die Aussage im Allgemeinen. □

Bemerkung 5.3.19. Ist $f\colon V^n \to W$ symmetrisch, so folgt unmittelbar aus Lemma 5.3.4, dass

$$f(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = f(v_1, \dots, v_n)$$

für alle $v_1, \dots, v_n \in V$ und alle $\sigma \in S_n$. Das ist tatsächlich der geschichtliche Grund dafür, dass die Gruppe S_n als symmetrische Gruppe bezeichnet wird. Ist andererseits $f\colon V^n \to W$ antisymmetrisch (z.B. alternierend), so ist f invariant gegenüber *geraden* Permutationen seiner Argumente, und deswegen wird die Gruppe A_n als alternierende Gruppe bezeichnet.

Lemma 5.3.20 (Funktorialität von Determinantenfunktionen)**.** *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen derselben endlichen Dimension n und sei $\delta \in \mathrm{Det}(W)$. Dann ist die Abbildung*

$$\begin{aligned} f^*(\delta)\colon V^n &\to K, \\ (v_1, \dots, v_n) &\mapsto \delta(f(v_1), \dots, f(v_n)),\end{aligned}$$

eine Determinantenfunktion auf V. Außerdem ist die so definierte Abbildung $f^*\colon \mathrm{Det}(W) \to \mathrm{Det}(V)$ linear.

Beweis. Man hat zu zeigen, dass $f^*(\delta)$ n-linear und alternierend ist. Die n-Linearität folgt aus der n-Linearität von δ und der Linearität von f, und es ist klar, dass $f^*(\delta)$ auch alternierend ist. Die Linearität von f^* ist auch klar. □

Satz 5.3.21 (Klassifikation von Determinantenfunktionen)**.** *Sei V ein K-Vektorraum der endlichen Dimension n und sei $B = (b_1, \dots, b_n)$ eine Basis von V. Dann ist die lineare Abbildung*

$$\begin{aligned} \mathrm{Det}(V) &\to K, \\ (\delta\colon V^n \to K) &\mapsto \delta(b_1, \dots, b_n),\end{aligned}$$

ein Isomorphismus. Die Umkehrabbildung schickt $1 \in K$ *auf die Determinantenfunktion*

$$\begin{aligned} \Delta_B \colon V^n &\to K, \\ (v_1, \dots, v_n) &\mapsto \Delta([v_1]_B, \dots, [v_n]_B). \end{aligned}$$

Beweis. Zur Injektivität. Sei $\delta \in \mathrm{Det}(V)$ eine Determinantenfunktion mit $\delta(b_1, \dots, b_n) = 0$. Es gilt dann $\delta(b_{i_1}, \dots, b_{i_n}) = 0$ für alle $i_1, \dots, i_n \in \{1, \dots, n\}$: Falls zwei Indizes i_k gleich sind, folgt das aus der Definition einer alternierenden Abbildung, sonst ist

$$\delta(b_{i_1}, \dots, b_{i_n}) = \pm\delta(b_1, \dots, b_n) = 0$$

nach Lemma 5.3.18. Seien nun $v_1, \dots, v_n \in V$ beliebige Vektoren, und sei $v_i = \sum_{j=1}^n \lambda_{ij} b_j$. Da δ multilinear ist, gilt

$$\delta(v_1, \dots, v_n) = \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n \lambda_{1j_1} \dots \lambda_{nj_n} \delta(b_{j_1}, \dots, b_{j_n}) = 0.$$

Also ist $\delta = 0$.

Zur Surjektvität. Es genügt zu zeigen, dass Δ_B eine Determinantenfunktion auf V mit $\Delta_B(b_1, \dots, b_n) = 1$ ist. Nach Definition ist $\Delta_B = (\varphi_B^{-1})^*(\Delta)$, wobei $\varphi_B \colon K^n \xrightarrow{\sim} V$ der der Basis B zugehörige Isomorphismus ist. Die gewünschten Eigenschaften von Δ_B folgen also aus dem Lemma 5.3.20 und der Proposition 5.3.17. □

Bemerkung 5.3.22. Sei $K = \mathbb{R}$. Für die Determinantenfunktion $\Delta \colon (\mathbb{R}^2)^2 \to \mathbb{R}$ gilt

$$|\Delta(v_1, v_2)| = \text{der Flächeninhalt des Parallelograms mit Eckpunkten } 0, v_1, v_2, v_1 + v_2.$$

Wenn $\Delta(v_1, v_2)$ nicht null ist, ist es genau dann positiv, wenn der kürzeste Winkel von v_1 nach v_2 gegen den Uhrzeigersinn läuft. Allgemeiner ist $|\Delta(v_1, \dots, v_n)|$ das Volumen des von $v_1, \dots, v_n$ aufgespannten Parallelotops in $\mathbb{R}^n$, im Sinne der Maßtheorie. Man kann deshalb Determinantenfunktionen als Varianten mit Vorzeichen des Volumens auffassen, die auch über beliebigen Körpern K sinnvoll sind.

5.3.3 Die Determinante einer Matrix

Definition 5.3.23 (Determinante einer Matrix)**.** Seien $n \in \mathbb{N}$ und $A \in M_n(K)$. Die *Determinante* von A ist der Skalar

$$\det(A) = \Delta(A_{*1}, \dots, A_{*n}),$$

wobei $\Delta \colon (K^n)^n \to K$ die Determinantenfunktion aus Proposition 5.3.17 ist.

Nach Definition von Δ gilt also

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot \prod_{i=1}^n A_{\sigma(i)i}. \tag{5.3.24}$$

Diese Formel für die Determinante heißt die *Leibniz-Formel*.

Beispiel 5.3.25. Es gilt

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Proposition 5.3.26 (Eigenschaften der Determinante)**.** *Sei* $n \in \mathbb{N}$ *und seien* $A, B \in M_n(K)$*. Dann gilt:*

(i) $\det(I_n) = 1$.

(ii) $\det(A \cdot B) = \det(A) \cdot \det(B)$.

(iii) *Ist A invertierbar, so ist* $\det(A^{-1}) = \det(A)^{-1}$.

(iv) $\det(A^\intercal) = \det(A)$.

Beweis. Zu (i). $\det(I_n) = \Delta(e_1, \ldots, e_n) = 1$.

Zu (ii). Nach Lemma 5.3.20 ist $L_A^*(\Delta)$ eine Determinantenfunktion auf K^n mit

$$L_A^*(\Delta)(e_1, \ldots, e_n) = \Delta(A_{*1}, \ldots, A_{*n}) = \det(A).$$

Nach der Klassifikation von Determinantenfunktionen (Satz 5.3.21) gilt $L_A^*(\Delta) = \det(A) \cdot \Delta$. Daher gilt

$$\det(A \cdot B) = L_A^*(\Delta)(B_{*1}, \ldots, B_{*n}) = \det(A) \cdot \Delta(B_{*1}, \ldots, B_{*n}) = \det(A) \cdot \det(B).$$

Zu (iii). Dies folgt unmittelbar aus (i) und (ii).

Zu (iv). Wir verwenden die Leibniz-Formel (5.3.24). Da S_n eine Gruppe ist, ist die Abbildung $S_n \to S_n$, $\tau \mapsto \tau^{-1}$, eine Permutation von S_n. Außerdem ist nach Definition jedes $\tau \in S_n$ eine Permutation der Indexmenge $\{1, \ldots, n\}$. Es gilt also

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} A_{\sigma(i)i} && \text{(Leibniz-Formel)} \\
&= \sum_{\tau \in S_n} \operatorname{sgn}(\tau^{-1}) \cdot \prod_{i=1}^{n} A_{\tau^{-1}(i)i} && \text{(Permutation } \tau \mapsto \tau^{-1}\text{)} \\
&= \sum_{\tau \in S_n} \operatorname{sgn}(\tau^{-1}) \cdot \prod_{j=1}^{n} A_{\tau^{-1}(\tau(j))\tau(j)} && \text{(Permutation } j \mapsto \tau(j)\text{)} \\
&= \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \cdot \prod_{j=1}^{n} A_{j\tau(j)} && \text{(sgn ist ein Gruppenhomomorphismus)} \\
&= \det(A^\intercal). && \text{(Leibniz-Formel)}
\end{aligned}$$

□

Bemerkung 5.3.27. Die Gleichung $\det(A) = \det(A^\intercal)$ bedeutet, dass

$$\Delta(A_{*1}, \ldots, A_{*n}) = \Delta(A_{1*}, \ldots, A_{n*}).$$

Das heißt, es macht keinen Unterschied, ob wir die Spalten oder die Zeilen von A in der Definition der Determinante verwenden.

Bemerkung 5.3.28. Nach Proposition 5.3.26(ii,iii) schränkt sich die Determinante zu einem Gruppenhomomorphismus $\det\colon \mathrm{GL}_n(K) \to K^\times$ ein.

Proposition 5.3.29 (Determinante und elementare Zeilenumformungen)**.** *Sei $n \in \mathbb{N}$ und $A \in M_n(K)$.*

(i) *Seien $i, j \in \{1, \ldots, n\}$ mit $i < j$. Dann ist*

$$\det(V_{ij} \cdot A) = -\det(A).$$

(ii) *Seien $i \in \{1, \ldots, n\}$ und $\lambda \in K^\times$. Dann ist*

$$\det(M_i(\lambda) \cdot A) = \lambda \cdot \det(A).$$

(iii) *Seien $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ und sei $\alpha \in K$. Dann ist*

$$\det(A_{ij}(\alpha) \cdot A) = \det(A).$$

Beweis. Nach Proposition 5.3.26(iv) ist $\det(A) = \Delta(A_{1*}, \dots, A_{n*})$. Die elementare Zeilenumformung $A \mapsto V_{ij} \cdot A$ vertauscht zwei Zeilen. Die erste Aussage folgt daraus, dass Δ antisymmetrisch ist. Die zweite und dritte Aussagen folgen ähnlich daraus, dass Δ n-linear und alternierend ist. □

Notation 5.3.30. Sei A eine $m \times n$-Matrix, $i \in \{1, \dots, m\}$ und $j \in \{1, \dots, n\}$. Wir bezeichnen mit $A[i, j]$ die $(m-1) \times (n-1)$-Matrix, die aus A entsteht, wenn man die i-te Zeile und j-te Spalte streicht.

Satz 5.3.31 (Laplacescher Entwicklungssatz)**.** *Sei $n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j}$ eine $n \times n$-Matrix über K.*

(i) (Entwicklung nach der j-ten Spalte) *Für jedes $j \in \{1, \dots, n\}$ gilt*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A[i, j]).$$

(ii) (Entwicklung nach der i-ten Zeile) *Für jedes $i \in \{1, \dots, n\}$ gilt*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A[i, j]).$$

Beweis. Die zweite Aussage folgt aus der ersten und Proposition 5.3.26(iv). Wir beweisen (i) mithilfe von der Leibniz-Formel. Sei $S_n^{j \mapsto i} \subset S_n$ die Teilmenge aller Permutationen, die j auf i abbilden. Für festes j ist dann $\{S_n^{j \mapsto 1}, \dots, S_n^{j \mapsto n}\}$ eine Partition von S_n. Es gibt außerdem eine Bijektion

$$S_n^{j \mapsto i} \to S_{n-1}, \quad \sigma \mapsto \sigma_{ij},$$

so dass folgendes Quadrat kommmutiert:

$$\begin{array}{ccc} \{1, \dots, n\} \setminus \{j\} & \xrightarrow{\sigma} & \{1, \dots, n\} \setminus \{i\} \\ \uparrow & & \uparrow \\ \{1, \dots, n-1\} & \xrightarrow{\sigma_{ij}} & \{1, \dots, n-1\}. \end{array}$$

Dabei sind die vertikalen Pfeile die ordnungserhaltenden Bijektionen.

Behauptung. Es gilt $\operatorname{sgn}(\sigma) = (-1)^{i+j} \operatorname{sgn}(\sigma_{ij})$.

Mit dieser Behauptung können wir berechnen:

$$\begin{aligned} \det(A) &= \sum_{i=1}^{n} \sum_{\sigma \in S_n^{j \mapsto i}} \operatorname{sgn}(\sigma) \cdot \prod_{k=1}^{n} a_{\sigma(k)k} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \sum_{\sigma \in S_n^{j \mapsto i}} \operatorname{sgn}(\sigma_{ij}) \cdot \prod_{l=1}^{n-1} A[i, j]_{\sigma_{ij}(l)l} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \sum_{\tau \in S_{n-1}} \operatorname{sgn}(\tau) \cdot \prod_{l=1}^{n-1} A[i, j]_{\tau(l)l} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A[i, j]). \end{aligned}$$

Es bleibt die Behauptung nachzuprüfen. Sei $\tilde{\sigma}_{ij} \in S_n$ die Fortsetzung von σ_{ij} auf $\{1, \dots, n\}$, die n auf n abbildet, so dass $\operatorname{sgn}(\sigma_{ij}) = \operatorname{sgn}(\tilde{\sigma}_{ij})$. Dann gilt

$$\sigma = (i \; i+1 \; \dots \; n) \circ \tilde{\sigma}_{ij} \circ (n \; \dots \; j+1 \; j).$$

Nach dem Satz 5.3.5(iii) ist $\operatorname{sgn}(i\ i+1\ \dots\ n) = (-1)^{n-i}$ und $\operatorname{sgn}(n\ \dots\ j+1\ j) = (-1)^{n-j}$, und damit

$$\operatorname{sgn}(\sigma) = (-1)^{2n-i-j}\operatorname{sgn}(\tilde{\sigma}_{ij}) = (-1)^{i+j}\operatorname{sgn}(\sigma_{ij}),$$

wie behauptet. □

Beispiel 5.3.32. Sei

$$A = \begin{pmatrix} 1 & 0 & -2 \\ 3 & 2 & -1 \\ -3 & 0 & 5 \end{pmatrix}.$$

Der Laplacesche Entwicklungssatz liefert sechs verschiedene Formeln für $\det(A)$, eine für jede Spalte und Zeile.

(i) Die einfachste Berechnung von $\det(A)$ ist voraussichtlich durch Entwicklung nach der zweiten Spalte, die nur einen Nicht-Null-Koeffizient enthält. Dies ergibt:

$$\det(A) = (-1)^{2+2} \cdot 2 \cdot \det \begin{pmatrix} 1 & -2 \\ -3 & 5 \end{pmatrix} = 2 \cdot (1 \cdot 5 - (-2) \cdot (-3)) = -2.$$

(ii) Entwicklung nach der ersten Zeile ergibt:

$$\det(A) = 1 \cdot \det \begin{pmatrix} 2 & -2 \\ 0 & 5 \end{pmatrix} + (-2) \cdot \det \begin{pmatrix} 3 & 2 \\ -3 & 0 \end{pmatrix} = 10 - 2 \cdot 6 = -2.$$

(iii) Entwicklung nach der dritten Spalte ergibt:

$$\begin{aligned} \det(A) &= (-2) \cdot \det \begin{pmatrix} 3 & 2 \\ -3 & 0 \end{pmatrix} - (-1) \cdot \det \begin{pmatrix} 1 & 0 \\ -3 & 0 \end{pmatrix} + 5 \cdot \det \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} \\ &= -2 \cdot 6 + 1 \cdot 0 + 5 \cdot 2 = -2. \end{aligned}$$

Beispiel 5.3.33. Sei

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Um die Determinante von A effektiv zu berechnen, entwickeln wir sie nach der dritten Spalte und danach nach der ersten Spalte:

$$\det(A) \overset{3.S}{=} -\det \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \overset{1.S}{=} -\det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -(-1) = 1.$$

Korollar 5.3.34 (Determinante einer Dreiecksmatrix)**.** *Sei $n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j}$ eine $n \times n$-Dreiecksmatrix, d.h., so dass $a_{ij} = 0$ für alle $i > j$ bzw. für alle $i < j$. Dann gilt*

$$\det(A) = \prod_{i=1}^{n} a_{ii}.$$

Beweis. Wir verwenden Induktion über n. Wenn $n = 0$ ist $A = I_n$ und damit $\det(A) = 1 = \prod_{i=1}^{0} a_{ii}$. Sei also $n \geq 1$. Entwicklung nach der ersten Spalte (bzw. nach der ersten Zeile) ergibt $\det(A) = a_{11} \cdot \det(A[1,1])$. Nach der Induktionsvoraussetzung ist $\det(A[1,1]) = \prod_{i=2}^{n} a_{ii}$, was die gewünschte Formel ergibt. □

Beispiel 5.3.35. Nach dem Korollar 5.3.34 und der Proposition 5.3.29 kann man auch die Determinante durch das Gaußsche Eliminationsverfahren berechnen. Denn eine quadratische Matrix in Zeilenstufenform ist insbesondere eine Dreiecksmatrix. Als Beispiel berechnen wir die Determinante der folgenden Matrix A:

$$A = \begin{pmatrix} 3 & 1 & 6 \\ 1 & 1 & 7 \\ 2 & 6 & -3 \end{pmatrix} \xrightarrow[A_{32}(-2)]{A_{12}(-3)} \begin{pmatrix} 1 & 1 & 7 \\ 0 & -2 & -15 \\ 0 & 4 & -17 \end{pmatrix}$$
$$\xrightarrow{A_{32}(2)} \begin{pmatrix} 1 & 1 & 7 \\ 0 & -2 & -15 \\ 0 & 0 & -47 \end{pmatrix} = A'.$$

Damit ist $\det(A) = -\det(A') = -94$.

Die folgende Verallgemeinerung des Korollars 5.3.34 ist auch nützlich:

Korollar 5.3.36 (Determinante einer Blockdreiecksmatrix)**.** *Seien $k \in \mathbb{N}$, $n_1, \ldots, n_k \in \mathbb{N} \setminus \{0\}$, $n = \sum_{i=1}^{k} n_i$ und sei A eine $n \times n$-Matrix der Gestalt*

$$A = \begin{pmatrix} A_1 & & & * \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{pmatrix} \quad \textit{bzw.} \quad A = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ * & & & A_k \end{pmatrix},$$

wobei $A_i \in M_{n_i}(K)$. Dann gilt

$$\det(A) = \prod_{i=1}^{k} \det(A_i).$$

Beweis. Es genügt den ersten Fall zu behandeln, indem man die transponierte Matrix betrachtet. Wir verwenden Induktion über n. Wenn $n = 0$ ist die Aussage trivial. Sei also $n \geq 1$. Entwicklung von $\det(A)$ nach der ersten Spalte liefert

$$\det(A) = \sum_{e=1}^{n_1} (-1)^{e+1} (A_1)_{e1} \det(A[e,1]).$$

Nach der Induktionsvoraussetzung ist $\det(A[e,1]) = \det(A_1[e,1]) \prod_{i=2}^{k} \det(A_i)$, und damit

$$\det(A) = \left(\sum_{e=1}^{n_1} (-1)^{e+1} (A_1)_{e1} \det(A_1[e,1]) \right) \cdot \prod_{i=2}^{k} \det(A_i) = \det(A_1) \cdot \prod_{i=2}^{k} \det(A_i),$$

wie gewünscht. □

Definition 5.3.37 (Kofaktor, Kofaktormatrix, adjunkte Matrix)**.** Seien $n \in \mathbb{N}$ und A eine $n \times n$-Matrix über K.

- Der (i,j)-*Kofaktor* von A ist $(-1)^{i+j} \det(A[i,j]) \in K$.
- Die $n \times n$-Matrix
$$\operatorname{cof}(A) = \left((-1)^{i+j} \det(A[i,j]) \right)_{i,j}$$
heißt die *Kofaktormatrix* von A.
- Die *adjunkte Matrix* zu A ist die $n \times n$-Matrix $\operatorname{adj}(A) = \operatorname{cof}(A)^{\intercal}$.

Beispiel 5.3.38. Es gilt

$$\operatorname{adj}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Korollar 5.3.39. *Sei $n \in \mathbb{N}$ und sei $A \in M_n(K)$. Dann gilt*

$$A \cdot \operatorname{adj}(A) = \operatorname{adj}(A) \cdot A = \det(A) \cdot I_n.$$

Beweis. Wir berechnen zunächst $A \cdot \operatorname{adj}(A)$. Es gilt

$$(A \cdot \operatorname{adj}(A))_{ij} = \sum_{k=1}^{n} A_{ik} \operatorname{adj}(A)_{kj} = \sum_{k=1}^{n} (-1)^{k+j} A_{ik} \det(A[j,k]).$$

Wenn $i = j$ ist diese Summe gleich $\det(A)$ nach dem Laplaceschen Entwicklungssatz (Entwicklung nach der i-ten Zeile). Es bleibt also zu zeigen, dass diese Summe null ist, wenn $i \neq j$. Sei A_j die Matrix, die aus A entsteht, wenn man die j-te Zeile durch A_{i*} ersetzt. Entwicklung nach der j-ten Zeile zeigt, dass die obige Summe gleich $\det(A_j)$ ist. Aber $\det(A_j) = \Delta((A_j)_{1*}, \ldots, (A_j)_{n*}) = 0$, da die i-te und j-te Zeilen von A_j gleich sind und Δ alternierend ist. Die Berechnung von $\operatorname{adj}(A) \cdot A$ erfolgt auf ähnliche Weise durch Spaltenentwicklung. □

Korollar 5.3.40 (Determinante und Invertierbarkeit)**.** *Sei $n \in \mathbb{N}$. Eine Matrix $A \in M_n(K)$ ist genau dann invertierbar, wenn $\det(A) \in K^\times$. In diesem Fall gilt*

$$A^{-1} = \det(A)^{-1} \cdot \operatorname{adj}(A).$$

Beweis. Wenn A invertierbar ist, dann ist $\det(A) \in K^\times$ nach Proposition 5.3.26(iii). Sei umgekehrt $\det(A) \in K^\times$. Dann ist $\det(A)^{-1} \cdot \operatorname{adj}(A)$ die inverse Matrix zu A nach Korollar 5.3.39. □

Beispiel 5.3.41. Die 2×2-Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ist genau dann invertierbar, wenn $ad - bc \neq 0$, in welchem Fall gilt

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(siehe Beispiele 5.3.25 und 5.3.38).

Beispiel 5.3.42. Die Matrix $A \in M_3(K)$ aus Beispiel 5.3.35 ist genau dann invertierbar, wenn $\operatorname{char}(K) \notin \{2, 47\}$. Denn 2 und 47 sind die Primzahlen, die $\det(A) = -94$ teilen.

Korollar 5.3.43 (Cramersche Regel)**.** *Sei $n \in \mathbb{N}$ und sei (A, b) ein lineares Gleichungssystem über K mit n Gleichungen und n Unbekannten. Sei $A_j[b]$ die Matrix, die aus A entsteht, wenn man die j-te Spalte durch b ersetzt. Ist $\det(A) \in K^\times$, so hat (A, b) genau eine Lösung $x \in K^n$, für die gilt*

$$x_j = \frac{\det(A_j[b])}{\det(A)}.$$

Beweis. Nach Korollar 5.3.40 ist A invertierbar, und die einzige Lösung von (A, b) ist

$$x = A^{-1} \cdot b = \det(A)^{-1} \cdot \operatorname{adj}(A) \cdot b.$$

Die j-te Koordinate ist also

$$x_j = \det(A)^{-1} \cdot \sum_{i=1}^{n} \operatorname{adj}(A)_{ji} b_i = \det(A)^{-1} \cdot \sum_{i=1}^{n} (-1)^{i+j} b_i \det(A[i,j]),$$

und die Summe ist genau die Entwicklung von $\det(A_j[b])$ nach der j-ten Spalte. □

Bemerkung 5.3.44. Wie schon gesagt, die Formeln aus Korollaren 5.3.40 und 5.3.43 sind nur von theoretischer Bedeutung. Zum Beispiel impliziert Korollar 5.3.40, dass die Koeffizienten von A^{-1} als Polynome in den Koeffizienten von A geschrieben werden können, was in der algebraischen Geometrie wichtig ist. Bei großen Matrizen ist aber das Gaußsche Eliminationsverfahren viel schneller, um A^{-1} zu berechnen oder das lineare Gleichungssystem (A, b) zu lösen, und es ist nicht sinnvoll, diese Korollare zu verwenden.

5.3.4 Die Determinante eines Endomorphismus

In diesem Abschnitt definieren wir die Determinante $\det(f)$ eines Endomorphismus $f\colon V \to V$ eines endlich-dimensionalen K-Vektorraums V. Mithilfe der Determinante von Matrizen kann man einfach $\det(f)$ als $\det([f]_B^B)$ definieren, wobei B eine beliebige Basis von V ist. Dabei muss man nachprüfen, dass $\det([f]_B^B)$ unabhängig von B ist, was aus der Basiswechselformel (Proposition 4.2.43) und der Multiplikativität der Determinante (Proposition 5.3.26) folgt. Im Folgenden geben wir aber eine begrifflichere Definition von $\det(f)$, die keine Wahl einer Basis von V erfordert, und danach beweisen wir, dass $\det(f) = \det([f]_B^B)$ (siehe Proposition 5.3.49).

Lemma 5.3.45. *Sei L ein K-Vektorraum der Dimension 1. Dann ist die Abbildung*

$$\begin{aligned} K &\to \operatorname{End}_K(L), \\ \lambda &\mapsto (v \mapsto \lambda \cdot v), \end{aligned}$$

ein Isomorphismus.

Beweis. Die Injektivität folgt aus Proposition 3.2.6(iv). Aus $L \cong K$ folgt $\operatorname{End}_K(L) \cong \operatorname{End}_K(K) \cong K$ (siehe Beispiel 4.1.45), und damit $\dim_K \operatorname{End}_K(L) = 1$. Die Sujektivität folgt dann aus Korollar 4.1.39. □

Zur Erinnerung: Ist V ein endlich-dimensionaler K-Vektorraum, so ist die Menge $\operatorname{Det}(V)$ aller Determinantenfunktionen auf V ein K-Vektorraum der Dimension 1 (Satz 5.3.21). Haben außerdem V und W dieselbe endliche Dimension, so induziert jede lineare Abbildung $f\colon V \to W$ eine lineare Abbildung $f^*\colon \operatorname{Det}(W) \to \operatorname{Det}(V)$ (Lemma 5.3.20).

Definition 5.3.46 (Determinante eines Endomorphismus)**.** Sei V ein endlich-dimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus. Die *Determinante* von f ist der Skalar $\det(f) \in K$, so dass die von f induzierte Abbildung $f^*\colon \operatorname{Det}(V) \to \operatorname{Det}(V)$ gleich der Skalarmultiplikation mit $\det(f)$ ist (siehe Lemma 5.3.45), d.h.: Es gilt $f^*(\delta) = \det(f) \cdot \delta$ für alle $\delta \in \operatorname{Det}(V)$.

Bemerkung 5.3.47. Sei $K = \mathbb{R}$ und $f \in \operatorname{End}_{\mathbb{R}}(\mathbb{R}^n)$. Nach der Definition von $\det(f)$ und der Bemerkung 5.3.22 skaliert f das Volumen eines Parallelotops in $\mathbb{R}^n$ um den Faktor $|\det(f)|$. Die Determinante von f misst also die durch f bewirkte Volumenveränderung mit einem geeigneten Vorzeichen.

Proposition 5.3.48 (Eigenschaften der Determinante)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und seien $f, g\colon V \to V$ zwei Endomorphismen. Dann gilt:*

(i) $\det(\mathrm{id}_V) = 1$.

(ii) $\det(g \circ f) = \det(g) \cdot \det(f)$.

(iii) *Ist f ein Automorphismus, so gilt* $\det(f^{-1}) = \det(f)^{-1}$.

(iv) *Für die duale Abbildung $f^*\colon V^* \to V^*$ gilt* $\det(f^*) = \det(f)$.

Beweis. Jede Aussage folgt aus Proposition 5.3.49(ii) unten und der entsprechenden Aussage über die Determinante von Matrizen (Proposition 5.3.26). Wir geben aber zusätzlich matrixfreie Beweise.

Zu (i). Dies folgt daraus, dass $\mathrm{id}_V^* \colon \mathrm{Det}(V) \to \mathrm{Det}(V)$ die Identität ist.

Zu (ii). Sei $\delta \in \mathrm{Det}(V)$. Dann gilt

$$(g \circ f)^*(\delta) = f^*(g^*(\delta)) = f^*(\det(g) \cdot \delta) = \det(g) \cdot f^*(\delta) = \det(g) \cdot \det(f) \cdot \delta.$$

Zu (iii). Dies folgt unmittelbar aus (i) und (ii).

Zu (iv). In diesem Beweis schreiben wir $\mathrm{Det}(f)$ für die von f induzierte lineare Abbildung $\mathrm{Det}(V) \to \mathrm{Det}(V)$, um sie nicht mit der dualen Abbildung f^* zu verwechseln. Sei n die Dimension von V. Wir definieren die Abbildung $\Theta \colon (V^*)^n \to \mathrm{Det}(V)$ durch

$$\Theta(\alpha_1, \dots, \alpha_n)(v_1, \dots, v_n) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \alpha_{\sigma(1)}(v_1) \dots \alpha_{\sigma(n)}(v_n).$$

Man kann leicht nachprüfen, dass $\Theta(\alpha_1, \dots, \alpha_n)$ eine Determinantenfunktion auf V ist, und dass Θ selbst n-linear und alternierend ist (vgl. den Beweis der Proposition 5.3.17). Die Abbildung Θ ist außerdem surjektiv: Ist $B = (b_1, \dots, b_n)$ eine Basis von V, so ist der Vektorraum $\mathrm{Det}(V)$ von Δ_B erzeugt (Satz 5.3.21), und es gilt $\Delta_B = \Theta(b_1^*, \dots, b_n^*)$. Man erhält daraus eine injektive lineare Abbildung

$$\begin{aligned} \vartheta \colon \mathrm{Det}(V)^* &\to \mathrm{Det}(V^*), \\ \varepsilon &\mapsto \varepsilon \circ \Theta. \end{aligned}$$

Da beide $\mathrm{Det}(V)^*$ und $\mathrm{Det}(V^*)$ die Dimension 1 haben, ist ϑ sogar ein Isomorphismus. Es gilt zudem $\mathrm{Det}(f^*) \circ \vartheta = \vartheta \circ \mathrm{Det}(f)^*$. Da die Abbildung $\mathrm{Det}(f)$ durch Skalarmultiplikation mit $\det(f)$ gegeben ist, gilt das Gleiche für $\mathrm{Det}(f)^*$ und damit für $\mathrm{Det}(f^*)$, d.h., $\det(f^*) = \det(f)$. □

Proposition 5.3.49.

(i) *Seien $n \in \mathbb{N}$ und $A \in M_n(K)$. Dann gilt* $\det(L_A) = \det(A)$.

(ii) *Sei f ein Endomorphismus eines endlich-dimensionalen K-Vektorraum V. Für alle Basen B von V gilt* $\det(f) = \det([f]_B^B)$.

Beweis. Die erste Aussage ist der Sonderfall der zweiten mit $V = K^n$ und $B = (e_1, \dots, e_n)$. Sei $\delta \in \mathrm{Det}(V)$. Nach Satz 5.3.21 ist $\delta = \lambda \cdot \Delta_B$ mit einem $\lambda \in K$. Sei $B = (b_1, \dots, b_n)$. Es gilt

$$\begin{aligned} f^*(\delta)(b_1, \dots, b_n) &= \lambda \cdot \Delta_B(f(b_1), \dots, f(b_n)) = \lambda \cdot \Delta([f(b_1)]_B, \dots, [f(b_n)]_B) \\ &= \lambda \cdot \Delta([f]_B^B \cdot [b_1]_B, \dots, [f]_B^B \cdot [b_n]_B) = \lambda \cdot \Delta([f]_B^B \cdot e_1, \dots, [f]_B^B \cdot e_n) \\ &= \lambda \cdot \det([f]_B^B) = \delta(b_1, \dots, b_n) \cdot \det([f]_B^B). \end{aligned}$$

Die Determinantenfunktionen $f^*(\delta)$ und $\det([f]_B^B) \cdot \delta$ stimmen also auf $(b_1, \dots, b_n)$ überein. Aus dem Satz 5.3.21 folgt, dass $f^*(\delta) = \det([f]_B^B) \cdot \delta$, und damit dass $\det(f) = \det([f]_B^B)$. □

Proposition 5.3.50. *Sei V ein endlich-dimensionaler K-Vektorraum. Ein Endomorphismus $f \colon V \to V$ ist genau dann ein Automorphismus, wenn* $\det(f) \in K^\times$.

Beweis. Sei B eine Basis von V. Dann ist f genau dann ein Automorphismus, wenn die Darstellungsmatrix $[f]_B^B$ invertierbar ist. Die Aussage folgt jetzt aus Proposition 5.3.49(ii) und Korollar 5.3.40. □

Kapitel 6

Eigenwerte und Diagonalisierbarkeit

In diesem Kapitel beschäftigen wir uns mit *Endomorphismen* von Vektorräumen. Das Endziel ist es, die Frage 4.2.45(ii) vollständig zu beantworten, d.h., Endomorphismen von endlichdimensionalen Vektorräumen bis auf Isomorphie zu klassifizieren. Das werden wir aber nur in der Vorlesung *Lineare Algebra II* schaffen. In diesem Kapitel führen wir die wichtigen Begriffe von *Eigenwert* und *Eigenvektoren* ein, und erhalten wir eine Klassifikation von sogenannten *diagonalisierbaren* Endomorphismen, die bis auf Isomorphie durch ihre Eigenwerte bestimmt sind. Außerdem entwickeln wir mehrere Hilfsmittel zum besseren Verständnis von Endomorphismen, wie zum Beispiel das *charakteristische Polynom*. Die meisten Begriffe in diesem Kapitel sind tatsächlich auch sinnvoll bei unendlich-dimensionalen Vektorräumen, und es gibt wie immer in diesem Fall interessante Beispiele aus der Analysis. Deshalb berücksichtigen wir in unserer Untersuchung so weit wie möglich beliebige Vektorräume.

Ein beliebiger Grundkörper K wird immer noch festgelegt.

6.1 Präliminarien zu Endomorphismen

6.1.1 Direkte Summen von Vektorräumen

Die direkte Summe zweier K-Vektorräume haben wir bereits definiert (Definition 3.3.45). Als Nächstes wollen wir diese Konstruktion auf mehr als zwei Vektorräume verallgemeinern. Bei unendlich vielen Vektorräumen muss man aber zwischen dem Produkt und der direkten Summe unterscheiden, die verschiedene universelle Eigenschaften haben.

Definition 6.1.1 (Produkt und direkte Summe einer Familie von Vektorräumen)**.** Sei I eine beliebige Menge und $(V_i)_{i\in I}$ eine Familie von K-Vektorräumen.

- Das *Produkt* von $(V_i)_{i\in I}$ ist der K-Vektorraum

$$\prod_{i\in I} V_i = \{(v_i)_{i\in I} \mid v_i \in V_i \text{ für alle } i \in I\}$$

  mit der punktweisen Addition bzw. Skalarmultiplikation. Ist $e \in I$, so bezeichnen wir mit $\pi_e \colon \prod_{i\in I} V_i \to V_e$ die lineare Abbildung mit $\pi_e((v_i)_{i\in I}) = v_e$.

- Die *direkte Summe* von $(V_i)_{i\in I}$ ist der Untervektorraum

$$\bigoplus_{i\in I} V_i \subset \prod_{i\in I} V_i$$

bestehend aus allen Familien $(v_i)_{i\in I}$, die null sind außerhalb einer endlichen Teilmenge von I. Ist $e \in I$, so bezeichnen wir mit $\iota_e \colon V_e \to \bigoplus_{i\in I} V_i$ die lineare Abbildung mit

$$\iota_e(v)_i = \begin{cases} v, & \text{falls } i = e, \\ 0, & \text{andernfalls.} \end{cases}$$

Bemerkung 6.1.2. Es ist $K^I = \prod_{i\in I} K$ und $K^{(I)} = \bigoplus_{i\in I} K$.

Proposition 6.1.3 (universelle Eigenschaften des Produkts und der direkten Summe)**.** *Sei $(V_i)_{i\in I}$ eine Familie von K-Vektorräumen.*

(i) *Zu jedem K-Vektorraum W und jeder Familie $(f_i \colon W \to V_i)_{i\in I}$ von linearen Abbildungen gibt es* genau eine *lineare Abbildung $f \colon W \to \prod_{i\in I} V_i$, so dass $\pi_i \circ f = f_i$ für alle $i \in I$.*

(ii) *Zu jedem K-Vektorraum W und jeder Familie $(f_i \colon V_i \to W)_{i\in I}$ von linearen Abbildungen gibt es* genau eine *lineare Abbildung $f \colon \bigoplus_{i\in I} V_i \to W$, so dass $f \circ \iota_i = f_i$ für alle $i \in I$.*

Beweis. Zu (i). Man definiert f durch $f(w) = (f_i(w))_{i\in I}$, so dass nach Definition gilt $\pi_i \circ f = f_i$. Die Abbildung f ist linear, da der Vektorraumstruktur auf $\prod_{i\in I} V_i$ punktweise definiert ist. Die Eindeutigkeit ist klar, da die i-te Koordinate von $f(w)$ gleich $\pi_i(f(w)) = f_i(w)$ sein muss.

Zu (ii). Man definiert f durch $f((v_i)_{i\in I}) = \sum_{i\in I} f_i(v_i)$; dies ist sinnvoll (und offensichtlich linear), da nur endlich viele Summanden nicht null sind. Nach Definition der direkten Summe ist jedes Element von $\bigoplus_{i\in I} V_i$ eine Linearkombination von Elemente der Gestalt $\iota_i(v_i)$ mit $i \in I$ und $v_i \in V_i$. Deswegen gibt es höchstens ein f mit den geforderten Eigenschaften. □

Bemerkung 6.1.4. Das Produkt und die Summe einer Mengenfamilie (Definition 1.2.15) haben analoge universelle Eigenschaften bzgl. beliebiger statt linearer Abbildungen.

Bemerkung 6.1.5 (Dimension einer direkten Summe)**.** Ist $(V_i)_{i\in I}$ eine endliche Familie von endlich-dimensionalen K-Vektorräumen, so gilt

$$\dim_K \left(\bigoplus_{i\in I} V_i \right) = \sum_{i\in I} \dim_K V_i.$$

Denn sind B_i Basen von V_i, so ist die Zusammensetzung der Familien $\iota_i(B_i)$ eine Basis von $\bigoplus_{i\in I} V_i$.

Die folgende Definition ist eine Verallgemeinerung von Definition 3.3.36:

Definition 6.1.6 (Summe einer Familie von Untervektorräumen)**.** Sei V ein K-Vektorraum und $(U_i)_{i\in I}$ eine Familie von Untervektorräumen. Die *Summe* der Familie $(U_i)_{i\in I}$ ist der Untervektorraum

$$\sum_{i\in I} U_i := \operatorname{Span}_K \left(\bigcup_{i\in I} U_i \right) \subset V.$$

Ist $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V, so gibt es nach der universellen Eigenschaft der direkten Summe eine kanonische surjektive lineare Abbildung

$$\bigoplus_{i\in I} U_i \to \sum_{i\in I} U_i.$$

Wenn diese Abbildung bijektiv ist, sagt man auch, dass die Summe $\sum_{i\in I} U_i$ *direkt* ist, und man schreibt dann oft $\bigoplus_{i\in I} U_i$ für diesen Untervektorraum von V.

Proposition 6.1.7. *Sei V ein K-Vektorraum und $(U_i)_{i\in I}$ eine Familie von Untervektorräumen. Dann sind die folgenden Aussagen äquivalent:*

(i) *Die kanonische Abbildung $\bigoplus_{i\in I} U_i \to \sum_{i\in I} U_i$ ist ein Isomorphismus.*

(ii) *Für jede endliche Teilmenge $J \subset I$ ist jede Familie $(u_i)_{i\in J}$ mit $u_i \in U_i \setminus \{0\}$ linear unabhängig.*

Beweis. Zu (i) $\Rightarrow$ (ii). Sei $\sum_{i\in J} \lambda_i u_i = 0$. Die Linearkombination $\sum_{i\in J} \lambda_i u_i$ ist das Bild von $(\lambda_i u_i)_{i\in J}$ unter der kanonischen Abbildung

$$\bigoplus_{i\in J} U_i \to \bigoplus_{i\in I} U_i \to \sum_{i\in I} U_i,$$

die injektiv ist. Daraus folgt, dass $\lambda_i u_i = 0$ und damit $\lambda_i = 0$ für alle $i \in J$.

Zu (ii) $\Rightarrow$ (i). Die kanonische Abbildung ist surjektiv nach Definition der Summe. Sei $(u_i)_{i\in I}$ ein Element ihres Kerns, d.h., $\sum_{i\in I} u_i = 0$. Sei $J \subset I$ die endliche Teilmenge aller Indizes i mit $u_i \neq 0$. Nach (ii) ist die Familie $(u_i)_{i\in J}$ linear unabhängig. Daraus folgt, dass $J = \varnothing$, sonst würde die Gleichung $\sum_{i\in J} u_i = 0$ im Widerspruch zu der linearen Unabhängigkeit stehen. □

6.1.2 Invariante Untervektorräume

Definition 6.1.8 (invarianter Untervektorraum)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Ein Untervektorraum $U \subset V$ heißt *f-invariant*, wenn $f(U) \subset U$.

Ist $U \subset V$ f-invariant, so schränkt sich f zu einem Endomorphismus $f_U \in \mathrm{End}_K(U)$. Nach der universellen Eigenschaft des Quotientenvektorraums (Proposition 4.1.33) induziert auch f einen Endomorphismus $\bar{f} \in \mathrm{End}_K(V/U)$ mit $\bar{f}(v + U) = f(v) + U$.

Beispiel 6.1.9. Sei $f\colon V \to V$ ein Endomorphismus. Dann sind $\{0\}$, V, $\ker f$ und $\operatorname{im} f$ f-invariante Untervektoräume von V.

Proposition 6.1.10. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $(U_i)_{i\in I}$ eine Familie von f-invarianten Untervektorräumen.*

(i) *Der Durchschnitt $\bigcap_{i\in I} U_i$ ist f-invariant.*

(ii) *Die Summe $\sum_{i\in I} U_i$ ist f-invariant.*

Beweis. Dies folgt aus den mengentheoretischen Formeln

$$f\left(\bigcap_{i\in I} U_i\right) \subset \bigcap_{i\in I} f(U_i) \quad \text{und} \quad f\left(\bigcup_{i\in I} U_i\right) = \bigcup_{i\in I} f(U_i)$$

und der Proposition 4.1.20(i). □

Weitere Beispiele erhalten wir durch die Potenzen von f:

Notation 6.1.11 (Potenzen eines Endomorphismus)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $k \in \mathbb{N}$. Man definiert die k-te Potenz f^k rekursiv wie folgt:

$$f^0 = \mathrm{id}_V, \quad f^{k+1} = f \circ f^k.$$

Proposition 6.1.12. *Sei V ein K-Vektorraum, $f \in \mathrm{End}_K(V)$ ein Endomorphismus und $g \in \mathrm{End}_K(V)$ eine Linearkombination der Familie $(f^k)_{k\in\mathbb{N}}$. Dann:*

(i) $\ker g$ *und* $\operatorname{im} g$ *sind f-invariant.*

(ii) *Ist $U \subset V$ ein f-invarianter Untervektorraum, so ist U auch g-invariant.*

Beweis. *Zu* (i). Sei $g = \sum_{k=0}^{n} \lambda_k f^k$. Die Linearität von f impliziert, dass $g \circ f = f \circ g$, denn:

$$g(f(v)) = \sum_{k=0}^{n} \lambda_k f^{k+1}(v) = f\left(\sum_{k=0}^{n} \lambda_k f^k(v)\right) = f(g(v)).$$

Sei $v \in \ker g$. Dann ist $g(f(v)) = f(g(v)) = f(0) = 0$, und damit ist $f(v) \in \ker g$. Sei nun $v \in \operatorname{im} g$, d.h., $v = g(w)$ mit einem $w \in V$. Dann ist $f(v) = f(g(w)) = g(f(w))$, und damit ist $f(w) \in \operatorname{im} g$.

Zu (ii). Durch Induktion über k schließen wir unmittelbar, dass $f^k(U) \subset U$ für alle $k \in \mathbb{N}$. Da U ein Untervektorraum ist, folgern wir, dass $g(U) \subset U$. □

6.1.3 Isomorphie von Endomorphismen

Definition 6.1.13 (Isomorphie von Endomorphismen)**.** Seien V, W Vektorräume über K und seien $f \in \operatorname{End}_K(V)$ und $g \in \operatorname{End}_K(W)$ Endomorphismen. Man sagt, dass die Paare (V, f) und (W, g) *isomorph* sind, und man schreibt $(V, f) \cong (W, g)$, wenn ein Isomorphismus $\varphi\colon V \xrightarrow{\sim} W$ existiert, so dass $\varphi \circ f = g \circ \varphi$:

$$\begin{array}{ccc} V & \xrightarrow[\sim]{\varphi} & W \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ V & \xrightarrow[\sim]{\varphi} & W. \end{array}$$

Bemerkung 6.1.14. Man kann leicht nachprüfen, dass Isomorphie eine Äquivalenzrelation zwischen Paaren (V, f) ist (vgl. Bemerkung 4.1.16).

Beispiel 6.1.15.

(i) Jedes (V, f) ist isomorph zu einem Paar der Gestalt $(K^{(I)}, g)$: Man wählt einen Isomorphismus $\varphi\colon V \xrightarrow{\sim} K^{(I)}$ und setzt $g = \varphi \circ f \circ \varphi^{-1}$.

(ii) Sei V ein n-dimensionaler K-Vektorraum mit einer Basis B. Durch den Isomorphismus $\varphi_B^{-1}\colon V \xrightarrow{\sim} K^n$ ist jedes Paar (V, f) zu $(K^n, L_{[f]_B^B})$ isomorph.

Bemerkung 6.1.16. Die Determinante von Endomorphismen (Definition 5.3.46) ist eine *Isomorphie-Invariante* im folgenden Sinne: Ist V endlich-dimensional und sind (V, f) und (W, g) isomorph, so ist $\det(f) = \det(g)$. Denn sei $\varphi\colon V \xrightarrow{\sim} W$ ein Isomorphismus mit $\varphi \circ f = g \circ \varphi$ und sei $\delta \in \operatorname{Det}(W)$ eine Determinantenfunktion. Dann gilt

$$\begin{aligned} g^*(\delta) &= (\varphi \circ f \circ \varphi^{-1})^*(\delta) \\ &= (\varphi^{-1})^*(f^*(\varphi^*(\delta))) \\ &= (\varphi^{-1})^*(\det(f) \cdot \varphi^*(\delta)) \\ &= \det(f) \cdot (\varphi^{-1})^*(\varphi^*(\delta)) \\ &= \det(f) \cdot (\varphi \circ \varphi^{-1})^*(\delta) \\ &= \det(f) \cdot \delta, \end{aligned}$$

und damit $\det(g) = \det(f)$.

Definition 6.1.17 (Ähnlichkeit von Matrizen)**.** Seien $n \in \mathbb{N}$ und $A, B \in M_n(K)$. Man sagt, dass A *ähnlich* oder *konjugiert* zu B ist, wenn eine invertierbare Matrix $S \in \mathrm{GL}_n(K)$ existiert, so dass $S^{-1} \cdot A \cdot S = B$.

Proposition 6.1.18. *Ähnlichkeit ist eine Äquivalenzrelation auf* $M_n(K)$.

Beweis. Sie ist reflexiv, da $A = I_n^{-1} \cdot A \cdot I_n$. Sie ist symmetrisch, denn:

$$B = S^{-1} \cdot A \cdot S \quad \Longrightarrow \quad A = (S^{-1})^{-1} \cdot B \cdot S^{-1}.$$

Zur Transitivät, seien $B = S^{-1} \cdot A \cdot S$ und $C = T^{-1} \cdot B \cdot T$. Dann gilt:

$$(S \cdot T)^{-1} \cdot A \cdot (S \cdot T) = T^{-1} \cdot (S^{-1} \cdot A \cdot S) \cdot T = T^{-1} \cdot B \cdot T = C. \qquad \square$$

Proposition 6.1.19. *Sei $n \in \mathbb{N}$.*

(i) *Seien $A, B \in M_n(K)$. Dann sind A und B genau dann ähnlich, wenn (K^n, L_A) und (K^n, L_B) isomorph sind.*

(ii) *Seien V und W n-dimensionale K-Vektorräume mit Endomorphismen $f \in \mathrm{End}_K(V)$ und $g \in \mathrm{End}_K(W)$ und mit Basen B und C. Dann sind (V, f) und (W, g) genau dann isomorph, wenn $[f]_B^B$ und $[g]_C^C$ ähnlich sind.*

Beweis. Die erste Aussage ist der Sonderfall der zweiten, mit $V = W = K^n$ und $B = C$ der Standardbasis. Sei $\varphi\colon V \to W$ ein Isomorphismus mit $\varphi \circ f \circ \varphi^{-1} = g$. Für $S = [\varphi^{-1}]_B^C$ gilt dann $S^{-1}[f]_B^B S = [g]_C^C$ nach Proposition 4.2.39. Sei umgekehrt $S \in \mathrm{GL}_n(K)$ mit $S^{-1}[f]_B^B S = [g]_C^C$. Sei $\varphi\colon V \to W$ der Isomorphismus mit $[\varphi]_C^B = S^{-1}$. Nach Proposition 4.2.39 gilt dann $[\varphi \circ f \circ \varphi^{-1}]_C^C = S^{-1}[f]_B^B S = [g]_C^C$, und daher $\varphi \circ f \circ \varphi^{-1} = g$. $\square$

Man kann die letzte Proposition wie folgt zusammenfassen: Es gibt eine bijektive Abbildung

$$\begin{aligned} M_n(K)/\text{Ähnlichkeit} &\xrightarrow{\sim} \{(V, f) \mid \dim_K V = n \text{ und } f \in \mathrm{End}_K(V)\}/\text{Isomorphie}, \\ [A] &\mapsto [(K^n, L_A)], \end{aligned}$$

deren Umkehrabbildung die Äquivalenzklasse von (V, f) auf die von $[f]_B^B$ abbildet, wobei B eine beliebige Basis von V ist. In der Praxis bedeutet das folgendes: Wenn wir eine Abbildung $\Phi\colon M_n(K) \to X$ definiert haben, so dass $\Phi(A) = \Phi(B)$ wann immer A und B ähnlich sind, dann erhalten wir aus jedem Paar (V, f) ein wohldefiniertes Element $\Phi(f) \in X$, so dass $\Phi(f) = \Phi([f]_B^B)$ für jede Basis B von V. Außerdem gilt dann $\Phi(f) = \Phi(g)$, wenn die Paare (V, f) und (W, g) isomorph sind, d.h., Φ ist automatisch eine *Isomorphie-Invariante*. Als Beispiel dieser Methode definieren wir jetzt die *Spur* eines Endomorphismus.

Definition 6.1.20 (Spur einer Matrix)**.** Seien $n \in \mathbb{N}$ und $A = (a_{ij})_{i,j}$ eine $n \times n$-Matrix über K. Die *Spur* von A ist

$$\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii} \in K.$$

Proposition 6.1.21 (zyklische Invarianz der Spur)**.** *Seien $m, n \in \mathbb{N}$, sei $A \in M_{m \times n}(K)$ und sei $B \in M_{n \times m}(K)$. Dann gilt*

$$\mathrm{tr}(A \cdot B) = \mathrm{tr}(B \cdot A).$$

Beweis. $\mathrm{tr}(A \cdot B) = \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} = \mathrm{tr}(B \cdot A).$ $\square$

Bemerkung 6.1.22. Seien $A_1, \ldots, A_k \in M_n(K)$ und sei $\sigma \in S_k$ ein Zyklus der Länge k. Die zyklische Invarianz der Spur impliziert, dass

$$\mathrm{tr}(A_1 \cdot \ldots \cdot A_k) = \mathrm{tr}(A_{\sigma(1)} \cdot \ldots \cdot A_{\sigma(k)}).$$

Im Gegensatz zu der Determinante, gilt dies aber im Allgemeinen nicht für eine beliebige Permutation σ.

Korollar 6.1.23. *Sei $n \in \mathbb{N}$ und seien $A, B \in M_n(K)$. Sind A und B ähnlich, so gilt* $\operatorname{tr}(A) = \operatorname{tr}(B)$.

Beweis. Sei $S^{-1} \cdot A \cdot S = B$. Nach Proposition 6.1.21 gilt nun

$$\operatorname{tr}(B) = \operatorname{tr}(S^{-1} \cdot A \cdot S) = \operatorname{tr}(A \cdot S \cdot S^{-1}) = \operatorname{tr}(A). \qquad \square$$

Definition 6.1.24 (Spur eines Endomorphismus)**.** Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}_K(V)$. Die *Spur* von f ist

$$\operatorname{tr}(f) := \operatorname{tr}([f]_B^B),$$

wobei B eine Basis von V ist. Nach Korollar 6.1.23 ist $\operatorname{tr}(f)$ unabhängig von der Wahl der Basis B.

Bemerkung 6.1.25. Es ist auch möglich, eine begrifflichere matrixfreie Definition der Spur $\operatorname{tr}(f)$ zu formulieren, wie bei der Determinante (siehe Abschnitt 5.3.4). Dazu braucht man aber den Begriff des *Tensorprodukts* von Vektorräumen, den wir noch nicht besprochen haben.

6.2 Eigenvektoren und Eigenwerte

Definition 6.2.1 (Eigenraum, Eigenvektor, Eigenwert)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $\lambda \in K$.

- Der *Eigenraum* zu λ von f ist der Untervektorraum

$$\operatorname{Eig}_\lambda(f) := \{v \in V \mid f(v) = \lambda \cdot v\} = \ker(\lambda \cdot \mathrm{id}_V - f) \subset V.$$

- Ein *Eigenvektor* zu λ von f ist ein Element von $\operatorname{Eig}_\lambda(f) \setminus \{0\}$, d.h., ein Vektor $v \neq 0$, so dass $f(v) = \lambda \cdot v$.

- λ heißt *Eigenwert* von f, wenn $\operatorname{Eig}_\lambda(f) \neq \{0\}$, d.h., wenn ein Eigenvektor zu λ existiert.

Ist $n \in \mathbb{N}$ und ist A eine $n \times n$-Matrix über K, so bezeichnen wir als *Eigenräume*, *Eigenvektoren* und *Eigenwerte* von A die Eigenräume, Eigenvektoren und Eigenwerte von $L_A\colon K^n \to K^n$.

Bemerkung 6.2.2. Es gilt $\operatorname{Eig}_0(f) = \ker f$.

Bemerkung 6.2.3. Seien $f \in \operatorname{End}_K(V)$ und $g \in \operatorname{End}_K(W)$ Endomorphismen, und sei $\varphi\colon V \xrightarrow{\sim} W$ ein Isomorphismus mit $\varphi \circ f = g \circ \varphi$. Für alle $\lambda \in K$ ist dann $\varphi(\operatorname{Eig}_\lambda(f)) = \operatorname{Eig}_\lambda(g)$, denn:

$$f(v) = \lambda \cdot v \iff \varphi(f(v)) = \varphi(\lambda \cdot v) \iff g(\varphi(v)) = \lambda \cdot \varphi(v).$$

Insbesondere haben f und g dieselben Eigenwerte.

Beispiel 6.2.4. Sei V ein K-Vektorraum.

(i) Für die Identität id_V gilt

$$\operatorname{Eig}_\lambda(\mathrm{id}_V) = \begin{cases} V, & \text{falls } \lambda = 1, \\ \{0\}, & \text{falls } \lambda \neq 1. \end{cases}$$

Falls $V \neq \{0\}$ ist also $1 \in K$ der einzige Eigenwert von id_V, und alle Vektoren $v \neq 0$ sind Eigenvektoren dazu.

(ii) Allgemeiner, ist $\mu \in K$ und $V \neq \{0\}$, so ist μ der einzige Eigenwert von $\mu \cdot \mathrm{id}_V$, und alle Vektoren $v \neq 0$ sind Eigenvektoren dazu.

(iii) Sei $t\colon V \oplus V \to V \oplus V$ die lineare Abbildung $t(v,w) = (w,v)$. Es gilt

$$\begin{aligned}\mathrm{Eig}_\lambda(t) &= \{(v,w) \in V \oplus V \mid w = \lambda \cdot v \text{ und } v = \lambda \cdot w\} \\ &= \{(v, \lambda \cdot v) \mid v \in V \text{ und } (\lambda^2 - 1) \cdot v = 0\}.\end{aligned}$$

Nach Proposition 3.2.6(iv) ist $(\lambda^2 - 1) \cdot v = 0$ nur möglich, wenn $v = 0$ oder $\lambda^2 - 1 = 0$. Daher gilt

$$\mathrm{Eig}_\lambda(t) = \{(0,0)\} \cup \{(v, \lambda \cdot v) \mid v \in V \setminus \{0\} \text{ und } \lambda^2 - 1 = 0\}.$$

Falls $V \neq \{0\}$ sind also die Eigenwerte von t alle Skalare λ mit $\lambda^2 - 1 = 0$. Aus der Gleichung $\lambda^2 - 1 = (\lambda - 1)(\lambda + 1)$ und der Nullteilerfreiheit von K folgt, dass 1 und -1 die einzigen Eigenwerte von t sind. Die Eigenvektoren zu 1 sind (v,v) mit $v \neq 0$ und die Eigenvektoren zu -1 sind $(v,-v)$ mit $v \neq 0$.

(iv) Sei $D = \mathrm{diag}(d_1, \ldots, d_n)$ eine $n \times n$-Diagonalmatrix über K. Dann ist der Eigenraum zu λ von D der Untervektorraum $\mathrm{Span}_K\{e_i \mid d_i = \lambda\} \subset K^n$. Insbesondere ist jeder Standardeinheitsvektor $e_i \in K^n$ ein Eigenvektor zum Eigenwert d_i von D.

Beispiel 6.2.5 (Drehungen)**.** Sei $\alpha \in \mathbb{R} \setminus \pi\mathbb{Z}$ eine reelle Zahl, die kein ganzzahliges Vielfaches von π ist, und sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die Drehung um den Winkel α um den Nullpunkt. Dann ist $\mathrm{Eig}_\lambda(f) = \{0\}$ für alle $\lambda \in K$. Denn für alle $v \neq 0$ liegen die Vektoren v und $f(v)$ auf keiner gemeinsamen Ursprungsgerade. Der Endomorphismus f hat also keine Eigenwerte und somit keine Eigenvektoren.

Beispiel 6.2.6 (Abhängigkeit vom Grundkörper)**.** Sei

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R}).$$

Die Abbildung $L_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ ist eine Drehung um den Winkel $\pi/2$, die keine Eigenwerte besitzt (Beispiel 6.2.5). Betrachtet man jedoch A als komplexe Matrix, so hat $L_A\colon \mathbb{C}^2 \to \mathbb{C}^2$ die Eigenwerte i und $-i$, mit zugehörigen Eigenvektoren $\left(\begin{smallmatrix} i \\ 1 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right)$:

$$A \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \begin{pmatrix} i \\ 1 \end{pmatrix}, \quad A \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} -i \\ 1 \end{pmatrix} = -i \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Beispiel 6.2.7 (Eigenwerte und Eigenvektoren der Differentiation)**.** Sei $C^\infty(\mathbb{R},\mathbb{R})$ der $\mathbb{R}$-Vektorraum aller beliebig oft differenzierbaren Funktionen auf $\mathbb{R}$, sei $D\colon C^\infty(\mathbb{R},\mathbb{R}) \to C^\infty(\mathbb{R},\mathbb{R})$ die Differentiationsabbildung (siehe Beispiel 4.1.31) und sei $\lambda \in \mathbb{R}$. Dann ist

$$\mathrm{Eig}_\lambda(D) = \{f \in C^\infty(\mathbb{R},\mathbb{R}) \mid f' = \lambda \cdot f\}.$$

In der Analysis wird gezeigt, dass der Eigenraum $\mathrm{Eig}_\lambda(D)$ 1-dimensional ist und von der Exponentialfunktion $x \mapsto \exp(\lambda x)$ erzeugt wird. Insbesondere sind alle reellen Zahlen Eigenwerte von D, und die Eigenvektoren zu einem λ sind die Funktionen $x \mapsto c\exp(\lambda x)$ mit $c \in \mathbb{R}^\times$.

Beispiel 6.2.8. Sei $D^2\colon C^\infty(\mathbb{R},\mathbb{R}) \to C^\infty(\mathbb{R},\mathbb{R})$ der Endomorphismus $f \mapsto f''$ und sei $\lambda \in \mathbb{R}$. Dann gilt $\dim_{\mathbb{R}} \mathrm{Eig}_\lambda(D^2) = 2$ und zwar

$$\mathrm{Eig}_\lambda(D^2) = \begin{cases} \mathrm{Span}_{\mathbb{R}}\{x \mapsto \cosh(\sqrt{\lambda}\,x), x \mapsto \sinh(\sqrt{\lambda}\,x)\}, & \text{falls } \lambda > 0, \\ \mathrm{Span}_{\mathbb{R}}\{x \mapsto 1, x \mapsto x\}, & \text{falls } \lambda = 0, \\ \mathrm{Span}_{\mathbb{R}}\{x \mapsto \cos(\sqrt{-\lambda}\,x), x \mapsto \sin(\sqrt{-\lambda}\,x)\}, & \text{falls } \lambda < 0. \end{cases}$$

Lemma 6.2.9. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V.*

(i) *Für alle $\lambda \in K$ ist der Eigenraum $\mathrm{Eig}_\lambda(f) \subset V$ f-invariant.*

(ii) *Ist $\lambda \neq \mu$, so gilt $\mathrm{Eig}_\lambda(f) \cap \mathrm{Eig}_\mu(f) = \{0\}$.*

Beweis. Die erste Aussage ist der Sonderfall der Proposition 6.1.12(i) mit $g = \lambda \cdot \mathrm{id}_V - f$. Sei $v \in \mathrm{Eig}_\lambda(f) \cap \mathrm{Eig}_\mu(f)$. Dann gilt $f(v) = \lambda \cdot v = \mu \cdot v$, und damit $(\lambda - \mu) \cdot v = 0$. Da $\lambda \neq \mu$ folgt aus Proposition 3.2.6(iv), dass $v = 0$. □

Proposition 6.2.10 (lineare Unabhängigkeit von Eigenvektoren)**.** *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Sei $\Lambda \subset K$ eine Teilmenge bestehend aus Eigenwerten von f, und zu jedem $\lambda \in E$ sei $v_\lambda \in V$ ein Eigenvektor zum Eigenwert λ. Dann ist die Familie $(v_\lambda)_{\lambda\in\Lambda}$ linear unabhängig.*

Beweis. Nach Definition der linearen Unabhängigkeit dürfen wir voraussetzen, dass Λ endlich ist. Dann verwenden wir Induktion über die Mächtigkeit von Λ. Wenn $\Lambda = \varnothing$ ist die Aussage trivial. Sei also $\lambda_0 \in \Lambda$ und sei $\sum_{\lambda\in\Lambda} \mu_\lambda \cdot v_\lambda = 0$ mit $\mu_\lambda \in K$. Dann gilt:

$$0 = (\lambda_0 \,\mathrm{id}_V - f)\left(\sum_{\lambda\in\Lambda} \mu_\lambda \cdot v_\lambda\right) = \sum_{\lambda\in\Lambda\setminus\{\lambda_0\}} \mu_\lambda \cdot (\lambda_0 \,\mathrm{id}_V - f)(v_\lambda).$$

Nach Lemma 6.2.9(i) liegt jeder Vektor $(\lambda_0 \,\mathrm{id}_V - f)(v_\lambda)$ in $\mathrm{Eig}_\lambda(f)$, und nach Lemma 6.2.9(ii) ist er nicht null, sonst wäre $v_\lambda \in \mathrm{Eig}_\lambda(f) \cap \mathrm{Eig}_{\lambda_0}(f) = \{0\}$, aber $v_\lambda \neq 0$ nach Definition von Eigenvektor. Also ist jedes $(\lambda_0 \,\mathrm{id}_V - f)(v_\lambda)$ wieder ein Eigenvektor zu λ. Aus der Induktionsvoraussetzung folgt, dass $\mu_\lambda = 0$ für alle $\lambda \in \Lambda \setminus \{\lambda_0\}$. Dann ist auch $\mu_{\lambda_0} \cdot v_{\lambda_0} = 0$ und damit $\mu_{\lambda_0} = 0$. □

Korollar 6.2.11. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $\Lambda \subset K$ eine Teilmenge. Dann ist die kanonische Abbildung*

$$\bigoplus_{\lambda\in\Lambda} \mathrm{Eig}_\lambda(f) \to \sum_{\lambda\in\Lambda} \mathrm{Eig}_\lambda(f)$$

ein Isomorphismus.

Beweis. Dies folgt aus Propositionen 6.2.10 und 6.1.7. □

Definition 6.2.12 (geometrische Vielfachheit)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $\lambda \in K$. Die Dimension von $\mathrm{Eig}_\lambda(f)$ heißt die *geometrische Vielfachheit* von λ bzgl. f und wird mit $\mu_f^{\mathrm{geom}}(\lambda)$ bezeichnet.

Beispiel 6.2.13. Sei $A = \mathrm{diag}(d_1, \ldots, d_n)$ eine $n \times n$-Diagonalmatrix über K. Aus Beispiel 6.2.4(iv) folgt, dass $\mu_A^{\mathrm{geom}}(\lambda) = |\{i \in \{1, \ldots, n\} \mid d_i = \lambda\}|$.

Proposition 6.2.14 (Charakterisierung von Eigenwerten)**.** *Sei V ein endlich-dimensionaler K-Vektorraum, $f \in \mathrm{End}_K(V)$ ein Endomorphismus und $\lambda \in K$. Dann sind folgende Aussagen äquivalent:*

(i) *λ ist ein Eigenwert von f.*

(ii) *Es gilt $\mathrm{rg}(\lambda \cdot \mathrm{id}_V - f) < \dim_K V$.*

(iii) *Es gilt $\det(\lambda \cdot \mathrm{id}_V - f) = 0$.*

Beweis. Nach Definition ist λ genau dann ein Eigenwert von f, wenn die Abbildung $\lambda \,\mathrm{id}_V - f$ nicht injektiv ist. Aber wenn V endlich-dimensional ist, sind die Injektivität, Surjektivität und Bijketivität von $\lambda \,\mathrm{id}_V - f$ äquivalent (Korollar 4.1.39). Aussage (ii) bedeutet, dass $\lambda \,\mathrm{id}_V - f$ nicht surjektiv ist (nach Proposition 3.3.35), und Aussage (iii) bedeutet, dass $\lambda \,\mathrm{id}_V - f$ nicht bijektiv ist (Proposition 5.3.50). Deswegen sind alle Aussagen äquivalent. □

Nach Proposition 6.2.14 kann man die Eigenwerte von f finden, indem man die Gleichung $\det(\lambda \cdot \mathrm{id}_V - f) = 0$ löst. Nach der Leibniz-Formel ist $\det(\lambda \cdot \mathrm{id}_V - f)$ eine Polynomfunktion von λ, und es gibt leider keine allgemeine Methode zur Lösung solcher Gleichungen. Wir kommen auf diesen Punkt im Abschnitt 6.3 zurück.

In der Funktionalanalysis spielt die folgende Verallgemeinerung von Eigenwerten eine wichtige Rolle:

Definition 6.2.15 (Spektrum, Spektralwert)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Das *Spektrum* $\sigma(f)$ von f ist die Menge aller Skalare λ, so dass die Abbildung $\lambda \cdot \mathrm{id}_V - f$ nicht bijektiv ist. Elemente von $\sigma(f)$ heißen *Spektralwerte* von f.

Bemerkung 6.2.16 (Eigenwerte vs. Spektralwerte)**.** Eigenwerte von f sind auch Spektralwerte von f. Die Umkehrung gilt, wenn V endlich-dimensional ist (nach Korollar 4.1.39), aber nicht im Allgemeinen. Zum Beispiel hat der Endomorphismus

$$f\colon K^{\mathbb{N}} \to K^{\mathbb{N}}, \quad (x_0, x_1, \dots) \mapsto (0, x_0, x_1, \dots),$$

keine Eigenwerte, aber 0 ist ein Spektralwert von f, da f nicht surjektiv ist. Es ist eigentlich $\sigma(f) = \{0\}$.

Korollar 6.2.17 (Mächtigkeit des Spektrums mit geometrischer Vielfachheit gerechnet)**.** *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V der endlichen Dimension n. Dann gelten $|\sigma(f)| \leq n$ und $\sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) \leq n$.*

Beweis. Dies folgt aus dem Korollar 6.2.11, da $\sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) = \dim_K\big(\bigoplus_{\lambda \in \sigma(f)} \mathrm{Eig}_\lambda(f)\big)$ und $\mu_f^{\mathrm{geom}}(\lambda) \geq 1$ für alle $\lambda \in \sigma(f)$. □

Mit dem Gaußschen Eliminationsverfahren können wir die geometrischen Vielfachheiten bzgl. einer Matrix sowie Basen der zugehörigen Eigenräume leicht bestimmen:

Rezept 6.2.18 (Berechnung der geometrischen Vielfachheit)**.** Gegeben seien eine Matrix $A \in M_n(K)$ und ein Skalar $\lambda \in K$. Gesucht ist $\mu_A^{\mathrm{geom}}(\lambda)$. Man berechnet den Rang von $\lambda I_n - A$ mit dem Rezept 5.2.14. Nach der Dimensionsformel für lineare Abbildungen ist dann $\mu_A^{\mathrm{geom}}(\lambda) = n - \mathrm{rg}(\lambda I_n - A)$.

Rezept 6.2.19 (Berechnung des Eigenraums)**.** Gegeben seien eine Matrix $A \in M_n(K)$ und ein Skalar $\lambda \in K$. Gesucht ist eine Basis von $\mathrm{Eig}_\lambda(A)$. Dazu verwendet man das Rezept 5.2.15 mit der Matrix $\lambda I_n - A$.

6.2.1 Diagonalisierbarkeit

Definition 6.2.20 (diagonalisierbarer Endomorphismus)**.** Sei V ein K-Vektorraum. Ein Endomorphismus $f \in \mathrm{End}_K(V)$ heißt *diagonalisierbar*, wenn eine Basis von V bestehend aus Eigenvektoren von f existiert.

Eine quadratische Matrix A heißt *diagonalisierbar*, wenn L_A diagonalisierbar ist.

Beispiel 6.2.21.

(i) Sei V ein K-Vektorraum und sei $t\colon V \oplus V \to V \oplus V$, $(v, w) \mapsto (w, v)$ (siehe Beispiel 6.2.4). Ist $(v_i)_{i \in I}$ eine Basis von V und ist $\mathrm{char}(K) \neq 2$, so bilden die Eigenvektoren (v_i, v_i) und $(v_i, -v_i)$ von t eine Basis von $V \oplus V$, und damit ist t diagonalisierbar. Aber wenn $\mathrm{char}(K) = 2$ und $V \neq \{0\}$, dann erzeugen die Eigenvektoren von t den Untervektorraum $\{(v, v) \mid v \in V\}$ von $V \oplus V$, und damit ist t nicht diagonalisierbar.

(ii) Sei $n \in \mathbb{N}$ und sei D eine $n \times n$-Diagonalmatrix über K. Dann besteht der Standardbasis von K^n aus Eigenvektoren von D, und insbesondere ist D diagonalisierbar.

Beispiel 6.2.22. Der Endomorphismus D von $C^\infty(\mathbb{R},\mathbb{R})$ ist nicht diagonalisierbar, da nicht alle Funktionen aus $C^\infty(\mathbb{R},\mathbb{R})$ Linearkombinationen der Funktionen $x \mapsto \exp(\lambda x)$ sind (siehe Beispiel 6.2.7). Denn eine solche Linearkombination muss unbeschränkt oder konstant sein, aber es existiert beschränkte beliebig oft differenzierbare Funktionen, die nicht konstant sind (z.B., cos und sin).

Proposition 6.2.23. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Die folgenden Aussagen sind äquivalent:*

(i) *f ist diagonalisierbar.*

(ii) *Jeder Vektor $v \in V$ ist eine Linearkombination von Eigenvektoren von f.*

(iii) *Die kanonische lineare Abbildung $\bigoplus_{\lambda\in K} \mathrm{Eig}_\lambda(f) \to V$ ist ein Isomorphismus.*

Beweis. Die Implikation (i) $\Rightarrow$ (ii) ist klar, und die Implikation (ii) $\Rightarrow$ (iii) folgt aus Korollar 6.2.11. Ist die Abbildung $\bigoplus_{\lambda\in K} \mathrm{Eig}_\lambda(f) \to V$ ein Isomorphismus, so erhält man eine Basis von V bestehend aus Eigenvektoren, indem man Basen aller Eigenräume $\mathrm{Eig}_\lambda(f)$ zusammensetzt. □

Satz 6.2.24 (Charakterisierung der Diagonalisierbarkeit)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und sei $f \in \mathrm{End}_K(V)$. Dann sind die folgenden Aussagen äquivalent:*

(i) *f ist diagonalisierbar.*

(ii) *Es existiert eine Basis B von V, so dass $[f]_B^B$ eine Diagonalmatrix ist.*

(iii) *Es gilt*
$$\sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) = \dim_K V.$$

Beweis. Zu (i) $\Leftrightarrow$ (ii). Sei B eine Basis von V. Dann ist $[f]_B^B$ genau dann eine Diagonalmatrix, wenn B aus Eigenvektoren von f besteht.

Zu (i) $\Leftrightarrow$ (iii). Nach Korollar 6.2.11 ist $\sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{geom}}(\lambda)$ genau dann gleich $\dim_K V$, wenn die kanonische Abbildung $\bigoplus_{\lambda\in\sigma(f)} \mathrm{Eig}_\lambda(f) \to V$ ein Isomorphismus ist. Nach Proposition 6.2.23 ist das Letztere zur Diagonalisierbarkeit von f äquivalent. □

Korollar 6.2.25. *Sei V ein K-Vektorraum der Dimension $n \in \mathbb{N}$ und sei $f \in \mathrm{End}_K(V)$. Hat f n paarweise verschiedene Eigenwerte, so ist f diagonalisierbar.*

Beweis. Nach Definition ist die geometrische Vielfachheit eines Eigenwerts mindestens 1. Wenn f n paarweise verschiedene Eigenwerte besitzt, dann ist $\sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) \geq n$. Andererseits ist $\sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) \leq n$ nach Korollar 6.2.17. Nach Satz 6.2.24 (iii) $\Rightarrow$ (i) ist also f diagonalisierbar. □

Korollar 6.2.26 (Diagonalisierbarkeit von Matrizen)**.** *Seien $n \in \mathbb{N}$ und $A \in M_n(K)$. Die folgenden Aussagen sind äquivalent:*

(i) *A ist diagonalisierbar.*

(ii) *A ist ähnlich zu einer Diagonalmatrix.*

Beweis. Dies folgt aus Satz 6.2.24 (i) $\Leftrightarrow$ (ii) und Proposition 6.1.19. □

Rezept 6.2.27 (Test auf Diagonalisierbarkeit)**.** Gegeben seien eine Matrix $A \in M_n(K)$ und ihre Eigenwerte $\lambda_1, \ldots, \lambda_k$. Zu bestimmen ist, ob A diagonalisierbar ist. Wenn $k = n$ ist, ist A diagonalisierbar nach Korollar 6.2.25. Sonst berechnet man die geometrischen Vielfachheiten $\mu_A^{\mathrm{geom}}(\lambda_i)$ mit Rezept 6.2.18. Nach Satz 6.2.24 ist die Matrix A genau dann diagonalisierbar, wenn $\sum_{i=1}^k \mu_A^{\mathrm{geom}}(\lambda_i) = n$.

Rezept 6.2.28 (Diagonalisierung einer Matrix)**.** Gegeben seien eine Matrix $A \in M_n(K)$ und ihre Eigenwerte $\lambda_1, \dots, \lambda_k$. Gesucht ist eine Matrix $S \in \mathrm{GL}_n(K)$, wenn sie existiert, so dass $S^{-1}AS$ eine Diagonalmatrix ist. Mit Rezept 6.2.19 findet man Basen der Eigenräume $\mathrm{Eig}_{\lambda_i}(A)$, und dadurch wird auch bestimmt, ob A diagonalisierbar ist (siehe Rezept 6.2.27). Wenn ja, erhält man eine Basis $B = (v_1, \dots, v_n)$ von K^n, indem man die gefundenen Basen der Eigenräume zusammensetzt. Sei S die Matrix mit Spalten $v_1, \dots, v_n$, d.h., die Basiswechselmatrix T_E^B von B nach der Standardbasis E. Nach der Basiswechselformel (Proposition 4.2.43) hat dann die Matrix S die gewünschte Eigenschaft. Man kann außerdem die inverse Matrix S^{-1} mit Rezept 5.2.18 berechnen.

Beispiel 6.2.29. Wir können jetzt das Beispiel 4.2.44 erklären. Sei

$$A = \begin{pmatrix} 3 & -1 \\ 2 & 0 \end{pmatrix} \in M_2(\mathbb{R}).$$

Um die Eigenwerte von A zu finden, müssen wir die Gleichung $\det(\lambda I_2 - A) = 0$ lösen:

$$\det(\lambda I_2 - A) = (\lambda - 3)\lambda + 2 = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2),$$

und daher sind 1 und 2 die Eigenwerte von A. Insbesondere ist A diagonalisierbar. Mit Rezept 6.2.19 berechnen wir die zugehörigen Eigenräume:

$$I_2 - A = \begin{pmatrix} -2 & 1 \\ -2 & 1 \end{pmatrix} \xrightarrow{A_{21}(-1)} \begin{pmatrix} -2 & 1 \\ 0 & 0 \end{pmatrix} \implies \mathrm{Eig}_1(A) = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$
$$2I_2 - A = \begin{pmatrix} -1 & 1 \\ -2 & 2 \end{pmatrix} \xrightarrow{A_{21}(-2)} \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \implies \mathrm{Eig}_2(A) = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Also ist $B = \left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)\right)$ eine Basis von $\mathbb{R}^2$ bestehend aus Eigenvektoren von A, und damit ist

$$S^{-1}AS = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{wobei} \quad S = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Wir berechnen noch die inverse Matrix S^{-1} mit Rezept 5.2.18:

$$(S|I_2) = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right) \xrightarrow{A_{21}(-1)} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array}\right) \xrightarrow{A_{12}(-1)} \left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right) = (I_2|S^{-1}).$$

Beispiel 6.2.30. Sei

$$A = \begin{pmatrix} -5 & 2 & -4 \\ -2 & 0 & -2 \\ 4 & -2 & 3 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Wir versuchen A mit Rezept 6.2.28 zu diagonalisieren. Es ist

$$\begin{aligned} \det(\lambda I_3 - A) &= \det \begin{pmatrix} \lambda + 5 & -2 & 4 \\ 2 & \lambda & 2 \\ -4 & 2 & \lambda - 3 \end{pmatrix} \\ &\overset{\text{1.S}}{=} (\lambda + 5)(\lambda(\lambda - 3) - 4) - 2(-2(\lambda - 3) - 8) - 4(-4 - 4\lambda) \\ &= \lambda^3 + 2\lambda^2 + \lambda = \lambda(\lambda + 1)^2. \end{aligned}$$

Die Eigenwerte von A sind also 0 und -1. Als Nächstes berechnen wir die Eigenräume:

- *Eigenraum zu* -1.

$$-I_3 - A = \begin{pmatrix} 4 & -2 & 4 \\ 2 & -1 & 2 \\ -4 & 2 & -4 \end{pmatrix} \xrightarrow[V_{12}]{\substack{A_{12}(-2) \\ A_{32}(2)}} \begin{pmatrix} 2 & -1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Der Rang von $-I_2 - A$ ist also gleich 1, so dass $\mu_A^{\text{geom}}(-1) = 3-1 = 2$. Daraus können wir bereits schließen, dass A diagonalisierbar ist. Die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

bilden eine Basis von $\text{Eig}_{-1}(A) = \mathcal{L}(-I_3 - A, 0)$.

- *Eigenraum zu* 0.

$$0I_3 - A = \begin{pmatrix} 5 & -2 & 4 \\ 2 & 0 & 2 \\ -4 & 2 & -3 \end{pmatrix} \xrightarrow[V_{12}]{M_2(\frac{1}{2})} \begin{pmatrix} 1 & 0 & 1 \\ 5 & -2 & 4 \\ -4 & 2 & -3 \end{pmatrix}$$
$$\xrightarrow[A_{31}(4)]{A_{21}(-5)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & -2 & -1 \\ 0 & 2 & 1 \end{pmatrix} \xrightarrow{A_{32}(1)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & -2 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Damit wird $\text{Eig}_0(A) = \mathcal{L}(A, 0)$ von folgendem Vektor erzeugt:

$$v_3 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}.$$

Daraus erhalten wir $S^{-1}AS = \text{diag}(-1, -1, 0)$, wobei

$$S = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 0 & 1 \\ 0 & 1 & -2 \end{pmatrix}.$$

Schließlich berechnen wir die inverse Matrix S^{-1}:

$$(S|I_3) = \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{21}(-2)} \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 2 & -3 & -2 & 1 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \end{array}\right)$$
$$\xrightarrow[V_{23}]{A_{23}(-2)} \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & -2 \end{array}\right) \xrightarrow[A_{13}(-2)]{A_{23}(2)} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 5 & -2 & 4 \\ 0 & 1 & 0 & -4 & 2 & -3 \\ 0 & 0 & 1 & -2 & 1 & -2 \end{array}\right)$$
$$\xrightarrow{A_{12}(1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -4 & 2 & -3 \\ 0 & 0 & 1 & -2 & 1 & -2 \end{array}\right) = (I_3|S^{-1}).$$

Beispiel 6.2.31. Seien $\lambda, \alpha \in K$. Falls $\alpha \neq 0$ ist die Dreiecksmatrix

$$A = \begin{pmatrix} \lambda & \alpha \\ 0 & \lambda \end{pmatrix}$$

nicht diagonalisierbar. Denn A hat den einzigen Eigenwert λ, und seine geometrische Vielfachheit ist nur 1:

$$\lambda I_2 - A = \begin{pmatrix} 0 & -\alpha \\ 0 & 0 \end{pmatrix} \quad \Longrightarrow \quad \text{Eig}_\lambda(A) = K \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Bemerkung 6.2.32 (Potenzen einer diagonalisierbaren Matrix)**.** Sei $A \in M_n(K)$ eine quadratische Matrix. Ist A diagonalisierbar, so können wir die Potenzen A^k von A wie folgt berechnen. Sei $S \in \text{GL}_n(K)$, so dass $S^{-1}AS = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Dann ist

$$A^k = (S\,\text{diag}(\lambda_1, \ldots, \lambda_n)S^{-1})^k = S\,\text{diag}(\lambda_1, \ldots, \lambda_n)^k S^{-1} = S\,\text{diag}(\lambda_1^k, \ldots, \lambda_n^k)S^{-1}.$$

Beispiel 6.2.33 (Fibonacci-Zahlen)**.** Sei $(F_n)_{n\in\mathbb{N}}$ die Folge der Fibonacci-Zahlen. Im Beispiel 4.2.19 haben wir die folgende Formel erreicht:

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = A^n \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{wobei} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R}).$$

Es ist $\det(\lambda I_2 - A) = \lambda(\lambda - 1) - 1 = \lambda^2 - \lambda - 1$, und damit hat A die zwei Eigenwerte

$$\frac{1 \pm \sqrt{5}}{2}.$$

Insbesondere ist A diagonalisierbar. Durch die Methode aus Bemerkung 6.2.32 erhalten wir die explizite Formel

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right).$$

Proposition 6.2.34 (Ähnlichkeit von Diagonalmatrizen)**.** *Sei $n \in \mathbb{N}$ und seien $D = \operatorname{diag}(d_1, \ldots, d_n)$ und $D' = \operatorname{diag}(d'_1, \ldots, d'_n)$ zwei Diagonalmatrizen über K. Dann sind die folgenden Aussagen äquivalent:*

(i) *D und D' sind ähnlich.*

(ii) *Es gibt eine Permutation $\sigma \in S_n$, so dass $d'_i = d_{\sigma(i)}$ für alle $i \in \{1, \ldots, n\}$.*

Beweis. Für die Elementarmatrix V_{ij} gilt $V_{ij}^{-1} D V_{ij} = \operatorname{diag}(d_{\tau(1)}, \ldots, d_{\tau(n)})$, wobei τ die Transposition $(i\ j)$ ist. Da jede Permutation σ eine Komposition von Transpositionen ist (Lemma 5.3.4), folgt daraus die Implikation (ii) $\Rightarrow$ (i). Sind umgekehrt D und D' ähnlich, so haben D und D' die gleichen Eigenwerte mit den gleichen geometrischen Vielfachheiten (siehe Bemerkung 6.2.3). Aber die Diagonalkoeffizienten einer Diagonalmatrix sind genau ihre Eigenwerte, und ein Eigenwert kommt so oft vor wie seine geometrische Vielfachheit (siehe Beispiel 6.2.13). Deshalb gilt die Implikation (i) $\Rightarrow$ (ii). □

Bemerkung 6.2.35 (Klassifikation von diagonalisierbaren Endomorphismen bis auf Isomorphie)**.** Sei $\mathcal{D}\mathrm{iag}\mathcal{E}\mathrm{nd}_n$ die Menge aller Isomorphieklassen von Paaren (V, f), wobei V ein n-dimensionaler K-Vektorraum ist und f ein diagonalisierbarer Endomorphismus von V ist. Sei $\sim$ die folgende Äquivalenzrelation auf K^n: $x \sim y$ genau dann, wenn eine Permutation $\sigma \in S_n$ existiert, so dass $y_i = x_{\sigma(i)}$ für alle $i \in \{1, \ldots, n\}$. Nach Proposition 6.2.34 gibt es dann eine bijektive Abbildung

$$\begin{aligned} K^n/\!\sim \ &\xrightarrow{\sim} \mathcal{D}\mathrm{iag}\mathcal{E}\mathrm{nd}_n, \\ [(d_1, \ldots, d_n)] &\mapsto [(K^n, L_{\operatorname{diag}(d_1,\ldots,d_n)})]. \end{aligned}$$

Die Umkehrabbildung schickt einen diagonalisierbaren Endomorphimus $f \in \operatorname{End}_K(V)$ auf das n-Tupel seiner Eigenwerte mit geometrischer Vielfachheit gezählt.

Proposition 6.2.36 (Determinante und Spur diagonalisierbarer Endomorphismen)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und sei $f \in \operatorname{End}_K(V)$ ein diagonalisierbarer Endomorphismus. Dann gilt:*

(i) $\det(f) = \prod_{\lambda \in \sigma(f)} \lambda^{\mu_f^{\mathrm{geom}}(\lambda)}$.

(ii) $\operatorname{tr}(f) = \sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) \cdot \lambda$.

Beweis. Sei B eine Basis von V, so dass $[f]_B^B$ eine Diagonalmatrix ist. Dann ist $\det(f)$ bzw. $\operatorname{tr}(f)$ das Produkt bzw. die Summe der Diagonalkoeffizienten von $[f]_B^B$, die genau die Eigenwerte von f sind, mit geometrischer Vielfachheit gezählt. □

6.3 Das charakteristische Polynom

6.3.1 Polynome

Zur Erinnerung ist $K^{(\mathbb{N})}$ der K-Vektorraum aller Folgen $(a_n)_{n\in\mathbb{N}}$ in K, die null außerhalb einer endlichen Teilmenge von $\mathbb{N}$ ist. Die Folgen $(\delta_{in})_{n\in\mathbb{N}}$ mit $i \in \mathbb{N}$ bilden eine Basis von $K^{(\mathbb{N})}$ (siehe Beispiel 3.3.13).

Sei T ein Symbol, das wir als *Variable* oder *Unbestimmte* benennen. Dann bezeichnen wir mit $K[T]$ der K-Vektorraum $K^{(\mathbb{N})}$, in dem wir die Folge $(\delta_{in})_{n\in\mathbb{N}}$ als T^i schreiben. Also ist $(T^i)_{i\in\mathbb{N}}$ eine Basis des K-Vektorraums $K[T]$, so dass jedes Element $p \in K[T]$ als Linearkombination

$$p = \sum_{i\in\mathbb{N}} a_i T^i$$

geschrieben werden kann, wobei $a_i \in K$ und nur endlich viele der a_i nicht null sind. Außerdem schreiben wir T anstelle von T^1 und identifizieren wir die von T^0 aufgespannte Gerade in $K[T]$ mit K durch die injektive lineare Abbildung

$$K \hookrightarrow K[T], \quad a \mapsto aT^0.$$

Definition 6.3.1 (Polynom, Monom, Glied, Absolutglied)**.** Mit der obigen Schreibweise bezeichnen wir Elemente von $K[T]$ als *Polynome über K* (oder *mit Koeffizienten in K*) in der Variablen T. Ist $p = \sum_{i\in\mathbb{N}} a_i T^i$ ein Polynom, so heißen die Skalare a_i die *Koeffizienten* von p. Der Nullvektor $0 \in K[T]$ heißt das *Nullpolynom*.

Ein Polynom mit höchstens einem Nicht-Null-Koeffizient heißt *Monom*. Die Monome a_iT^i heißen die *Glieder* des Polynoms $\sum_{i\in\mathbb{N}} a_iT^i$, und der Koeffizient a_0 heißt das *Absolutglied*.

Definition 6.3.2 (Multiplikation von Polynomen)**.** Seien $p = \sum_{i\in\mathbb{N}} a_iT^i$ und $q = \sum_{i\in\mathbb{N}} b_iT^i$ zwei Polynome über K. Man definiert das Produkt $p \cdot q$ durch

$$p \cdot q = \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} a_i b_j T^{i+j} = \sum_{k\in\mathbb{N}} \left(\sum_{i+j=k} a_i b_j \right) T^k.$$

Man beachte dabei, dass diese Summen nur endlich viele Summanden enthalten, die nicht null sind, so dass $p \cdot q$ ein wohldefiniertes Polynom ist.

Beispiel 6.3.3.

(i) Sind $\lambda, \mu \in K$, so gilt $(T + \lambda) \cdot (T + \mu) = T^2 + (\lambda + \mu)T + \lambda\mu$. Insbesondere ist $(T + \lambda) \cdot (T - \lambda) = T^2 - \lambda^2$.

(ii) Für jedes $n \in \mathbb{N} \setminus \{0\}$ gilt $T^n - 1 = (T - 1) \cdot (T^{n-1} + \ldots + T + 1)$ (die rechte Seite ist eine „Teleskopsumme“).

Proposition 6.3.4. *Die Multiplikation auf $K[T]$ ist assoziativ und kommutativ, sie hat das neutrale Element $1 = T^0$ und sie ist distributiv über die Addition.*

Beweis. Seien $p = \sum_{i\in\mathbb{N}} a_iT^i$, $q = \sum_{i\in\mathbb{N}} b_iT^i$ und $r = \sum_{i\in\mathbb{N}} c_iT^i$ Polynome über K. Die Assoziativität folgt daraus, dass beide $p \cdot (q \cdot r)$ und $(p \cdot q) \cdot r$ gleich

$$\sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} \sum_{k\in\mathbb{N}} a_i b_j c_k T^{i+j+k}$$

sind. Die Kommutativität und die Neutralität von T^0 sind klar. Zur Distributivität berechnen wir:

$$\begin{aligned} p \cdot (q + r) &= \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} a_i(b_j + c_j)T^{i+j} \\ &= \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} a_i b_j T^{i+j} + \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} a_i c_j T^{i+j} \\ &= p \cdot q + p \cdot r. \end{aligned}$$

□

Bemerkung 6.3.5. Insbesondere ist $(K[T], +, \cdot)$ ein kommutativer Ring (siehe Bemerkung 2.3.5), den wir als *Polynomring* über K bezeichnen. Außerdem ist die Multiplikation auf $K[T]$ mit der Skalarmultiplikation kompatibel: Sind $p, q \in K[T]$ und ist $\lambda \in K$, so gilt

$$(\lambda \cdot p) \cdot q = \lambda \cdot (p \cdot q) = p \cdot (\lambda \cdot q).$$

Ein kommutativer Ring mit einer kompatiblen Struktur von K-Vektorraum in diesem Sinne heißt eine *kommutative K-Algebra*.

Definition 6.3.6 (Grad, Leitkoeffizient, monisches Polynom)**.** Der *Grad* eines Polynoms $p = \sum_{i \in \mathbb{N}} a_i T^i$ über K ist

$$\deg(p) := \sup\{i \in \mathbb{N} \mid a_i \neq 0\} \in \{-\infty\} \cup \mathbb{N},$$

wobei $\sup \varnothing = -\infty$. Das Nullpolynom ist also das einzige Polynom vom Grad $-\infty$, und ein Polynom $p \in K[T]$ vom Grad $d \geq 0$ kann wie folgt geschrieben werden, mit $a_d \neq 0$:

$$p = a_d T^d + a_{d-1} T^{d-1} + \cdots + a_1 T + a_0.$$

Der Koeffizient $a_d \in K \setminus \{0\}$ heißt der *Leitkoeffizient* von p. Ein Polynom $p \in K[T]$ heißt *monisch* oder *normiert*, wenn $\deg(p) \geq 0$ und der Leitkoeffizient von p gleich 1 ist.

Proposition 6.3.7 (Eigenschaften des Grades)**.** *Seien $p, q \in K[T]$ Polynome und $\lambda \in K$.*

(i) $\deg(p) = -\infty \iff p = 0$.

(ii) *Es gilt* $\deg(p \cdot q) = \deg(p) + \deg(q)$.

(iii) *Es gilt* $\deg(p + q) \leq \max\{\deg(p), \deg(q)\}$. *Die Gleichheit gilt, wenn* $\deg(p) \neq \deg(q)$.

(iv) *Es gilt* $\deg(\lambda \cdot p) \leq \deg(p)$. *Die Gleichheit gilt, wenn $\lambda \in K^\times$.*

Beweis. Dies folgt unmittelbar aus den Definitionen. □

Definition 6.3.8 (Einsetzung, Polynomfunktion)**.** Sei $p = \sum_{i \in \mathbb{N}} a_i T^i \in K[T]$ und $\lambda \in K$. Die *Einsetzung* von λ in p ist $p(\lambda) := \sum_{i \in \mathbb{N}} a_i \lambda^i \in K$. Die Abbildung

$$\begin{aligned} K &\to K, \\ \lambda &\mapsto p(\lambda), \end{aligned}$$

heißt die dem Polynom p zugehörige *Polynomfunktion* auf K. Man bezeichnet mit $\mathrm{Poly}(K, K)$ die Menge aller Polynomfunktionen auf K, d.h., das Bild der Abbildung $K[T] \to \mathrm{Abb}(K, K)$, die p auf $\lambda \mapsto p(\lambda)$ abbildet. Man kann leicht nachprüfen, dass die letztere Abbildung linear ist, so dass $\mathrm{Poly}(K, K)$ ein Untervektorraum von $\mathrm{Abb}(K, K)$ ist.

Beispiel 6.3.9. Die Polynomfunktionen vom Grad $\leq n$ auf $\mathbb{R}$ bilden den Kern der $(n+1)$-ten Differentiationsabbildung $D^{n+1} \colon C^\infty(\mathbb{R}, \mathbb{R}) \to C^\infty(\mathbb{R}, \mathbb{R})$.

Bemerkung 6.3.10 (Polynome vs. Polynomfunktionen)**.** Man sollte Polynome über K nicht mit Polynomfunktionen auf K verwechseln: Ist K ein endlicher Körper, so gibt es nur endlich viele Polynomfunktionen $K \to K$, aber trotzdem unendlich viele Polynome über K. Insbesondere gibt es in diesem Fall viele verschiedene Polynome über K, die dieselbe Polynomfunktion induzieren. Zum Beispiel induzieren alle Polynome T^{2i+1} über $\mathbb{F}_3$ die Identität auf $\mathbb{F}_3$.

Definition 6.3.11 (Teilbarkeit)**.** Seien $f, g \in K[T]$. Man sagt, dass g *teilt* f oder dass f durch g *teilbar* ist, und man schreibt $g | f$, wenn ein Polynom $q \in K[T]$ existiert, so dass $f = qg$.

Satz 6.3.12 (Polynomdivision mit Rest)**.** *Seien $f, g \in K[T]$ zwei Polynome mit $g \neq 0$. Dann existieren eindeutige Polynome $q, r \in K[T]$ mit $f = qg + r$ und $\deg(r) < \deg(g)$.*

Beweis. Zur Eindeutigkeit. Seien $f = q_1 g + r_1$ und $f = q_2 g + r_2$ mit $\deg(r_1) < \deg(g)$ und $\deg(r_2) < \deg(g)$. Dann gilt:

$$0 = f - f = (q_1 - q_2)g + (r_1 - r_2)$$

und damit $(q_1 - q_2)g = r_2 - r_1$. Nach Proposition 6.3.7 gelten $\deg((q_1 - q_2)g) = \deg(q_1 - q_2) + \deg(g)$ und $\deg(r_2 - r_1) \leq \max\{\deg(r_1), \deg(r_2)\} < \deg(g)$. Folglich ist $\deg(q_1 - q_2) < 0$, d.h., $q_1 = q_2$. Dann ist auch $r_1 = f - q_1 g = f - q_2 g = r_2$.

Zur Existenz. Seien $n = \deg(f)$ und $m = \deg(g) \geq 0$. Wir beweisen die Existenz von q und r durch vollständige Induktion über n. Falls $n < m$ kann man $q = 0$ und $r = f$ nehmen. Falls $n \geq m$ schreibt man $f = aT^n + \bar{f}$ und $g = bT^m + \bar{g}$ mit $a, b \in K \setminus \{0\}$, $\deg(\bar{f}) < n$ und $\deg(\bar{g}) < m$. Man setzt $q_1 = b^{-1}aT^{n-m}$ und $f_1 = f - q_1 g$, so dass $f = q_1 g + f_1$. Dann ist

$$\deg(f_1) = \deg(\bar{f} - b^{-1}aT^{n-m}\bar{g}) \leq \max\{\deg(\bar{f}), \deg(\bar{g}) + n - m\} < n.$$

Nach der Induktionsvoraussetzung existieren $q', r \in K[T]$ mit $f_1 = q'g + r$ und $\deg(r) < \deg(g)$, und damit ist

$$f = (q_1 + q')g + r,$$

wie gewünscht. □

Beispiel 6.3.13. Der Beweis der Existenz von q und r im Satz 6.3.12 ist völlig konstruktiv und liefert einen Algorithmus, der ganz ähnlich wie die gewöhnliche schriftliche Division mit Rest von ganzen Zahlen ist. Als Beispiel sei $f = T^3 + T + 1$ und $g = T - 2$. Dann erhalten wir $q = T^2 + 2T + 5$ und $r = 11$:

$$\begin{array}{rl}
\big(T^3 & +T \quad +1\big) : \big(T-2\big) = T^2 + 2T + 5 \text{ mit Rest } 11. \\
\underline{-T^3 + 2T^2} & \\
2T^2 & +T \quad +1 \\
\underline{-2T^2 + 4T} & \\
5T & +1 \\
\underline{-5T + 10} & \\
11 &
\end{array}$$

Insbesondere ist $T^3 + T + 1$ genau dann durch $T - 2$ teilbar, wenn $\mathrm{char}(K) = 11$.

Definition 6.3.14 (Nullstelle)**.** Sei $p \in K[T]$ ein Polynom. Eine *Nullstelle* von p ist ein Skalar $a \in K$, so dass $p(a) = 0$.

Proposition 6.3.15. *Sei $p \in K[T]$ und $a \in K$. Die folgenden Aussagen sind äquivalent:*

(i) *a ist eine Nullstelle von p.*

(ii) *p ist durch $T - a$ teilbar.*

Beweis. Ist $p = (T - a)q$, so ist $p(a) = (a - a)q(a) = 0$. Sei umgekehrt a eine Nullstelle von p. Nach Satz 6.3.12 gibt es Polynome $q, r \in K[T]$ mit $p = (T - a)q + r$ und $\deg(r) < \deg(T - a) = 1$, d.h., $r \in K$. Dann gilt $0 = p(a) = (a - a)q(a) + r = r$, und damit ist $p = (T - a)q$. □

Beispiel 6.3.16. Sei $p \in \mathbb{N}$ eine Primzahl. Nach dem *kleinen Fermatschen Satz*, der in der *Algebra* Vorlesung bewiesen wird, gilt $n^p \equiv n \mod p$ für jede ganze Zahl n. Anders

gesagt ist jedes $a \in \mathbb{F}_p$ eine Nullstelle des Polynoms $T^p - T$. Durch p Anwendungen der Proposition 6.3.15 erhalten wir die Gleichung

$$T^p - T = \prod_{a \in \mathbb{F}_p} (T - a)$$

in $\mathbb{F}_p[T]$.

Korollar 6.3.17. *Sei $p \in K[T]$ ein Polynom vom Grad $d \geq 0$. Dann hat p höchstens d verschiedene Nullstellen.*

Beweis. Wir verwenden Induktion über d. Ein Polynom vom Grad 0 hat keine Nullstellen. Falls $d \geq 1$ und λ eine Nullstelle von p ist, dann existiert $q \in K[T]$ mit $p = (T - \lambda)q$ nach Proposition 6.3.15. Nach der Nullteilerfreiheit von K müssen alle anderen Nullstellen von p auch Nullstellen von q sein. Nach der Induktionsvoraussetzung hat q höchstens $d - 1$ Nullstellen, und damit hat p höchstens d Nullstellen. □

Korollar 6.3.18. *Ist K unendlich, so ist die lineare Abbildung $K[T] \to \mathrm{Poly}(K, K)$ aus Definition 6.3.8 ein Isomorphismus.*

Beweis. Sei p ein Polynom im Kern dieser Abbildung. Dann ist jedes $a \in K$ eine Nullstelle von p. Da K unendlich ist, folgt aus Korollar 6.3.17, dass $\deg(p) = -\infty$, d.h., dass $p = 0$. □

Definition 6.3.19 (Vielfachheit)**.** Sei $p \in K[T]$ und $a \in K$. Die *Vielfachheit* von a in p ist

$$v_a(p) := \sup\{n \in \mathbb{N} \mid (T - a)^n \text{ teilt } p\} \in \mathbb{N} \cup \{+\infty\}.$$

Bemerkung 6.3.20. Nach Proposition 6.3.15 ist a ist genau dann eine Nullstelle von p, wenn $v_a(p) \geq 1$. Für das Nullpolynom 0 gilt $v_a(0) = +\infty$.

Sind $p, q \in K[T]$, so gilt

$$v_a(pq) = v_a(p) + v_a(q).$$

Die Ungleichung $v_a(p)+v_a(q) \leq v_a(pq)$ ist klar. Angenommen, es wäre $v_a(p)+v_a(q) < v_a(pq)$. Dann gibt es Polynome $r, \bar{p}, \bar{q} \in K[T]$, so dass $pq = (T - a)^{v_a(p)+v_a(q)} r$, $p = (T - a)^{v_a(p)}\bar{p}$ und $q = (T - a)^{v_a(q)}\bar{q}$ mit $r(a) = 0$, $\bar{p}(a) \neq 0$ und $\bar{q}(a) \neq 0$. Aus der Eindeutigkeitsaussage im Satz 6.3.12 folgt $r = \bar{p}\bar{q}$. Insbesondere ist $r(a) = \bar{p}(a)\bar{q}(a) \neq 0$, was ein Widerspruch ist.

Definition 6.3.21 (Zerfall in Linearfaktoren)**.** Man sagt, dass ein Polynom $p \in K[T]$ *in seine Linearfaktoren zerfällt*, wenn p ein Produkt von Polynomen vom Grad ≤ 1 ist.

Bemerkung 6.3.22. Sei $p \in K[T] \setminus \{0\}$. Nach dem Satz 6.3.12 und der Bemerkung 6.3.20 kann man schreiben

$$p = q \cdot \prod_{a \in K} (T - a)^{v_a(p)}$$

mit einem eindeutig bestimmten Polynom q, das keine Nullstellen hat. Insbesondere ist

$$\sum_{a \in K} v_a(p) \leq n,$$

und die Gleichheit gilt genau dann, wenn p in seine Linearfaktoren zerfällt.

Korollar 6.3.23. *Ist K algebraisch abgeschlossen (Definition 2.4.5), so zerfällt jedes Polynom $p \in K[T]$ in seine Linearfaktoren.*

Beweis. Wir verwenden Induktion über $\deg(p)$. Wenn $\deg(p) \leq 1$ gibt es nichts zu zeigen. Falls $\deg(p) \geq 2$, dann existiert eine Nullstelle a von p nach Definition eines algebraisch abgeschlossenen Körpers. Nach Korollar 6.3.17 ist $p = (T - a)q$ mit $\deg(q) = \deg(p) - 1$. Nach der Induktionsvoraussetzung zerfällt q in seine Linearfaktoren, und somit auch p. □

Bemerkung 6.3.24. Sei $p = aT^2 + bT + c \in K[T]$ ein Polynom vom Grad 2 (d.h., $a \neq 0$). Falls $\mathrm{char}(K) \neq 2$ kann man die Nullstellen von p durch die gewöhnliche Mitternachtsformel ausdrücken:

$$\frac{-b \pm \sqrt{\Delta}}{2a}, \quad \Delta = b^2 - 4ac,$$

wobei $\pm\sqrt{\Delta}$ die Quadratwurzeln von Δ sind (wenn sie existieren; sonst hat p keine Nullstellen in K). Wenn $\mathrm{char}(K) \notin \{2, 3\}$ gibt es auch kompliziertere Wurzelausdrücke für die Nullstellen eines allgemeinen Polynoms vom Grad 3 oder 4. Bei Polynomen des fünften Grades und höher existiert aber keine allgemeine Wurzelausdruck für die Nullstellen (das wird in der Vorlesung *Algebra* genauer formuliert und auch bewiesen).

6.3.2 Das charakteristische Polynom

Zu jedem Endomorphismus $f\colon V \to V$ eines endlich-dimensionalen K-Vektorraums V gibt es ein kanonisches monisches Polynom $\chi_f \in K[T]$ vom Grad $\dim_K V$, das *charakteristische Polynom* von f, mit folgender Eigenschaft: Die Nullstellen von χ_f sind genau die Eigenwerte von f. Außerdem sind die Koeffizienten von χ_f eine gemeinsame Verallgemeinerung der Determinante $\det(f)$ und der Spur $\mathrm{tr}(f)$.

Wir definieren zunächst das charakteristische Polynom χ_A einer quadratischen Matrix A, und danach zeigen wir, dass $\chi_{[f]_B^B}$ unabhängig von der Wahl der Basis B ist. Dazu braucht man *Matrizen von Polynomen* zu betrachten: Eine $m \times n$-Matrix mit Koeffizienten in $K[T]$ ist einfach eine $\{1, \dots, m\} \times \{1, \dots, n\}$-indizierte Familie in $K[T]$.

Definition 6.3.25 (Determinante einer Matrix von Polynomen)**.** Sei $n \in \mathbb{N}$ und sei A eine $n \times n$-Matrix mit Koeffizienten in $K[T]$. Die *Determinante* von A ist das Polynom

$$\det(A) := \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot \prod_{i=1}^{n} A_{\sigma(i)i} \in K[T].$$

Obwohl $(K[T], +, \cdot)$ kein Körper ist, viele (aber nicht alle) der Resultate über die Determinante von Matrizen, die wir im Abschnitt 5.3.3 bewiesen haben, bleiben gültig für Matrizen von Polynomen, mit denselben Beweisen. Zum Beispiel gelten der Laplacesche Entwicklungssatz und seine Korollare, d.h., man kann die Determinante einer Matrix von Polynomen durch Spalten- oder Zeilenentwicklung berechnen. Für $A, B \in M_n(K[T])$ gilt auch $\det(A \cdot B) = \det(A) \cdot \det(B)$ in $K[T]$, denn man diese Formel direkt aus der Leibniz-Formel nachrechnen kann.

Definition 6.3.26 (charakteristisches Polynom einer Matrix)**.** Sei A eine $n \times n$-Matrix über K. Das *charakteristische Polynom* von A ist das Polynom

$$\chi_A := \det(T \cdot I_n - A) \in K[T].$$

Bemerkung 6.3.27. Das charakteristische Polynom von A wird manchmal als die Determinante von $A - T \cdot I_n$ definiert. Nach Proposition 5.3.29(ii) gilt

$$\det(A - T \cdot I_n) = (-1)^n \chi_A.$$

Für viele Zwecke (z.B. um die Nullstellen zu bestimmen) macht das Vorzeichen $(-1)^n$ keinen Unterschied. Der Vorteil unserer Definition ist, dass χ_A immer ein monisches Polynom ist (siehe Proposition 6.3.32(i)).

Beispiel 6.3.28.

(i) Ist $A = (a_{ij})_{i,j}$ eine $n \times n$-Dreiecksmatrix (Definition 4.2.11), so ist

$$\chi_A = \prod_{i=1}^{n} (T - a_{ii})$$

nach Korollar 5.3.34.

(ii) Sei

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & -2 \\ 3 & -1 & 5 \end{pmatrix}.$$

Entwicklung nach der zweiten Spalte ergibt:

$$\begin{aligned} \chi_A &= \det \begin{pmatrix} T-2 & 0 & 1 \\ 0 & T-1 & 2 \\ -3 & 1 & T-5 \end{pmatrix} \\ &= (T-1)\det \begin{pmatrix} T-2 & 1 \\ -3 & T-5 \end{pmatrix} - \det \begin{pmatrix} T-2 & 1 \\ 0 & 2 \end{pmatrix} \\ &= (T-1)((T-2)(T-5)+3) - 2(T-2) = T^3 - 8T^2 + 18T - 9. \end{aligned}$$

Proposition 6.3.29. *Sei $n \in \mathbb{N}$ und seien $A, B \in M_n(K)$. Sind A und B ähnlich, so gilt $\chi_A = \chi_B$.*

Beweis. Sei $S \in \mathrm{GL}_n(K)$ mit $B = S^{-1} \cdot A \cdot S$. Dann ist

$$TI_n - B = TI_n - S^{-1}AS = S^{-1}(TI_n - A)S$$

in $M_n(K[T])$, und damit $\chi_B = \det(S)^{-1} \cdot \chi_A \cdot \det(S) = \chi_A$. □

Definition 6.3.30 (charakteristisches Polynom eines Endomorphismus)**.** Sei V ein K-Vektorraum der endlichen Dimension n und sei $f \in \mathrm{End}_K(f)$ ein Endomorphismus von V. Das *charakteristische Polynom* von f ist das Polynom

$$\chi_f := \chi_{[f]_B^B},$$

wobei B eine Basis von V ist. Nach Proposition 6.3.29 ist χ_f unabhängig von der Wahl der Basis B.

Bemerkung 6.3.31. Es ist auch möglich, eine matrixfreie Definition von χ_f zu geben. Dazu braucht man aber mehrere Erweiterungen der bisherigen betrachteten Begriffe. Man betrachtet nämlich die Menge $V[T]$ von Polynomen mit Koeffizienten in V, die ein „Vektorraum über $K[T]$“ ist. Man kann dann die Menge $\mathrm{Det}(V[T])$ von $K[T]$-multilinearen Determinantenfunktionen auf $V[T]$ definieren, und man kann zeigen, dass $\mathrm{End}_{K[T]}(\mathrm{Det}(V[T]))$ zu $K[T]$ kanonisch isomorph ist (vgl. Lemma 5.3.45). Die Determinante einer $K[T]$-linearen Abbildung $F\colon V[T] \to V[T]$ ist dann das eindeutige Polynom $\det(F)$, so dass $F^*(\Delta) = \det(F) \cdot \Delta$ für alle Determinantenfunktionen $\Delta \in \mathrm{Det}(V[T])$. Dann ist $\chi_f = \det(T \cdot \mathrm{id}_{V[T]} - f_{V[T]})$, wobei die Abbildung $f_{V[T]}\colon V[T] \to V[T]$ ein Polynom $\sum_{i\in\mathbb{N}} v_i T^i$ auf $\sum_{i\in\mathbb{N}} f(v_i)T^i$ abbildet.

Proposition 6.3.32 (Koeffizienten und Nullstellen des charakteristischen Polynoms)**.** *Sei V ein K-Vektorraum der endlichen Dimension n und sei $f \in \mathrm{End}_K(V)$.*

(i) *χ_f ist ein monisches Polynom vom Grad n.*

(ii) *Der Koeffizient von T^{n-1} in χ_f ist $-\mathrm{tr}(f)$.*

(iii) *Das Absolutglied von χ_f ist $(-1)^n \det(f)$.*

(iv) *Die Nullstellen von χ_f sind genau die Eigenwerte von f.*

Beweis. Es genügt die entsprechenden Aussagen über das charakteristische Polynom einer $n \times n$-Matrix A zu beweisen. Nach Definition ist

$$\chi_A = \sum_{\sigma \in S_n} \chi_A^\sigma, \quad \text{wobei} \quad \chi_A^\sigma := \mathrm{sgn}(\sigma) \cdot \prod_{i=1}^{n} (TI_n - A)_{\sigma(i)i}.$$

Wenn $\chi_A^\sigma \neq 0$ ist der Grad von χ_A^σ gleich der Anzahl der Fixpunkte von σ, d.h., der Indizes i mit $\sigma(i) = i$. Also ist $\deg(\chi_A^\sigma) \leq n$ für alle $\sigma \in S_n$, und es kann nur $\geq n-1$ sein, wenn $\sigma = \mathrm{id}$. In diesem Fall ist

$$\chi_A^{\mathrm{id}} = \prod_{i=1}^{n}(T - A_{ii}) = T^n - \left(\sum_{i=1}^{n} A_{ii}\right) T^{n-1} + \cdots,$$

und damit ist χ_A monisch und ist der Koeffizient von T^{n-1} gleich $-\operatorname{tr}(A)$. Das Absolutglied erhält man, indem man 0 für T in χ_A einsetzt:

$$\chi_A(0) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n}(-A)_{\sigma(i)i} = \det(-A) = (-1)^n \det(A).$$

Allgemeiner ist $\chi_A(\lambda) = \det(\lambda I_n - A)$. Nach Proposition 6.2.14 ist deshalb ein Skalar genau dann eine Nullstelle von χ_A, wenn er ein Eigenwert von A ist. □

Beispiel 6.3.33. Für einen Endomorphismus f eines 2-dimensionalen K-Vektorraums gilt $\chi_f = T^2 - \operatorname{tr}(f)T + \det(f)$.

Definition 6.3.34 (algebraische Vielfachheit)**.** Sei V ein endlich-dimensionaler K-Vektorraum, sei $f \in \mathrm{End}_K(V)$ und sei $\lambda \in K$. Die *algebraische Vielfachheit* von λ bzgl. f ist die Vielfachheit von λ in χ_f im Sinne der Definition 6.3.19. Sie wird mit $\mu_f^{\mathrm{alg}}(\lambda) \in \mathbb{N}$ bezeichnet.

Proposition 6.3.35 (Zerlegung des charakteristischen Polynoms)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus.*

(i) *Sei $U \subset V$ ein f-invarianter Untervektorraum und seien $f_U \in \mathrm{End}_K(U)$ und $\bar{f} \in \mathrm{End}_K(V/U)$ die von f induzierten Endomorphismen. Dann gilt:*

$$\chi_f = \chi_{f_U} \cdot \chi_{\bar{f}}.$$

(ii) *Seien $U, W \subset V$ komplementäre f-invariante Untervektorräume und seien $f_U \in \mathrm{End}_K(U)$ und $f_W \in \mathrm{End}_K(W)$ die von f induzierten Endomorphismen. Dann gilt:*

$$\chi_f = \chi_{f_U} \cdot \chi_{f_W}.$$

Beweis. Man wählt eine Basis $C = (v_1, \ldots, v_m)$ von U und eine Familie $D = (v_{m+1}, \ldots, v_n)$, so dass $B = (v_1, \ldots, v_n)$ eine Basis von V ist. Dann ist $\bar{D} = (v_{m+1} + U, \ldots, v_n + U)$ eine Basis von V/U, und die Darstellungsmatrix von f bzgl. B hat die Form

$$[f]_B^B = \begin{pmatrix} [f_U]_C^C & * \\ 0 & [\bar{f}]_{\bar{D}}^{\bar{D}} \end{pmatrix}.$$

Aus Korollar 5.3.36 folgt, dass $\chi_f = \chi_{f_U} \cdot \chi_{\bar{f}}$. Falls W ein f-invariantes direktes Komplement von U ist, kann man für D eine Basis von W wählen. Dann ist

$$[f]_B^B = \begin{pmatrix} [f_U]_C^C & 0 \\ 0 & [f_W]_D^D \end{pmatrix},$$

und damit $\chi_f = \chi_{f_U} \cdot \chi_{f_W}$. □

Korollar 6.3.36 (geometrische vs. algebraische Vielfachheit)**.** *Sei $f\colon V \to V$ ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V.*

(i) *Für alle $\lambda \in K$ gilt $\mu_f^{\mathrm{geom}}(\lambda) \leq \mu_f^{\mathrm{alg}}(\lambda)$.*

(ii) *Ist f diagonalisierbar, so gilt $\mu_f^{\mathrm{alg}}(\lambda) = \mu_f^{\mathrm{geom}}(\lambda)$ für alle $\lambda \in K$.*

Beweis. Zu (i). Der Eigenraum $\mathrm{Eig}_\lambda(f) \subset V$ ist f-invariant nach Lemma 6.2.9(i). Sei $g \in \mathrm{End}_K(\mathrm{Eig}_\lambda(f))$ die Einschränkung von f. Nach Proposition 6.3.35(i) ist χ_f durch χ_g teilbar. Aber $g = \lambda \cdot \mathrm{id}_{\mathrm{Eig}_\lambda(f)}$ und damit ist $\chi_g = (T-\lambda)^{\mu_f^{\mathrm{geom}}(\lambda)}$. Deswegen ist die Vielfachheit von λ in χ_f mindestens $\mu_f^{\mathrm{geom}}(\lambda)$.

Zu (ii). Sei f diagonalisierbar. Nach (i) gilt:

$$\dim_K V = \sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) \leq \sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{alg}}(\lambda) \leq \deg(\chi_f) = \dim_K V.$$

Dabei haben wir auch Satz 6.2.24, Bemerkung 6.3.22 und Proposition 6.3.32(i) verwendet. Daraus folgt

$$\sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{geom}}(\lambda) = \sum_{\lambda \in \sigma(f)} \mu_f^{\mathrm{alg}}(\lambda),$$

und nach (i) ist dies nur möglich, wenn $\mu_f^{\mathrm{alg}}(\lambda) = \mu_f^{\mathrm{geom}}(\lambda)$ für alle $\lambda \in K$. □

Definition 6.3.37 (Einsetzung von Endomorphismen)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Ist $p = \sum_{i=0}^n a_i T^i$ ein Polynom über K, so definieren wir die *Einsetzung* von f in p durch

$$p(f) := \sum_{i=0}^n a_i f^i \in \mathrm{End}_K(V).$$

Man definiert auf ähnliche Weise die Einseztung $p(A) \in M_n(K)$ einer Matrix $A \in M_n(K)$ in p.

Dabei muss man sich daran erinnern, dass f^0 die Identität auf V ist. Ist zum Beispiel $p = \lambda - T$ mit $\lambda \in K$, so ist $p(f) = \lambda \cdot \mathrm{id}_V - f$.

Lemma 6.3.38. *Seien $p, q \in K[T]$ Polynome über K, $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und A eine quadratische Matrix über K.*

(i) *Es ist $(p \cdot q)(f) = p(f) \circ q(f)$.*

(ii) *Es ist $(p \cdot q)(A) = p(A) \cdot q(A)$.*

(iii) *Ist V endlich-dimensional mit einer Basis B, so ist $[p(f)]_B^B = p([f]_B^B)$.*

Beweis. Ist p eine Linearkombination von Polynomen p_i und gilt (i) für jedes p_i, so gilt (i) für p. Deswegen können wir annehmen, dass $p = T^d$ mit einem $d \in \mathbb{N}$. Ist $q = \sum_{i=0}^n b_i T^i$, so gilt

$$(T^d \cdot q)(f) = \sum_{i=0}^n b_i f^{d+i} = f^d \circ \left(\sum_{i=0}^n b_i f^i \right) = f^d \circ q(f),$$

da f^d linear ist. Die dritte Aussage folgt aus Proposition 4.2.39(ii,iii), und die zweite Aussage folgt aus (i) und (iii) (und kann auch leicht direkt überprüft werden). □

Satz 6.3.39 (Satz von Cayley-Hamilton)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus. Dann gilt $\chi_f(f) = 0$ in $\mathrm{End}_K(V)$.*

Beweis. Sei B eine Basis von V und sei $A = [f]_B^B$. Nach Lemma 6.3.38(iii) ist dann $[\chi_f(f)]_B^B = \chi_A(A)$. Es genügt also zu zeigen, dass $\chi_A(A) = 0$ für jede $n \times n$-Matrix A. Nach dem Korollar 5.3.39 existiert eine Matrix $B \in M_n(K[T])$ mit

$$(TI_n - A) \cdot B = \chi_A I_n,$$

nämlich $B = \operatorname{adj}(TI_n - A)$. Nach Definition der adjunkten Matrix sind die Koeffizienten von B Polynome vom Grad $\leq n-1$. Man kann also schreiben $B = \sum_{i=0}^{n-1} T^i B_i$ mit $B_i \in M_n(K)$. Man setzt auch $B_n = 0$ und $B_{-1} = 0$. Dann erhalten wir

$$\chi_A I_n = \sum_{i=0}^{n-1} (T^{i+1} B_i - T^i A B_i) = \sum_{i=0}^{n} T^i (B_{i-1} - A B_i).$$

Ist $\chi_A = \sum_{i=0}^{n} c_i T^i$, so folgt

$$c_i I_n = B_{i-1} - A B_i \quad \text{und daher} \quad c_i A^i = A^i B_{i-1} - A^{i+1} B_i$$

für alle $i \in \{0, \ldots, n\}$. Nimmt man die Summe über i, so erhält man

$$\chi_A(A) = \sum_{i=0}^{n} c_i A^i = \sum_{i=0}^{n} (A^i B_{i-1} - A^{i+1} B_i).$$

Die rechte Seite ist jetzt eine Teleskopsumme, die gleich Null ist, wie gewünscht. □

Bemerkung 6.3.40. Es mag scheinen, dass der Satz von Cayley-Hamilton trivial sein soll, da

$$\chi_A(A) \overset{!?}{=} \det(AI_n - A) = \det(0) = 0.$$

Das Problem mit diesem „Beweis“ ist, dass $\chi_A(A)$ eine $n \times n$-Matrix ist während $\det(AI_n - A)$ ein Skalar ist. Es macht also gar keinen Sinn, sie gleichzusetzen. Nach dem Satz von Cayley-Hamilton gilt eigentlich

$$\chi_A(A) = \det(AI_n - A) \cdot I_n,$$

da beide Seiten die Nullmatrix sind, aber diese Gleichung folgt *nicht* aus allgemeinen Gründen: Ist $p = \sum_{i=0}^{d} T^i C_i \in M_n(K[T])$, so ist im Allgemeinen $\det(p)(A) \neq \det(p(A)) \cdot I_n$.

6.4 Hauptvektoren

Hauptvektoren sind eine Verallgemeinerung von Eigenvektoren, die relevant bei nicht-diagonalisierbaren Endomorphismen sind. Das einfachste Beispiel einer nicht-diagonalisierbaren Matrix ist die Scherungsmatrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

(siehe Beispiel 6.2.31). Die Matrix A hat den einzigen Eigenwert 1 (da $\chi_A = (T-1)^2$), aber der zugehörige Eigenraum ist die Gerade $K \cdot e_1$. Man kann aber bemerken, dass für den Vektor e_2 gilt $(I_2 - A)e_2 = -e_1$, und daher $(I_2 - A)^2 e_2 = 0$.

In diesem Beispiel gibt es einen Eigenwert λ von einem $f \in \operatorname{End}_K(V)$ und einen Vektor $v \in V$, der nicht im Kern von $\lambda \cdot \mathrm{id}_V - f$ liegt, aber der im Kern einer *Potenz* von $\lambda \cdot \mathrm{id}_V - f$ liegt. Es stellt sich heraus, dass ein solcher Vektor v kein Eigenvektor zu einem anderen $\mu \neq \lambda$ sein kann (siehe Lemma 6.4.5(ii)), und deswegen ist die Existenz von v eine Obstruktion zur Diagonalisierbarkeit von f.

Diese Beobachtung führt zum Begriff von *Hauptvektor* zu einem Eigenwert und dem zusammenhängenden Begriff der *Trigonalisierbarkeit*, die wir in diesem Abschnitt untersuchen. Als Konsequenz werden wir auch eine geometrische Interpretation der algebraischen Vielfachheit erhalten (Proposition 6.4.12).

Lemma 6.4.1. *Sei V ein K-Vektorraum und $(U_n)_{n\in\mathbb{N}}$ eine Folge von Untervektorräumen von V, so dass $U_n \subset U_{n+1}$ für alle $n \in \mathbb{N}$. Dann ist die Vereinigung $\bigcup_{n\in\mathbb{N}} U_n$ ein Untervektorraum von V.*

Beweis. Wir verwenden das Kriterium 3.2.8. Es ist klar, dass $\bigcup_{n\in\mathbb{N}} U_n$ nicht leer ist. Seien $u, u' \in \bigcup_{n\in\mathbb{N}} U_n$ und $\lambda \in K$. Nach Definition der Vereinigung existieren $n, n' \in \mathbb{N}$, so dass $u \in U_n$ und $u' \in U_{n'}$. Dann liegen $\lambda \cdot u$ in U_n und $u + u'$ in $U_{\max\{n,n'\}}$. □

Definition 6.4.2 (Hauptraum, Hauptvektor, Stufe)**.** Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $\lambda \in K$.

- Der *Hauptraum* (oder *verallgemeinerte Eigenraum*) zu λ von f ist der Untervektorraum

$$\begin{aligned}\mathrm{Hau}_\lambda(f) &:= \{v \in V \mid \text{es existiert } n \in \mathbb{N} \text{ mit } (\lambda \cdot \mathrm{id}_V - f)^n(v) = 0\} \\ &= \bigcup_{n\in\mathbb{N}} \ker\big((\lambda \cdot \mathrm{id}_V - f)^n\big) \subset V\end{aligned}$$

  (siehe Lemma 6.4.1).

- Ein *Hauptvektor* (oder *verallgemeinerter Eigenvektor*) zu λ von f ist ein Element von $\mathrm{Hau}_\lambda(f) \setminus \{0\}$. Die *Stufe* eines Hauptvektors v zu λ ist das kleinste $n \in \mathbb{N} \setminus \{0\}$ mit $(\lambda \cdot \mathrm{id}_V - f)^n(v) = 0$.

Ist $n \in \mathbb{N}$ und ist A eine $n \times n$-Matrix über K, so bezeichnen wir als *Haupträume* und *Hauptvektoren* von A die Haupträume und Hauptvektoren von $L_A\colon K^n \to K^n$.

Bemerkung 6.4.3. Nach Definition gilt $\mathrm{Eig}_\lambda(f) \subset \mathrm{Hau}_\lambda(f)$, und die Eigenvektoren von f sind genau die Hauptvektoren von f der Stufe 1. Außerdem ist ein Skalar $\lambda \in K$ genau dann Eigenwert von f, wenn $\mathrm{Hau}_\lambda(f) \neq \{0\}$, denn: Ist v ein Hauptvektor zu λ der Stufe n, so ist $(\lambda \cdot \mathrm{id}_V - f)^{n-1}(v)$ ein Eigenvektor zu λ.

Beispiel 6.4.4.

(i) Sei $\lambda \in K$ und sei $A \in M_n(K)$ eine Dreiecksmatrix der Gestalt

$$A = \begin{pmatrix} \lambda & & * \\ & \lambda & \\ & \ddots & \\ 0 & & \lambda \end{pmatrix} \quad \text{bzw.} \quad A = \begin{pmatrix} \lambda & & 0 \\ & \lambda & \\ & \ddots & \\ * & & \lambda \end{pmatrix}.$$

Für die Matrix $\lambda I_n - A$ gilt dann $(\lambda I_n - A)^n = 0$. Deswegen ist $\mathrm{Hau}_\lambda(A) = K^n$, d.h., jeder Vektor $v \in K^n \setminus \{0\}$ ist ein Hauptvektor zu λ von A.

(ii) Sei $D\colon C^\infty(\mathbb{R},\mathbb{R}) \to C^\infty(\mathbb{R},\mathbb{R})$ die Differentiationsabbildung. Dann ist

$$\mathrm{Hau}_0(D) = \{f \in C^\infty(\mathbb{R},\mathbb{R}) \mid \text{es gibt } n \in \mathbb{N} \text{ mit } D^n(f) = 0\} = \mathrm{Poly}(\mathbb{R},\mathbb{R}).$$

Lemma 6.4.5. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V.*

(i) *Für alle $\lambda \in K$ ist der Hauptraum $\mathrm{Hau}_\lambda(f) \subset V$ f-invariant.*

(ii) *Ist $\lambda \neq \mu$, so gilt $\mathrm{Hau}_\lambda(f) \cap \mathrm{Hau}_\mu(f) = \{0\}$.*

Beweis. Zu (i). Aus Proposition 6.1.12(i) folgt, dass alle Untervektorräume $\ker((\lambda\cdot\mathrm{id}_V - f)^n)$ f-invariant sind, und daher auch ihre Vereinigung.

Zu (ii). Sei $v \in \mathrm{Hau}_\lambda(f) \cap \mathrm{Hau}_\mu(f)$. Angenommen ist $v \neq 0$. Sei n die Stufe von v als Hauptvektor zu λ. Dann ist $w = (\lambda \cdot \mathrm{id}_V - f)^{n-1}(v)$ ein Eigenvektor zu λ. Nach (i) gilt auch $w \in \mathrm{Hau}_\mu(f)$. Sei m die Stufe von w als Hauptvektor zu μ. Dann ist $u = (\mu \cdot \mathrm{id}_V - f)^{m-1}(w)$ ein Eigenvektor zu μ, und nach Lemma 6.2.9(i) ist u auch ein Eigenvektor zu λ. Das steht aber im Widerspruch zum Lemma 6.2.9(ii). □

Proposition 6.4.6 (lineare Unabhängigkeit von Hauptvektoren)**.** *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Sei $\Lambda \subset K$ eine Teilmenge bestehend aus Eigenwerten von f, und zu jedem $\lambda \in E$ sei $v_\lambda \in V$ ein Hauptvektor zum Eigenwert λ. Dann ist die Familie $(v_\lambda)_{\lambda\in\Lambda}$ linear unabhängig.*

Beweis. Nach Definition der linearen Unabhängigkeit dürfen wir voraussetzen, dass Λ endlich ist. Dann verwenden wir Induktion über die Mächtigkeit von Λ. Wenn $\Lambda = \varnothing$ ist die Aussage trivial. Sei also $\lambda_0 \in \Lambda$ und sei $\sum_{\lambda\in\Lambda} \mu_\lambda \cdot v_\lambda = 0$ mit $\mu_\lambda \in K$. Es existiert $n \in \mathbb{N}$, so dass $(\lambda_0 \operatorname{id}_V - f)^n(v_{\lambda_0}) = 0$. Dann gilt:

$$0 = (\lambda_0 \operatorname{id}_V - f)^n \left(\sum_{\lambda\in\Lambda} \mu_\lambda \cdot v_\lambda \right) = \sum_{\lambda\in\Lambda\setminus\{\lambda_0\}} \mu_\lambda \cdot (\lambda_0 \operatorname{id}_V - f)^n(v_\lambda).$$

Nach Lemma 6.4.5(i) liegt jeder Vektor $(\lambda_0 \operatorname{id}_V - f)^n(v_\lambda)$ in $\operatorname{Hau}_\lambda(f)$, und nach Lemma 6.4.5(ii) ist er nicht null, sonst wäre $v_\lambda \in \operatorname{Hau}_\lambda(f) \cap \operatorname{Hau}_{\lambda_0}(f) = \{0\}$, aber $v_\lambda \neq 0$ nach Definition von Hauptvektor. Also ist jedes $(\lambda_0 \operatorname{id}_V - f)^n(v_\lambda)$ wieder ein Hauptvektor zu λ. Aus der Induktionsvoraussetzung folgt, dass $\mu_\lambda = 0$ für alle $\lambda \in \Lambda \setminus \{\lambda_0\}$. Dann ist auch $\mu_{\lambda_0} \cdot v_{\lambda_0} = 0$ und damit $\mu_{\lambda_0} = 0$. □

Korollar 6.4.7. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V und sei $\Lambda \subset K$ eine Teilmenge. Dann ist die kanonische Abbildung*

$$\bigoplus_{\lambda\in\Lambda} \operatorname{Hau}_\lambda(f) \to \sum_{\lambda\in\Lambda} \operatorname{Hau}_\lambda(f)$$

ein Isomorphismus.

Beweis. Dies folgt aus Propositionen 6.4.6 und 6.1.7. □

6.4.1 Trigonalisierbarkeit

Der Begriff der Trigonalisierbarkeit erhalten wir, indem wir Eigenvektoren durch Hauptvektoren in der Definition der Diagonalisierbarkeit ersetzt:

Definition 6.4.8 (trigonalisierbarer Endomorphismus)**.** Sei V ein K-Vektorraum. Ein Endomorphismus $f \in \operatorname{End}_K(V)$ heißt *trigonalisierbar*, wenn eine Basis von V bestehend aus Hauptvektoren von f existiert.

Eine quadratische Matrix A heißt *trigonalisierbar*, wenn L_A trigonalisierbar ist.

Beispiel 6.4.9.

(i) Da jeder Eigenvektor ein Hauptvektor ist, ist jeder diagonalisierbare Endomorphismus auch trigonalisierbar.

(ii) Sei $V \neq \{0\}$ ein K-Vektorraum und sei $t\colon V \oplus V \to V \oplus V$, $(v, w) \mapsto (w, v)$. Nach Beispiel 6.2.21(i) ist der Endomorphismus t genau dann diagonalisierbar, wenn $\operatorname{char}(K) \neq 2$. Aber t ist immer trigonalisierbar: Ist $\operatorname{char}(K) = 2$, so sind alle Vektoren $(v, w) \neq (0, 0)$ Hauptvektoren zum Eigenwert 1 von t, denn es gilt $(\operatorname{id} - t)(v, w) = (v - w, w - v)$ und daher $(\operatorname{id} - t)^2(v, w) = (2v - 2w, 2w - 2v) = (0, 0)$.

(iii) Sei $\lambda \in K$ und sei A eine $n \times n$-Dreiecksmatrix über K mit allen Diagonalkoeffizienten gleich λ. Nach Beispiel 6.4.4(i) ist dann $\operatorname{Hau}_\lambda(A) = K^n$ und insbesondere ist A trigonalisierbar. Wir werden später beweisen, dass *alle* Dreiecksmatrizen trigonalisierbar sind (Satz 6.4.14).

Proposition 6.4.10. *Sei $f\colon V \to V$ ein Endomorphismus eines K-Vektorraums V. Die folgenden Aussagen sind äquivalent:*

(i) *f ist trigonalisierbar.*

(ii) *Jeder Vektor $v \in V$ ist eine Linearkombination von Hauptvektoren von f.*

(iii) *Die kanonische lineare Abbildung $\bigoplus_{\lambda\in K} \mathrm{Hau}_\lambda(f) \to V$ ist ein Isomorphismus.*

Beweis. Die Implikation (i) $\Rightarrow$ (ii) ist klar, und die Implikation (ii) $\Rightarrow$ (iii) folgt aus Korollar 6.4.7. Ist die Abbildung $\bigoplus_{\lambda\in K} \mathrm{Hau}_\lambda(f) \to V$ ein Isomorphismus, so erhält man eine Basis von V bestehend aus Hauptvektoren, indem man Basen aller Haupträume $\mathrm{Hau}_\lambda(f)$ zusammensetzt. □

Lemma 6.4.11. *Seien $p \in K[T]$, $a \in K$ und $n \in \mathbb{N}$. Ist $p(a) \neq 0$, so existieren Polynome $u, v \in K[T]$, so dass $u(T-a)^n + vp = 1$.*

Beweis. Wenn $n = 0$ leisten die Polynome $u = 1$ und $v = 0$ das Gewünschte. Nach Satz 6.3.12 existieren $q \in K[T]$ und $r \in K$ mit $p = (T-a)q + r$. Aus $p(a) \neq 0$ folgt $r \neq 0$. Setzt man $u = -r^{-1}q$ und $v = r^{-1}$, so erhält man $u(T-a) + vp = 1$. Ist $n \geq 1$, so hat die n-te Potenz von $u(T-a) + vp$ die Form $u^n(T-a)^n + wp$, und sie ist immer noch gleich 1. □

Proposition 6.4.12 (Haupträume und algebraische Vielfachheit)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus. Für alle $\lambda \in K$ gilt:*

(i) $\dim_K \mathrm{Hau}_\lambda(f) = \mu_f^{\mathrm{alg}}(\lambda)$.

(ii) $\mathrm{Hau}_\lambda(f) = \ker\left((\lambda \cdot \mathrm{id}_V - f)^{\mu_f^{\mathrm{alg}}(\lambda)}\right)$.

Beweis. Sei $m = \mu_f^{\mathrm{alg}}(\lambda)$. Nach Definition der algebraischen Vielfachheit ist $\chi_f = (T-\lambda)^m p$ mit einem $p \in K[T]$, so dass $p(\lambda) \neq 0$. Nach Lemma 6.4.11 gibt es zu jedem $n \in \mathbb{N}$ Polynome $u_n, v_n \in K[T]$, so dass

$$u_n(T-\lambda)^n + v_n p = 1. \tag{6.4.13}$$

Seien $U_n = \ker((f - \lambda\,\mathrm{id})^n)$, $U = U_m$ und $W = \ker p(f)$. Nach Definition ist $\mathrm{Hau}_\lambda(f) = \bigcup_{n\in\mathbb{N}} U_n$. Wir behaupten, dass U und W komplementär in V sind:

- $\mathrm{Hau}_\lambda(f) \cap W = \{0\}$. Sei $x \in U_n \setminus \{0\}$. Nach (6.4.13) und Lemma 6.3.38(i) ist $(v_n(f) \circ p(f))(x) = x$, und insbesondere ist $p(f)(x) \neq 0$, d.h., $x \notin W$.
- $U + W = V$. Sei $x \in V$. Nach (6.4.13) und Lemma 6.3.38(i) ist

  $$x = (v_m(f) \circ p(f))(x) + (u_m(f) \circ (f - \lambda\,\mathrm{id}_V)^m)(x).$$

  Der erste Summand liegt in U, da $(T-\lambda)^m v_m p = v_m \chi_f$ und $\chi_f(f) = 0$ nach dem Satz von Cayley-Hamilton. Genauso liegt der zweite Summand in W.

Daraus folgt die zweite Aussage, denn: Jedes $x \in \mathrm{Hau}_\lambda(f)$ kann als $x = y + z$ mit $y \in U$ und $z \in W$ geschrieben werden. Da $U \subset \mathrm{Hau}_\lambda(f)$ liegt auch z in $\mathrm{Hau}_\lambda(f)$, so dass $z \in \mathrm{Hau}_\lambda(f) \cap W = \{0\}$.

Nach Proposition 6.1.12(i) sind U und W f-invariant. Mit der Proposition 6.3.35 erhalten wir die Zerlegung

$$\chi_f = \chi_{f_U} \cdot \chi_{f_W}.$$

Der Endomorphismus f_U ist trigonalisierbar mit dem einzigen Eigenwert λ. Nach der Implikation (i) $\Rightarrow$ (iv) im Satz 6.4.14 (deren Beweis unabhängig von der aktuellen Proposition ist), gilt $\chi_{f_U} = (T-\lambda)^{\dim_K U}$, und somit $m \geq \dim_K U$. Auf der anderen Seite ist $\chi_{f_W}(\lambda) \neq 0$, weil W keinen Eigenvektor zu λ enthält, und deshalb muss $\chi_{f_U}(f)$ durch $(T-\lambda)^m$ teilbar sein. Daraus folgt $m \leq \dim_K U$, und damit $m = \dim_K U$. □

Satz 6.4.14 (Charakterisierung der Trigonalisierbarkeit)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und sei $f \in \mathrm{End}_K(V)$. Dann sind die folgenden Aussagen äquivalent:*

(i) *f ist trigonalisierbar.*

(ii) *Es existiert eine Basis B von V, so dass $[f]_B^B$ eine obere Dreiecksmatrix ist.*

(iii) *Es existiert eine Basis B von V, so dass $[f]_B^B$ eine untere Dreiecksmatrix ist.*

(iv) *Das charakteristische Polynom χ_f zerfällt in seine Linearfaktoren.*

(v) *Es gilt*

$$\sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{alg}}(\lambda) = \dim_K V.$$

Beweis. Zu (i) $\Rightarrow$ (ii). Wir beweisen die Aussage durch Induktion über $n = \dim_K V$. Sie ist trivial, wenn $n = 0$; sonst existiert nach Bemerkung 6.4.3 ein Eigenvektor $v \in V$ zu einem Eigenwert λ. Da $f(Kv) \subset Kv$ induziert f nach der universellen Eigenschaft des Quotientenvektorraums einen Endomorphismus $\bar{f}$ von V/Kv. Im Quotientenvektorraum V/Kv ist wieder jeder Vektor eine Linearkombination von Hauptvektoren, d.h., $\bar{f}$ ist wieder trigonalisierbar (Proposition 6.4.10). Da $\dim_K(V/Kv) = n-1$ gibt es nach Induktionsvoraussetzung eine Basis $C = (\bar{v}_2, \dots, \bar{v}_n)$ von V/Kv, so dass die Matrix $[\bar{f}]_C^C$ eine obere Dreiecksmatrix ist. Seien $v_2, \dots, v_n \in V$ Urbilder der Vektoren $\bar{v}_2, \dots, \bar{v}_n$. Dann ist $B = (v, v_2, \dots, v_n)$ eine Basis von V, so dass

$$[f]_B^B = \begin{pmatrix} \lambda & * \\ 0 & [\bar{f}]_C^C \end{pmatrix}.$$

Insbesondere ist $[f]_B^B$ eine obere Dreiecksmatrix.

Zu (ii) $\Rightarrow$ (iii). Sei $B = (b_1, \dots, b_n)$ und sei $B' = (b_n, \dots, b_1)$. Ist $[f]_B^B$ eine obere Dreiecksmatrix, so ist $[f]_{B'}^{B'}$ eine untere Dreiecksmatrix.

Zu (iii) $\Rightarrow$ (iv). Dies folgt aus dem Korollar 5.3.34.

Zu (iv) $\Rightarrow$ (v). Dies folgt aus der Definition von μ_f^{alg}.

Zu (v) $\Rightarrow$ (i). Nach Korollar 6.4.7 und Proposition 6.4.12(i) gilt

$$\dim_K \left(\sum_{\lambda\in\sigma(f)} \mathrm{Hau}_\lambda(f) \right) = \dim_K \left(\bigoplus_{\lambda\in\sigma(f)} \mathrm{Hau}_\lambda(f) \right) = \sum_{\lambda\in\sigma(f)} \mu_f^{\mathrm{alg}}(\lambda) = \dim_K V,$$

und damit ist $V = \sum_{\lambda\in\sigma(f)} \mathrm{Hau}_\lambda(f)$. Aus Proposition 6.4.10 folgt nun, dass f trigonalisierbar ist. □

Korollar 6.4.15. *Sei K ein algebraisch abgeschlossener Körper und V ein endlich-dimensionaler K-Vektorraum. Dann ist jeder Endomorphismus $f \in \mathrm{End}_K(V)$ trigonalisierbar.*

Beweis. Nach Korollar 6.3.23 zerfällt χ_f in seine Linearfaktoren. Nach Satz 6.4.14 (v) $\Rightarrow$ (i) ist f trigonalisierbar. □

Korollar 6.4.16 (Trigonalisierbarkeit von Matrizen)**.** *Seien $n \in \mathbb{N}$ und $A \in M_n(K)$. Dann sind die folgenden Aussagen äquivalent:*

(i) *A ist trigonalisierbar.*

(ii) *A ist ähnlich zu einer oberen Dreiecksmatrix.*

(iii) *A ist ähnlich zu einer unteren Dreiecksmatrix.*

Beweis. Dies folgt aus Satz 6.4.14 (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) und Proposition 6.1.19. □

Bemerkung 6.4.17. Um eine Klassifikation von trigonalisierbaren Endomorphismen bis auf Isomorphie zu erhalten, braucht man noch Dreiecksmatrizen bis auf Ähnlichkeit zu klassifizieren. Das werden wir in der Vorlesung *Lineare Algebra II* weiter untersuchen.

Bemerkung 6.4.18. In der Vorlesung *Algebra* wird gezeigt, dass jeder Körper K ein Teilkörper eines algebraisch abgeschlossenen Körpers $\bar{K}$ ist (zum Beispiel: $\mathbb{R} \subset \mathbb{C}$). Nach Korollaren 6.4.15 und 6.4.16 ist dann jede quadratische Matrix über K ähnlich zu einer Dreiecksmatrix über $\bar{K}$.

Rezept 6.4.19 (Test auf Trigonalisierbarkeit)**.** Gegeben seien eine Matrix $A \in M_n(K)$ und ihre Eigenwerte $\lambda_1, \ldots, \lambda_k$. Zu bestimmen ist, ob A trigonalisierbar ist. Wenn $k = n$ ist, ist A sogar diagonalisierbar nach Korollar 6.2.25. Sonst berechnet man die algebraischen Vielfachheiten $\mu_A^{\text{alg}}(\lambda_i)$, indem man χ_A durch $T - \lambda_i$ so oft wie möglich dividiert. Nach Satz 6.4.14 ist die Matrix A genau dann trigonalisierbar, wenn $\sum_{i=1}^{k} \mu_A^{\text{alg}}(\lambda_i) = n$.

Proposition 6.4.20 (Determinante und Spur trigonalisierbarer Endomorphismen)**.** *Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}_K(V)$ ein trigonalisierbarer Endomorphismus. Dann gilt:*

(i) $\det(f) = \prod_{\lambda \in \sigma(f)} \lambda^{\mu_f^{\text{alg}}(\lambda)}$.

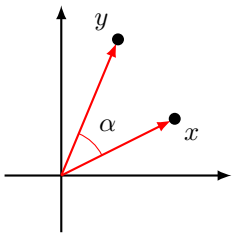
(ii) $\operatorname{tr}(f) = \sum_{\lambda \in \sigma(f)} \mu_f^{\text{alg}}(\lambda) \cdot \lambda$.

Beweis. Sei B eine Basis von V, so dass $[f]_B^B$ eine Dreiecksmatrix ist. Dann ist $\det(f)$ bzw. $\operatorname{tr}(f)$ das Produkt bzw. die Summe der Diagonalkoeffizienten von $[f]_B^B$, die genau die Eigenwerte von f sind, mit algebraischer Vielfachheit gezählt. □

Kapitel 7

Euklidische und unitäre Vektorräume

Seien $x, y \in \mathbb{R}^2$ zwei Vektoren, die wir als Pfeile darstellen:



Nach dem Satz des Pythagoras ist die Länge des Pfeils von x gleich

$$\|x\| := \sqrt{x_1^2 + x_2^2}.$$

Allgemeiner ist der Abstand zwischen x und y gleich $\|x - y\|$. Falls x und y nicht null sind, gibt es auch einen wohldefinierten Winkel $\alpha \in [0, \pi]$ zwischen den Pfeilen von x und y. Mit ein wenig elementarer Trigonometrie kann man leicht nachrechnen, dass

$$\cos \alpha = \frac{x_1 y_1 + x_2 y_2}{\|x\| \|y\|}.$$

Beide Maße (Länge und Winkel) lassen sich durch das sogenannte *Standardskalarprodukt* auf $\mathbb{R}^2$ ausdrücken:

$$\langle -, - \rangle \colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, \quad (x, y) \mapsto \langle x, y \rangle := x_1 y_1 + x_2 y_2.$$

Es gilt nämlich:

$$\|x\| = \sqrt{\langle x, x \rangle} \quad \text{und} \quad \cos \alpha = \frac{\langle x, y \rangle}{\|x\| \|y\|}.$$

Die Abbildung $\langle -, - \rangle$ ist nicht linear, sondern *bilinear* (Definition 5.3.11). Da der Zielbereich gleich dem Grundkörper $\mathbb{R}$ ist, spricht man in diesem Fall von einer *Bilinearform* auf $\mathbb{R}^2$. Zudem hat die Bilinearform $\langle -, - \rangle$ die wichtige Eigenschaft, dass das Selbstprodukt $\langle x, x \rangle$ immer nichtnegativ ist, damit man seine Quadratwurzel ziehen darf.

Ein *Skalarprodukt* auf einem $\mathbb{R}$-Vektorraum V ist eine Abbildung $V \times V \to \mathbb{R}$, die ähnliche Eigenschaften wie das Standardskalarprodukt auf $\mathbb{R}^2$ besitzt. Aus einem Skalarprodukt kann man insbesondere die Begriffe von Länge und Winkel auch in höherer Dimension (und sogar in unendlich-dimensionalen Vektorräumen) vernünftig definieren. Ein $\mathbb{R}$-Vektorraum mit einem Skalarprodukt heißt *euklidischer Vektorraum*.

In vielen Anwendungen braucht man auch solche Begriffe für *komplexe* Vektorräume. Hierzu muss man berücksichtigen, dass komplexe Zahlen selbst als zweidimensionale Vektoren aufgefasst werden können. Im einfachsten Fall des eindimensionalen $\mathbb{C}$-Vektorraums $\mathbb{C}$ ist das Standardskalarprodukt wie folgt definiert:

$$\langle -, - \rangle \colon \mathbb{C} \times \mathbb{C} \to \mathbb{C}, \quad (z, w) \mapsto \langle z, w \rangle := \bar{z} w.$$

Das Selbstprodukt $\langle z, z \rangle$ ist dann immer *reell* und nichtnegativ, und es gilt $|z| = \sqrt{\langle z, z \rangle}$. In diesem Fall ist die Abbildung $\langle -, - \rangle$ nicht einmal bilinear, denn $(\overline{\lambda z})w = \bar{\lambda}(\bar{z}w) \neq \lambda(\bar{z}w)$; sie ist linear in ihrem zweiten Argument aber „linear bis auf komplexe Konjugation“ in ihrem ersten Argument. Eine solche Abbildung $V \times V \to \mathbb{C}$ mit einem $\mathbb{C}$-Vektorraum V heißt *Sesquilinearform* auf V. Ein $\mathbb{C}$-Vektorraum mit einem sesquilinearen Skalarprodukt heißt *unitärer Vektorraum*.

In diesem Kapitel untersuchen wir zunächst Bilinearformen und Sesquilinearformen über beliebigen Körpern. Dann führen wir Skalarprodukte auf $\mathbb{R}$- und $\mathbb{C}$-Vektorräumen ein. Schließlich beweisen wir zwei wichtige Sätze der linearen Algebra: den *Spektralsatz* (Satz 7.3.7) und den *Trägheitssatz von Sylvester* (Satz 7.3.19). Der erste sagt unter anderem, dass symmetrische Matrizen über $\mathbb{R}$ diagonalisierbar sind. Der zweite ist eine vollständige Klassifikation von symmetrischen Bilinearformen auf endlich-dimensionalen $\mathbb{R}$-Vektorräumen.

7.1 Bilinearformen

Seien $V_1, \ldots, V_n, W$ Vektorräume über K. In der Definition 5.3.11 haben wir schon *n-lineare Abbildungen*

$$V_1 \times \cdots \times V_n \to W$$

definiert. Wenn der Zielbereich W gleich dem Grundkörper K ist, nennt man üblicherweise eine solche Abbildung eine *Form*. Im Fall $n = 1$ erhalten wir den Begriff der *Linearform*, der mit dem Begriff des *Dualraums* zusammenhängt (Definition 4.1.47). In diesem Abschnitt wollen wir den Fall $n = 2$ näher betrachten:

Definition 7.1.1 (Bilinearform)**.** Sei K ein Körper und seien V, W Vektorräume über K. Eine Abbildung

$$b \colon V \times W \to K$$

heißt *Bilinearform*, wenn sie bilinear im Sinne der Definition 5.3.11 ist, d.h.:

(i) Für alle $v, v' \in V$, $w \in W$ und $\lambda \in K$ gilt:

$$b(v + v', w) = b(v, w) + b(v', w) \quad \text{und} \quad b(\lambda v, w) = \lambda b(v, w).$$

(ii) Für alle $v \in V$, $w, w' \in W$ und $\mu \in K$ gilt:

$$b(v, w + w') = b(v, w) + b(v, w') \quad \text{und} \quad b(v, \mu w) = \mu b(v, w).$$

Wenn $V = W$ spricht man von einer Bilinearform *auf* V.

Beispiel 7.1.2 (Bilinearformen aus Matrizen)**.** Seien $m, n \in \mathbb{N}$ und $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K. Die Matrix A induziert eine Bilinearform

$$b_A \colon K^m \times K^n \to K, \quad b_A(x, y) = x^\intercal \cdot A \cdot y = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j.$$

Die Matrix A ist durch die Form b_A bestimmt, denn es gilt $a_{ij} = b_A(e_i, e_j)$. Außerdem folgt aus Proposition 7.1.12(i), dass jede Bilinearform $K^m \times K^n \to K$ gleich b_A ist, mit einer geeigneten Matrix A.

Notation 7.1.3. Die Menge aller Bilinearformen $V \times W \to K$ wird mit $\mathrm{Bil}_K(V,W)$ bezeichnet. Wenn $V = W$ schreibt man auch $\mathrm{Bil}_K(V)$. Nach Bemerkung 5.3.15 ist $\mathrm{Bil}_K(V,W)$ ein Untervektorraum von $\mathrm{Abb}(V \times W, K)$, d.h.: Sind $b, b' \in \mathrm{Bil}_K(V,W)$ und ist $\lambda \in K$, so sind die Summe $b + b'$ und das Skalarvielfache λb wieder Bilinearformen $V \times W \to K$.

Eine Bilinearform $b\colon V \times W \to K$ induziert lineare Abbildungen

$$\begin{aligned} b_l &\colon V \to W^*, & b_r &\colon W \to V^*, \\ v &\mapsto b(v,-), & w &\mapsto b(-,w). \end{aligned}$$

Umgekehrt ist b durch entweder b_l oder b_r bestimmt, denn:

$$b(v,w) = b_l(v)(w) = b_r(w)(v).$$

Definition 7.1.4 (nicht ausgeartet, perfekte Paarung)**.** Sei $b\colon V \times W \to K$ eine Bilinearform.

- b heißt *nicht ausgeartet*, wenn die linearen Abbildungen b_l und b_r injektiv sind.
- b heißt *perfekte Paarung*, wenn die linearen Abbildungen b_l und b_r bijektiv sind.

Bemerkung 7.1.5. Sei $b\colon V \times W \to K$ eine nicht ausgeartete Bilinearform. Ist V oder W endlich-dimensional, so folgt aus Proposition 4.1.56(ii) und Korollar 4.1.39, dass $\dim_K(V) = \dim_K(W)$ und dass b eine perfekte Paarung ist. Ist umgekehrt b eine perfekte Paarung, so sind V und W endlich-dimensional (derselben Dimension) nach Bemerkung 4.1.61.

Definition 7.1.6 (symmetrische Bilinearform)**.** Sei V ein K-Vektorraum. Eine Bilinearform b auf V heißt *symmetrisch*, wenn für alle $v, w \in V$ gilt:

$$b(v,w) = b(w,v).$$

Bemerkung 7.1.7. Ein Bilinearform b auf V ist genau dann symmetrisch, wenn $b_l = b_r$.

Beispiel 7.1.8.

(i) Sei $a \in K$. Die Abbildung

$$K \times K \to K, \quad (x,y) \mapsto axy,$$

ist eine symmetrische Bilinearform auf K. Sie ist genau dann nicht ausgeartet, wenn $a \neq 0$.

(ii) Die Bilinearform

$$e\colon V \times V^* \to K, \quad (v,\alpha) \mapsto \alpha(v),$$

ist nicht ausgeartet: Die zugehörigen linearen Abbildungen e_l und e_r sind $\mathrm{ev}\colon V \to V^{**}$, die nach Proposition 4.1.59 injektiv ist, und id_{V^*}. Ist V unendlich-dimensional, so ist aber ev nicht surjektiv, und damit ist e keine perfekte Paarung.

(iii) Sei $n \in \mathbb{N}$. Die Bilinearform

$$b\colon K^n \times K^n \to K, \quad (x,y) \mapsto \sum_{i=1}^{n} x_i y_i,$$

ist eine symmetrische perfekte Paarung. Die linearen Abbildungen $b_l, b_r\colon K^n \to (K^n)^*$ sind beide gleich dem Isomorphismus ε_E aus Proposition 4.1.56, wobei E die Standardbasis von K^n ist.

(iv) Sei allgemeiner I eine Menge. Dann ist die symmetrische Bilinearform

$$b\colon K^{(I)} \times K^{(I)} \to K, \quad ((x_i)_{i\in I}, (y_i)_{i\in I}) \mapsto \sum_{i\in I} x_i y_i,$$

nicht ausgeartet, und sie ist genau dann eine perfekte Paarung, wenn I endlich ist.

Bemerkung 7.1.9 (quadratische Formen)**.** Sei V ein K-Vektorraum. Eine *quadratische Form* auf V ist eine Abbildung $q\colon V \to K$, so dass:

(i) Es ist $q(\lambda v) = \lambda^2 q(v)$ für alle $v \in V$ und $\lambda \in K$.

(ii) Die Abbildung $b_q\colon V \times V \to K$, $b_q(v, w) = q(v + w) - q(v) - q(w)$, ist bilinear.

Man beachte dabei, dass die Bilinearform b_q symmetrisch ist. Wenn $\mathrm{char}(K) \neq 2$ ist die Abbildung $q \mapsto b_q$ eine Bijektion zwischen quadratischen Formen auf V und symmetrischen Bilinearformen auf V: Die Umkehrabbildung bildet b auf die quadratische Form $v \mapsto \frac{1}{2}b(v, v)$ ab. Wenn $\mathrm{char}(K) = 2$ sind aber quadratische Formen und symmetrische Bilinearformen ganz unterschiedliche Begriffe (die beide interessant sind).

7.1.1 Darstellung von Bilinearformen

Wie bei linearen Abbildungen kann man auch Bilinearformen mit Matrizen darstellen.

Definition 7.1.10 (Darstellungsmatrix einer Bilinearform)**.** Seien V und W endlich-dimensionale Vektorräume über K mit Basen $B = (v_1, \ldots, v_m)$ und $C = (w_1, \ldots, w_n)$ und sei $b\colon V \times W \to K$ eine Bilinearform. Die $m \times n$-Matrix

$$[b]_{B,C} := (b(v_i, w_j))_{i,j} \in M_{m\times n}(K)$$

heißt die *Darstellungsmatrix* von b bzgl. der Basen B und C.

Beispiel 7.1.11. Sei $A \in M_{m\times n}(K)$. Für die Bilinearform $b_A \in \mathrm{Bil}_K(K^m, K^n)$ aus Beispiel 7.1.2 gilt $[b_A]_{E_m,E_n} = A$, wobei E_m bzw. E_n die Standardbasis von K^m bzw. K^n ist.

Proposition 7.1.12 (Eigenschaften der Darstellungsmatrix)**.** *Seien V und W endlich-dimensionale Vektorräume über K mit Basen $B = (v_1, \ldots, v_m)$ und $C = (w_1, \ldots, w_n)$.*

(i) *Die Abbildung*

$$\begin{aligned}\mathrm{Bil}_K(V, W) &\to M_{m\times n}(K),\\ b &\mapsto [b]_{B,C}\end{aligned}$$

ist ein Isomorphismus von K-Vektorräumen. Die Umkehrabbildung schickt $A \in M_{m\times n}(K)$ auf die Bilinearform

$$\begin{aligned}V \times W &\to K,\\ (v, w) &\mapsto [v]_B^{\intercal} \cdot A \cdot [w]_C.\end{aligned}$$

(ii) *Sei $b\colon V \times W \to K$ eine Bilinearform, $v \in V$ und $w \in W$. Dann gilt:*

$$b(v, w) = [v]_B^{\intercal} \cdot [b]_{B,C} \cdot [w]_C.$$

(iii) *Seien B^* und C^* die dualen Basen zu B und C. Ist $b\colon V \times W \to K$ eine Bilinearform, so gilt:*

$$[b]_{B,C} = [b_r]_{B^*}^{C} = ([b_l]_{C^*}^{B})^{\intercal}.$$

(iv) *Eine Bilinearform $b\colon V \times W \to K$ ist genau dann nicht ausgeartet, wenn $m = n$ und die Matrix $[b]_{B,C}$ invertierbar ist.*

Beweis. Zu (i). Dass die Abbildung $b \mapsto [b]_{B,C}$ linear ist folgt unmittelbar aus der Definition von $[b]_{B,C}$. Es folgt aus (ii), dass die Komposition

$$\mathrm{Bil}_K(V,W) \to M_{m\times n}(K) \to \mathrm{Bil}_K(V,W)$$

der gegebenen Abbildungen gleich der Identität ist. Sei umgekehrt $A \in M_{m\times n}(K)$. Dann ist

$$[v_i]_B^\intercal \cdot A \cdot [w_j]_C = e_i^\intercal \cdot A \cdot e_j = A_{ij},$$

und somit ist die umgekehrte Komposition auch die Identität auf $M_{m\times n}(K)$.

Zu (ii). Die Formel gilt nach Definition von $[b]_{B,C}$, wenn v und w aus den Basen B und C ausgewählt werden. Beliebige v und w sind Linearkombinationen dieser Basisvektoren. Da beide Seiten bilinear in (v,w) sind, gilt die Formel im Allgemeinen.

Zu (iii). Die Gleichung $[b]_{B,C} = [b_r]_{B^*}^C$ bedeutet, dass

$$b_r(w_j) = \sum_{k=1}^m b(v_k, w_j) \cdot v_k^*$$

im Dualraum V^*. Dies kann man durch Auswertung in den Basisvektoren v_i prüfen. Nach Definition der dualen Basis ist $v_k^*(v_i) = \delta_{ik}$, so dass

$$\left(\sum_{k=1}^m b(v_k, w_j) \cdot v_k^*\right)(v_i) = \sum_{k=1}^m b(v_k, w_j)\delta_{ik} = b(v_i, w_j) = b_r(w_j)(v_i),$$

wie gewünscht. Der Beweis der Gleichung $[b]_{B,C} = ([b_l]_{C^*}^B)^\intercal$ ist ähnlich.

Zu (iv). Falls $m \neq n$ ist b auf keinen Fall nicht ausgeartet (Bemerkung 7.1.5). Falls $m = n$ folgt die Aussage aus (iii): Die Abbildungen b_l und b_r sind genau dann bijektiv, wenn die Darstellungsmatrizen $[b_l]_{C^*}^B$ und $[b_r]_{B^*}^C$ invertierbar sind. □

Seien B, B' zwei Basen eines K-Vektorraums V der endlichen Dimension n. Zur Erinnerung (Definition 4.2.41) ist die *Basiswechselmatrix* $T_{B'}^B \in \mathrm{GL}_n(K)$ von B nach B' die Darstellungsmatrix der Identität bezüglich dieser Basen:

$$T_{B'}^B := [\mathrm{id}_V]_{B'}^B \in \mathrm{GL}_n(K).$$

Die Spalten von $T_{B'}^B$ sind konkret die Koordinatenvektoren der Vektoren aus B bezüglich der Basis B'.

Proposition 7.1.13 (Basiswechselformel für Bilinearformen)**.** *Seien V und W endlich-dimensionale Vektorräume über K, B, B' Basen von V und C, C' Basen von W. Für alle Bilinearformen $b\colon V \times W \to K$ gilt:*

$$[b]_{B',C'} = (T_B^{B'})^\intercal \cdot [b]_{B,C} \cdot T_C^{C'}.$$

Beweis. Seien $v \in V$ und $w \in W$. Nach Proposition 7.1.12(ii) gilt:

$$\begin{aligned} b(v,w) &= [v]_B^\intercal \cdot [b]_{B,C} \cdot [w]_C \\ &= (T_B^{B'} \cdot [v]_{B'})^\intercal \cdot [b]_{B,C} \cdot (T_C^{C'} \cdot [w]_{C'}) \\ &= [v]_{B'}^\intercal \cdot ((T_B^{B'})^\intercal \cdot [b]_{B,C} \cdot T_C^{C'}) \cdot [w]_{C'}. \end{aligned}$$

Nach Proposition 7.1.12(i) ist $(T_B^{B'})^\intercal \cdot [b]_{B,C} \cdot T_C^{C'}$ die Darstellungsmatrix von b bzgl. der Basen B' und C', wie gewünscht. □

Beispiel 7.1.14. Sei

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} \in M_2(\mathbb{R}).$$

Die zugehörige Bilinearform b_A auf $\mathbb{R}^2$ ist

$$b_A(x, y) = 5x_1y_1 - 2x_1y_2 - 2x_2y_1 + x_2y_2.$$

Sei E die Standardbasis von $\mathbb{R}^2$ und sei B die Basis

$$B = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right), \quad \text{so dass} \quad T_E^B = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}.$$

Dann gilt $[b_A]_{E,E} = A$ und somit

$$[b_A]_{B,B} = (T_E^B)^\intercal \cdot A \cdot T_E^B = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Die Form b_A sieht also besonders einfach aus, wenn man die Basis B als Koordinatensystem benutzt: Nach Proposition 7.1.12(ii) gilt nämlich

$$b_A(x, y) = [x]_B^\intercal \cdot [y]_B \quad \text{für alle } x, y \in \mathbb{R}^2.$$

Wie man eine solche Basis B finden kann, wird später im Abschnitt 7.3 erklärt.

Definition 7.1.15 (symmetrische Matrix)**.** Eine quadratische Matrix A über K heißt *symmetrisch*, wenn $A^\intercal = A$.

Proposition 7.1.16 (Symmetriekriterium)**.** *Sei V ein endlich-dimensionaler Vektorraum über K und sei B eine Basis von V. Eine Bilinearform b auf V ist genau dann symmetrisch, wenn die Matrix $[b]_{B,B}$ symmetrisch ist.*

Beweis. Sei $B = (v_1, \ldots, v_n)$, so dass $[b]_{B,B} = (b(v_i, v_j))_{i,j}$. Ist b symmetrisch, so gilt insbesondere $b(v_i, v_j) = b(v_j, v_i)$ für alle i, j, d.h., $[b]_{B,B}$ ist symmetrisch. Gilt umgekehrt die Gleichung $b(v, w) = b(w, v)$ für alle v, w aus der Basis B, so gilt sie für alle $v, w \in V$ wegen der Bilinearität von b. □

7.1.2 Sesquilinearformen

Definition 7.1.17 (Körperinvolution, Fixkörper, Körper mit Involution)**.** Sei K ein Körper. Ein *Körperinvolution* auf K ist eine Abbildung $\sigma\colon K \to K$ mit folgenden Eigenschaften:

(i) σ ist eine *Involution*, d.h., $\sigma \circ \sigma = \mathrm{id}_K$.

(ii) Für alle $x, y \in K$ gilt:

$$\sigma(x + y) = \sigma(x) + \sigma(y) \quad \text{und} \quad \sigma(x \cdot y) = \sigma(x) \cdot \sigma(y).$$

Der *Fixkörper* von σ ist

$$K^\sigma := \{x \in K \mid \sigma(x) = x\}.$$

Ein *Körper mit Involution* ist ein Paar (K, σ) bestehend aus einem Körper K und einer Körperinvolution σ auf K.

Bemerkung 7.1.18. Ein Körperinvolution erfüllt $\sigma(0) = 0$ und $\sigma(1) = 1$: Wegen der Surjektivität von σ sind $\sigma(0)$ und $\sigma(1)$ wieder neutrale Elemente bzgl. der Addition und der Multiplikation, und die sind eindeutig (siehe Proposition 2.1.3(i)). Insbesondere ist σ ein selbstinverser Körperhomomorphismus im Sinne der Bemerkung 4.1.2. Daraus folgt leicht, dass der Fixkörper K^σ ein Teilkörper von K ist.

Beispiel 7.1.19.

(i) Das wichtigste Beispiel einer Körperinvolution ist die komplexe Konjugation $z \mapsto \bar{z}$ auf dem Körper $\mathbb{C}$. Der Fixkörper ist gleich $\mathbb{R}$. Wenn man $\mathbb{C}$ als Körper mit Involution betrachtet, ist die Involution immer die komplexe Konjugation, sofern nicht anders angegeben.

(ii) Die komplexe Konjugation schränkt sich zu einer Körperinvolution auf den rationalen komplexen Zahlen $\mathbb{Q}(i)$ (Beispiel 2.4.4), deren Fixkörper $\mathbb{Q}$ ist.

(iii) Die einzige Körperinvolution auf $\mathbb{Q}$ ist die Identität, da $\mathbb{Q}$ der einzige Teilkörper von $\mathbb{Q}$ ist. Es zeigt sich, dass die Identität auch die einzige Körperinvolution auf $\mathbb{R}$ ist.

(iv) Die folgende Abbildung σ ist eine Körperinvolution auf dem Körper $\mathbb{F}_4$ (siehe Bemerkung 2.4.11):

$$\sigma(0) = 0, \quad \sigma(1) = 1, \quad \sigma(\alpha) = \beta, \quad \sigma(\beta) = \alpha.$$

Ihr Fixkörper ist $\mathbb{F}_2$. Allgemeiner gibt es eine eindeutige Körperinvolution auf dem endlichen Körper $\mathbb{F}_{q^2}$ mit Fixkörper $\mathbb{F}_q$.

Bemerkung 7.1.20. Sei (K, σ) ein Körper mit Involution. Für alle $x \in K$ liegen $x + \sigma(x)$ und $x \cdot \sigma(x)$ im Fixkörper K^σ, denn:

$$\sigma(x + \sigma(x)) = \sigma(x) + \sigma^2(x) = \sigma(x) + x,$$

und genauso mit der Multiplikation. Zum Beispiel: Für jede komplexe Zahl z sind bekanntlich beide $z + \bar{z} = 2 \operatorname{Re} z$ und $z \cdot \bar{z} = |z|^2$ reell.

Definition 7.1.21 (semilineare Abbildung)**.** Sei (K, σ) ein Körper mit Involution und seien V, W Vektorräume über K. Eine Abbildung $f \colon V \to W$ heißt *semilinear*, oder genauer *σ-semilinear*, wenn folgendes gilt:

(i) Für alle $v, v' \in V$ gilt:

$$f(v + v') = f(v) + f(v').$$

(ii) Für alle $v \in V$ und $\lambda \in K$ gilt:

$$f(\lambda \cdot v) = \sigma(\lambda) \cdot f(v).$$

Eine semilineare Abbildung heißt auch *antilinear* oder *konjugiert linear*.

Beispiel 7.1.22. Die Abbildung $\sigma \colon K \to K$ selbst ist semilinear (und nicht linear, außer wenn $\sigma = \mathrm{id}_K$). Allgemeiner ist die Abbildung

$$K^n \to K^n, \quad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} \sigma(x_1) \\ \vdots \\ \sigma(x_n) \end{pmatrix},$$

semilinear.

Bemerkung 7.1.23. Eine σ-semilineare Abbildung ist insbesondere K^σ-linear. Bezüglich Komposition verhalten sich lineare und semilineare Abbildungen wie gerade und ungerade Zahlen bezüglich Addition: Die Komposition einer semilinearen Abbildung mit einer linearen Abbildung ist wieder semilinear, und die Komposition zweier semilinearen Abbildungen ist linear (da $\sigma^2 = \mathrm{id}_K$).

Definition 7.1.24 (Sesquilinearform)**.** Sei (K, σ) ein Körper mit Involution und seien V, W Vektorräume über K. Eine Abbildung

$$s\colon V \times W \to K$$

heißt *Sesquilinearform*, wenn sie semilinear in ihrem ersten Argument und linear in ihrem zweiten Argument ist, d.h.:

(i) Für alle $v, v' \in V$, $w \in W$ und $\lambda \in K$ gilt:

$$s(v + v', w) = s(v, w) + s(v', w) \quad \text{und} \quad s(\lambda v, w) = \sigma(\lambda) s(v, w).$$

(ii) Für alle $v \in V$, $w, w' \in W$ und $\mu \in K$ gilt:

$$s(v, w + w') = s(v, w) + s(v, w') \quad \text{und} \quad s(v, \mu w) = \mu s(v, w).$$

Wenn $V = W$ spricht man von einer Sesquilinearform *auf* V.

Beispiel 7.1.25 (Sesquilinearformen aus Matrizen)**.** Sei (K, σ) ein Körper mit Involution und sei $A = (a_{ij})_{i,j}$ eine $m \times n$-Matrix über K. Dann ist die folgende Abbildung eine Sesquilinearform:

$$s_A\colon K^m \times K^n \to K, \quad s_A(x, y) = x^{\mathsf{H}} \cdot A \cdot y = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \sigma(x_i) y_j.$$

Dabei ist x^{H} der Zeilenvektor $\begin{pmatrix} \sigma(x_1) & \dots & \sigma(x_m) \end{pmatrix}$, siehe Definition 7.1.33. Wenn $\sigma = \mathrm{id}_K$ gewinnen wir die Bilinearform b_A aus Beispiel 7.1.2 zurück.

Notation 7.1.26. Die Menge aller Sesquilinearformen $V \times W \to K$ wird mit $\mathrm{Sil}_K(V, W)$ bezeichnet. Wenn $V = W$ schreibt man auch $\mathrm{Sil}_K(V)$. Man kann leicht nachprüfen, dass $\mathrm{Sil}_K(V, W)$ ein Untervektorraum von $\mathrm{Abb}(V \times W, K)$ ist.

Bemerkung 7.1.27. Die lateinischen Präfixe *semi-* und *sesqui-* bedeuten „halb“ und „eineinhalb“. Die Bemerkung 7.1.23 erklärt, in welchem Sinne semilineare Abbildungen „halblinear“ sind.

Im endlich-dimensionalen Fall möchten wir auch Sesquilinearformen mit Matrizen darstellen. Dazu kann man beobachten, dass eine Sesquilinearform $V \times W \to K$ als Bilinearform aufgefasst werden kann, wenn man die Vektorraumstruktur von V in geeigneter Weise anpasst:

Definition 7.1.28 (konjugierter Vektorraum, adjungierter Vektorraum)**.** Sei (K, σ) ein Körper mit Involution und V ein K-Vektorraum.

- Der *konjugierte Vektorraum* $\bar{V}$ zu V ist der K-Vektorraum mit derselben unterliegenden abelschen Gruppe $(V, +)$ und mit der Skalarmultiplikation

  $$(\lambda, v) \mapsto \sigma(\lambda) \cdot v.$$

  Ist $f\colon V \to W$ eine lineare Abbildung, so ist f auch eine lineare Abbildung von $\bar{V}$ nach $\bar{W}$, die man auch mit $\bar{f}\colon \bar{V} \to \bar{W}$ bezeichnet.

- Der *adjungierte Vektorraum* $V^\dagger$ zu V ist der Dualraum von $\bar{V}$. Ist $f\colon V \to W$ eine lineare Abbildung, so schreibt man $f^\dagger\colon W^\dagger \to V^\dagger$ für die duale Abbildung zu $\bar{f}$.

Bemerkung 7.1.29.

(i) Dass $\bar{V}$ mit den obigen Verknüpfungen ein K-Vektorraum ist, folgt unmittelbar daraus, dass σ ein Körperhomomorphismus ist. Aus $\sigma^2 = \mathrm{id}_K$ folgt außerdem, dass $\bar{\bar{V}} = V$.

(ii) Ist $B = (v_i)_{i\in I}$ eine Basis von V, so ist B auch eine Basis von $\bar{V}$. Man beachte dabei, dass der Koordinatenvektor $[v]_B \in K^{(I)}$ hängt davon ab, ob wir v als Vektor von V oder von $\bar{V}$ betrachten. Ist $(\lambda_i)_{i\in I}$ der Koordinatorvektor von v als Vektor von V, so ist $(\sigma(\lambda_i))_{i\in I}$ der Koordinatenvektor von v als Vektor von $\bar{V}$.

Mit der Definition 7.1.28 ist eine semilineare Abbildung $V \to W$ das Gleiche wie eine lineare Abbildung $\bar{V} \to W$ (oder $V \to \bar{W}$). Insbesondere gilt:

$$V^\dagger = \{\text{semilineare Abbildungen } V \to K\}.$$

Zudem ist eine Sesquilinearform $V \times W \to K$ das Gleiche wie eine Bilinearform $\bar{V} \times W \to K$, das heißt:

$$\mathrm{Sil}_K(V, W) = \mathrm{Bil}_K(\bar{V}, W).$$

Insbesondere induziert ein $s \in \mathrm{Sil}_K(V, W)$ die linearen Abbildungen

$$\begin{aligned} s_l \colon \bar{V} &\to W^*, & s_r \colon W &\to V^\dagger, \\ v &\mapsto s(v, -), & w &\mapsto s(-, w). \end{aligned}$$

Sind V und W endlich-dimensional mit Basen B und C, so erhalten wir die Darstellungsmatrix $[s]_{B,C}$ einer Sesquilinearform $s \colon V \times W \to K$, indem wir s als Bilinearform $\bar{V} \times W \to K$ betrachten. Alle Resultate vom Abschnitt 7.1.1 können wir nun auf Sesquilinearformen anwenden, indem man V durch $\bar{V}$ ersetzt. Zum Beispiel gibt es einen Isomorphismus

$$\mathrm{Sil}_K(V, W) \xrightarrow{\sim} M_{m\times n}(K), \quad s \mapsto [s]_{B,C} = (s(v_i, w_j))_{i,j},$$

wobei $B = (v_1, \ldots, v_m)$ und $C = (w_1, \ldots, w_n)$ (Proposition 7.1.12(i)).

Definition 7.1.30 (hermitesche Form)**.** Sei (K, σ) ein Körper mit Involution und sei V ein Vektorraum über K. Eine Sesquilinearform s auf V heißt *hermitesch* oder eine *hermitesche Form*, wenn für alle $v, w \in V$ gilt:

$$s(v, w) = \sigma(s(w, v)).$$

Bemerkung 7.1.31. Für eine hermitesche Form s auf V und einen Vektor $v \in V$ gilt insbesondere $s(v, v) = \sigma(s(v, v))$, d.h., $s(v, v) \in K^\sigma$. Wenn $\sigma = \mathrm{id}_K$ ist eine hermitesche Form das Gleiche wie eine symmetrische Bilinearform.

Beispiel 7.1.32.

(i) Sei $a \in K$. Die Sesquilinearform $K \times K \to K$, $(x, y) \mapsto a\sigma(x)y$, ist genau dann hermitesch, wenn $a \in K^\sigma$.

(ii) Sei $n \in \mathbb{N}$. Die Abbildung

$$s \colon K^n \times K^n \to K, \quad (x, y) \mapsto \sum_{i=1}^{n} \sigma(x_i)y_i,$$

ist eine hermitesche Sesquilinearform auf K^n.

Definition 7.1.33 (konjugierte Matrix, adjungierte Matrix)**.** Sei (K, σ) ein Körper mit Involution, seien $m, n \in \mathbb{N}$ und sei $A \in M_{m\times n}(K)$.

- Die *konjugierte Matrix* zu A ist die $m \times n$-Matrix $\bar{A}$ mit $\bar{A}_{ij} = \sigma(A_{ij})$.
- Die *adjungierte Matrix* (oder *transponiert-konjugierte Matrix*, oder *hermitesch transponierte Matrix*) zu A ist die $n \times m$-Matrix

$$A^{\mathsf{H}} := (\bar{A})^{\mathsf{T}} = \overline{A^{\mathsf{T}}}.$$

Bemerkung 7.1.34. Die Abbildungen

$$\begin{aligned} & M_{m\times n}(K) \to M_{m\times n}(K), \quad A \mapsto \bar{A}, \\ \text{und} \quad & M_{m\times n}(K) \to M_{n\times m}(K), \quad A \mapsto A^{\mathsf{H}}, \end{aligned}$$

sind nicht linear (außer wenn $\sigma = \mathrm{id}_K$ oder $m = n = 0$) sondern semilinear. Sie sind außerdem mit der Matrixmultiplikation kompatibel: Sind $A \in M_{m\times n}(K)$ und $B \in M_{n\times p}(K)$, so gilt

$$\overline{AB} = \bar{A}\bar{B} \quad \text{und} \quad (AB)^{\mathsf{H}} = B^{\mathsf{H}}A^{\mathsf{H}}.$$

Die erste Gleichung folgt unmittelbar daraus, dass σ ein Körperhomomorphismus ist. Die zweite Gleichung folgt aus der ersten und $(AB)^{\mathsf{T}} = B^{\mathsf{T}}A^{\mathsf{T}}$ (Proposition 4.2.16(iv)).

Proposition 7.1.35 (Darstellungsmatrizen der konjugierten/adjungierten Abbildung)**.** *Sei (K, σ) ein Körper mit Involution und seien V und W endlich-dimensionale K-Vektorräume mit Basen B und C.*

(i) *Für alle linearen Abbildungen $f\colon V \to W$ gilt:*

$$[\bar{f}]_C^B = \overline{[f]_C^B}.$$

(ii) *Für alle linearen Abbildungen $f\colon V \to W$ gilt:*

$$[f^\dagger]_{B^\dagger}^{C^\dagger} = ([f]_C^B)^{\mathsf{H}},$$

wobei $B^\dagger$ die duale Basis zu der Basis B von $\bar{V}$ ist.

Beweis. Zu (i). Seien $B = (v_1, \dots, v_n)$, $C = (w_1, \dots, w_m)$, und $[f]_C^B = (a_{ij})_{i,j}$. Nach Definition der Darstellungsmatrix gilt:

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Aber das Skalarvielfache $a_{ij}w_i$ in W ist das Skalarvielfache $\sigma(a_{ij})w_i$ in $\bar{W}$ (da $\sigma^2 = \mathrm{id}_K$). Damit ist die Darstellungsmatrix von $\bar{f}$ die konjugierte Matrix $(\sigma(a_{ij}))_{i,j}$.

Zu (ii). Man kombiniert (i) und Proposition 4.2.47. □

Bemerkung 7.1.36. Seien V, W endlich-dimensionale K-Vektorräume mit Basen B, C, und sei $s\colon V \times W \to K$ eine Sesquilinearform. Nach Proposition 7.1.12(ii), angewendet mit $\bar{V}$ und W, gilt

$$s(v, w) = [v]_B^{\mathsf{H}} \cdot [s]_{B,C} \cdot [w]_C$$

für alle $v \in V$ und $w \in W$.

Proposition 7.1.37 (Basiswechselformel für Sesquilinearformen)**.** *Sei (K, σ) ein Körper mit Involution, V und W endlich-dimensionale Vektorräume über K, B, B' Basen von V und C, C' Basen von W. Für alle Sesquilinearformen $s\colon V \times W \to K$ gilt:*

$$[s]_{B',C'} = (T_B^{B'})^{\mathsf{H}} \cdot [s]_{B,C} \cdot T_C^{C'}.$$

Beweis. Das ist der Sonderfall von Proposition 7.1.13, indem man s als Bilinearform $\bar{V} \times W \to K$ betrachtet: Wenn man B' und B als Basen von $\bar{V}$ betrachtet, ist die entsprechende Basiswechselmatrix gleich $\overline{T_B^{B'}}$, und ihre transponierte Matrix gleich $(T_B^{B'})^{\mathsf{H}}$. □

Definition 7.1.38 (hermitesche Matrix)**.** Sei (K, σ) ein Körper mit Involution und sei $n \in \mathbb{N}$. Eine Matrix $A \in M_n(K)$ heißt *hermitesch*, wenn $A^{\mathsf{H}} = A$.

Bemerkung 7.1.39. Ist $A \in M_n(K)$ hermitesch, so liegen alle Diagonalkoeffizienten A_{ii} im Fixkörper K^σ, denn $\sigma(A_{ii}) = (A^{\mathsf{H}})_{ii} = A_{ii}$.

Proposition 7.1.40 (Kriterium für hermitesche Formen)**.** *Sei (K, σ) ein Körper mit Involution, sei V ein endlich-dimensionaler K-Vektorraum und sei B eine Basis von V. Eine Sesquilinearform s auf V ist genau dann hermitesch, wenn die Darstellungsmatrix $[s]_{B,B}$ hermitesch ist.*

Beweis. Sei $B = (v_1, \dots, v_n)$, so dass $[s]_{B,B} = (s(v_i, v_j))_{i,j}$. Ist s hermitesch, so gilt $s(v_i, v_j) = \overline{s(v_j, v_i)}$ für alle i, j, d.h., $[s]_{B,B}$ ist hermitesch. Gilt umgekehrt die Gleichung $s(v, w) = \overline{s(w, v)}$ für alle v, w aus der Basis B, so gilt sie für alle $v, w \in V$ wegen der Sesquilinearität von s. □

Beispiel 7.1.41. Die komplexe Matrix

$$A = \begin{pmatrix} \pi & 1+i \\ 1-i & -1 \end{pmatrix}$$

ist hermitesch. Damit ist die zugehörige Sesquilinearform

$$s_A \colon \mathbb{C}^2 \times \mathbb{C}^2 \to \mathbb{C}, \quad s_A(x, y) = \pi \bar{x}_1 y_1 + (1+i)\bar{x}_1 y_2 + (1-i)\bar{x}_2 y_1 - \bar{x}_2 y_2,$$

auch hermitesch.

Bemerkung 7.1.42. In der Mathematik schreibt man üblicherweise V^*, f^* und A^* anstelle von $V^\dagger$, $f^\dagger$ und A^{H}, aber dies kann zu Verwechslung mit der Notation für den Dualraum und die duale Abbildung führen. Die †-Schreibweise wird häufig in der Quantenmechanik verwendet.

7.1.3 Isomorphie von Sesquilinearformen

Wir legen einen Körper mit Involution (K, σ) fest, zum Beispiel $(\mathbb{R}, \mathrm{id}_{\mathbb{R}})$ oder $(\mathbb{C}, {}^-)$.

Definition 7.1.43 (Isomorphie von Sesquilinearformen)**.** Seien V, W Vektorräume über K, s_V eine Sesquilinearform auf V und s_W eine Sesquilinearform auf W.

- Ein *Isomorphismus* von (V, s_V) nach (W, s_W) ist ein Isomorphismus $\varphi \colon V \to W$ von K-Vektorräumen, so dass für alle $v, v' \in V$ gilt:

$$s_V(v, v') = s_W(\varphi(v), \varphi(v')).$$

- Man sagt, dass die Paare (V, s_V) und (W, s_W) *isomorph* sind, in Zeichen $(V, s_V) \cong (W, s_W)$, wenn ein Isomorphismus von (V, s_V) nach (W, s_W) existiert.

Bemerkung 7.1.44. Sei s_V eine Sesquilinearform auf V und sei $\varphi \colon V \xrightarrow{\sim} W$ ein Isomorphismus von K-Vektorräumen. Dann gibt es genau eine Sesquilinearform s_W auf W, so dass φ ein Isomorphismus von (V, s_V) nach (W, s_W) ist, nämlich:

$$s_W \colon W \times W \to K, \quad (w, w') \mapsto s_V(\varphi^{-1}(w), \varphi^{-1}(w')).$$

Beispiel 7.1.45. Sei $(K, \sigma) = (\mathbb{R}, \mathrm{id}_{\mathbb{R}})$. Für $\lambda \in \mathbb{R}$ sei $b_\lambda \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ die Bilinearform mit $b_\lambda(x, y) = \lambda xy$ (nach Proposition 7.1.12(i) sind alle Bilinearformen auf $\mathbb{R}$ von dieser Gestalt). Die Paare $(\mathbb{R}, b_\lambda)$ und $(\mathbb{R}, b_\mu)$ sind genau dann isomorph, wenn λ und μ das gleiche Vorzeichen haben (positiv, negativ, oder null). Auf der einen Seite, sind λ und μ beide positiv oder beide negativ, so ist

$$\varphi \colon \mathbb{R} \to \mathbb{R}, \quad \varphi(x) = \sqrt{\frac{\lambda}{\mu}} \cdot x,$$

ein Isomorphismus von $(\mathbb{R}, b_\lambda)$ nach $(\mathbb{R}, b_\mu)$. Auf der anderen Seite ist λ genau dann positiv bzw. negativ, wenn ein Vektor $x \in \mathbb{R} \setminus \{0\}$ mit $b_\lambda(x,x) > 0$ bzw. $b_\lambda(x,x) < 0$ existiert. Damit können $(\mathbb{R}, b_\lambda)$ und $(\mathbb{R}, b_\mu)$ nicht isomorph sein, wenn λ und μ verschiedene Vorzeichen haben. Es gibt also genau drei Isomorphieklassen von Bilinearformen auf eindimensionalen $\mathbb{R}$-Vektorräumen. Der Trägheitssatz von Sylvester ist eine Verallgemeinerung dieser Aussage auf höhere Dimension, siehe Satz 7.3.19.

Beispiel 7.1.46. Seien A und B wie im Beispiel 7.1.14, so dass $[b_A]_{B,B} = I_2$. Dann ist $\varphi_B \colon \mathbb{R}^2 \xrightarrow{\sim} \mathbb{R}^2$ ein Isomorphismus von $(\mathbb{R}^2, b_{I_2})$ nach $(\mathbb{R}^2, b_A)$, denn:

$$b_{I_2}(x,y) = x^\mathsf{T} \cdot y = [\varphi_B(x)]_B^\mathsf{T} \cdot [\varphi_B(y)]_B = b_A(\varphi_B(x), \varphi_B(y)).$$

Definition 7.1.47 (Kongruenz von Matrizen)**.** Seien $n \in \mathbb{N}$ und $A, B \in M_n(K)$. Man sagt, dass A *kongruent* zu B ist, wenn eine invertierbare Matrix $S \in \mathrm{GL}_n(K)$ existiert, so dass $S^\mathsf{H} \cdot A \cdot S = B$.

Proposition 7.1.48. *Kongruenz ist eine Äquivalenzrelation auf $M_n(K)$.*

Beweis. Siehe den Beweis von Proposition 6.1.18. □

Die Kongruenzrelation auf $M_n(K)$ spielt bei Sesquilinearformen eine ähnliche Rolle wie die Ähnlichkeitsrelation bei Endomorphismen (siehe Abschnitt 6.1.3). Die folgende Proposition soll insbesondere mit Proposition 6.1.19 verglichen werden:

Proposition 7.1.49. *Sei $n \in \mathbb{N}$.*

(i) *Seien $A, B \in M_n(K)$ und seien s_A und s_B die zugehörigen Sesquilinearformen auf K^n (siehe Beispiel 7.1.25). Dann sind A und B genau dann kongruent, wenn (K^n, s_A) und (K^n, s_B) isomorph sind.*

(ii) *Seien V, W Vektorräume über K der endlichen Dimension n mit Basen B, C, s_V eine Sesquilinearform auf V und s_W eine Sesquilinearform auf W. Dann sind (V, s_V) und (W, s_W) genau dann isomorph, wenn $[s_V]_{B,B}$ und $[s_W]_{C,C}$ kongruent sind.*

Beweis. Die erste Aussage folgt aus der zweiten mit $V = W = K^n$ und $B = C$ der Standardbasis. Sei $\varphi \colon V \xrightarrow{\sim} W$ ein Isomorphismus mit $s_V(v,v') = s_W(\varphi(v), \varphi(v'))$ und sei $S = [\varphi^{-1}]_B^C \in \mathrm{GL}_n(K)$. Wir behaupten, dass $S^\mathsf{H}[s_V]_{B,B}S = [s_W]_{C,C}$. Ist $C = (w_1, \ldots, w_n)$, so ist die i-te Spalte von S gleich $[\varphi^{-1}(w_i)]_B$. Für $i, j \in \{1, \ldots, n\}$ gilt:

$$\begin{aligned}
(S^\mathsf{H} \cdot [s_V]_{B,B} \cdot S)_{ij} &= e_i^\mathsf{H} \cdot (S^\mathsf{H} \cdot [s_V]_{B,B} \cdot S) \cdot e_j \\
&= (S \cdot e_i)^\mathsf{H} \cdot [s_V]_{B,B} \cdot (S \cdot e_j) \\
&= [\varphi^{-1}(w_i)]_B^\mathsf{H} \cdot [s_V]_{B,B} \cdot [\varphi^{-1}(w_j)]_B \\
&= s_V(\varphi^{-1}(w_i), \varphi^{-1}(w_j)) \\
&= s_W(w_i, w_j) \\
&= ([s_W]_{C,C})_{ij},
\end{aligned}$$

wie gewünscht.

Seien umgekehrt $[s_V]_{B,B}$ und $[s_W]_{C,C}$ kongruent, d.h., es existiert $S \in \mathrm{GL}_n(K)$ mit $S^\mathsf{H}[s_V]_{B,B}S = [s_W]_{C,C}$. Sei $\varphi \colon V \xrightarrow{\sim} W$ der Isomorphismus mit $[\varphi^{-1}]_B^C = S$. Für alle $v, v' \in V$ gilt dann:

$$\begin{aligned}
s_V(v,v') &= [v]_B^\mathsf{H} \cdot [s_V]_{B,B} \cdot [v']_B \\
&= (S \cdot [\varphi(v)]_C)^\mathsf{H} \cdot [s_V]_{B,B} \cdot (S \cdot [\varphi(v')]_C) \\
&= [\varphi(v)]_C^\mathsf{H} \cdot (S^\mathsf{H} \cdot [s_V]_{B,B} \cdot S) \cdot [\varphi(v')]_C \\
&= [\varphi(v)]_C^\mathsf{H} \cdot [s_W]_{C,C} \cdot [\varphi(v')]_C \\
&= s_W(\varphi(v), \varphi(v')).
\end{aligned}$$

Also ist φ ein Isomorphismus von (V, s_V) nach (W, s_W), wie gewünscht. □

7.2 Skalarprodukte

Skalarprodukte sind besondere Sesquilinearformen über den reellen oder komplexen Zahlen, die die Begriffe von *Abstand* und *Winkel* zwischen Vektoren ermöglichen.

Im Folgenden sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, d.h., $\mathbb{K}$ ist entweder der Körper der reellen Zahlen oder der der komplexen Zahlen. Zudem versehen wir $\mathbb{K}$ mit der Körperinvolution $z \mapsto \bar{z}$, deren Fixkörper gleich $\mathbb{R}$ ist (falls $\mathbb{K} = \mathbb{R}$ ist also diese Involution die Identität $\mathrm{id}_{\mathbb{R}}$). Damit können wir von Sesquilinearformen auf $\mathbb{K}$-Vektorräumen sprechen (siehe Definition 7.1.24). Falls $\mathbb{K} = \mathbb{R}$ ist also eine (hermitesche) Sesquilinearform das Gleiche wie eine (symmetrische) Bilinearform.

Definition 7.2.1 (Definitheit)**.** Sei V ein $\mathbb{K}$-Vektorraum und sei $s\colon V \times V \to \mathbb{K}$ eine hermitesche Sesquilinearform, so dass $s(v, v) \in \mathbb{R}$ für alle $v \in V$ (siehe Bemerkung 7.1.31). Die Form s heißt:

- *positiv semidefinit*, wenn $s(v, v) \geq 0$ für alle $v \in V$;
- *negativ semidefinit*, wenn $s(v, v) \leq 0$ für alle $v \in V$;
- *positiv definit*, wenn $s(v, v) > 0$ für alle $v \in V \setminus \{0\}$;
- *negativ definit*, wenn $s(v, v) < 0$ für alle $v \in V \setminus \{0\}$;
- *indefinit*, wenn sie weder positiv noch negativ semidefinit ist.

Eine hermitesche Matrix $A \in M_n(\mathbb{K})$ heißt positiv/negativ (semi)definit bzw. indefinit, wenn die zugehörige Sesquilinearform s_A auf $\mathbb{K}^n$ die entsprechende Eigenschaft hat.

Bemerkung 7.2.2. Eine hermitesche Form s auf V, die entweder positiv oder negativ definit ist, ist nicht ausgeartet. Denn für einen Vektor $v \in V$ mit $s(v, -) = 0$ oder $s(-, v) = 0$ gilt insbesondere $s(v, v) = 0$, und somit $v = 0$. Eine solche Form induziert also eine injektive lineare Abbildung

$$s_r\colon V \hookrightarrow V^\dagger, \quad v \mapsto s(-, v),$$

die sogar einen Isomorphismus ist, wenn V endlich-dimensional ist (Bemerkung 7.1.5).

Definition 7.2.3 (Skalarprodukt)**.** Sei V ein Vektorraum über $\mathbb{K}$. Ein *Skalarprodukt* auf V ist eine positiv definite hermitesche Sesquilinearform auf V, d.h., eine Abbildung

$$V \times V \to \mathbb{K}, \quad (v, w) \mapsto \langle v, w \rangle,$$

mit folgenden Eigenschaften, für alle $v, v', w, w' \in V$ und alle $\lambda, \mu \in \mathbb{K}$:

(i) $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$ und $\langle \lambda v, w \rangle = \bar{\lambda} \langle v, w \rangle$.

(ii) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$ und $\langle v, \mu w \rangle = \mu \langle v, w \rangle$.

(iii) $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

(iv) Ist $v \neq 0$, so ist $\langle v, v \rangle > 0$. (Man beachte dabei, dass aus (iii) folgt bereits $\langle v, v \rangle \in \mathbb{R}$.)

Bemerkung 7.2.4. In der Definition 7.2.3 genügt es eine der Bedingungen (i) und (ii) zu erfordern: Die andere folgt dann automatisch aus (iii).

Beispiel 7.2.5 (Standardskalarprodukt)**.**

(i) Sei $n \in \mathbb{N}$. Die Abbildung

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x, y) \mapsto x^\intercal \cdot y = \sum_{i=1}^{n} x_i y_i,$$

ist ein Skalarprodukt und heißt das *Standardskalarprodukt* auf $\mathbb{R}^n$.

(ii) Sei $n \in \mathbb{N}$. Die Abbildung

$$\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}, \quad (x, y) \mapsto x^{\mathsf{H}} \cdot y = \sum_{i=1}^{n} \bar{x}_i y_i,$$

ist ein Skalarprodukt und heißt das *Standardskalarprodukt* auf $\mathbb{C}^n$. Die positiv Definitheit folgt daraus, dass für $z \in \mathbb{C} \setminus \{0\}$ gilt $\bar{z}z = |z|^2 > 0$.

Beispiel 7.2.6. Seien $a < b$ reelle Zahlen. Auf dem $\mathbb{R}$-Vektorraum $C^0([a, b], \mathbb{R})$ von stetigen Funktionen auf $[a, b]$ können wir ein Skalarprodukt wie folgt definieren:

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx.$$

Die positiv Definitheit folgt daraus, dass $\int_a^b f(x)^2\,dx > 0$ für alle $f \in C^0([a, b], \mathbb{R}) \setminus \{0\}$. Diese Bilinearform lässt sich auf dem Vektorraum $\mathcal{L}^2([a, b], \mathbb{R})$ von *quadratisch integrierbaren* Funktionen fortsetzen, aber sie ist dann nur positiv semidefinit und kein Skalarprodukt (das Integral von f^2 kann null sein, ohne dass f identisch null ist).

Für komplexe Funktionen definiert auf ähnliche Weise die Formel

$$\langle f, g \rangle = \int_a^b \overline{f(x)}g(x)\,dx$$

ein Skalarprodukt auf dem $\mathbb{C}$-Vektorraum $C^0([a, b], \mathbb{C})$.

Definition 7.2.7 (euklidischer/unitärer Vektorraum)**.**

- Ein *euklidischer Vektorraum* ist ein Paar $(V, \langle -, - \rangle)$ bestehend aus einem $\mathbb{R}$-Vektorraum V und einem Skalarprodukt $\langle -, - \rangle$ auf V.
- Ein *unitärer Vektorraum* ist ein Paar $(V, \langle -, - \rangle)$ bestehend aus einem $\mathbb{C}$-Vektorraum V und einem Skalarprodukt $\langle -, - \rangle$ auf V.

Definition 7.2.8 (Norm)**.** Sei $(V, \langle -, - \rangle)$ ein euklidischer/unitärer Vektorraum. Die von $\langle -, - \rangle$ induzierte *Norm* auf V ist die wie folgt definierte Abbildung:

$$\begin{aligned} \|-\| \colon V &\to \mathbb{R}_{\geq 0}, \\ v &\mapsto \sqrt{\langle v, v \rangle}. \end{aligned}$$

Dies ist wohldefiniert, da $\langle v, v \rangle \in \mathbb{R}_{\geq 0}$ nach Definition 7.2.3(iv).

Notation 7.2.9. Oft unterdrückt man das Skalarprodukt in der Notation für einen euklidischen/unitären Vektorraum $(V, \langle -, - \rangle)$. Das heißt, man sagt einfach „Sei V ein euklidischer/unitärer Vektorraum“. Das zugehörige Skalarprodukt wird dann immer mit $\langle -, - \rangle$ bezeichnet, und die zugehörige Norm mit $\|-\|$. Sofern nicht anders angegeben, betrachten wir immer $\mathbb{K}^n$ als euklidischer (falls $\mathbb{K} = \mathbb{R}$) bzw. unitärer (falls $\mathbb{K} = \mathbb{C}$) Vektorraum mit dem Standardskalarprodukt aus Beispiel 7.2.5.

Beispiel 7.2.10. Im euklidischen Vektorraum $\mathbb{R}^n$ mit dem Standardskalarprodukt gilt:

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

Im unitären Vektorraum $\mathbb{C}^n$ mit dem Standardskalarprodukt gilt:

$$\|x\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2} = \sqrt{(\operatorname{Re} x_1)^2 + (\operatorname{Im} x_1)^2 + \cdots + (\operatorname{Re} x_n)^2 + (\operatorname{Im} x_n)^2}.$$

In beiden Fällen ist also $\|x\|$ die Länge des Vektors x im Sinne der euklidischen Geometrie. Im eindimensionalen unitären Vektorraum $\mathbb{C}$ ist insbesondere die Norm $\|-\|$ gleich dem üblichen Betrag $|-|$.

Satz 7.2.11 (Ungleichung von Cauchy-Schwarz)**.** *Sei $(V, \langle -, - \rangle)$ ein euklidischer/unitärer Vektorraum. Für alle $v, w \in V$ gilt:*

$$|\langle v, w \rangle| \leq \|v\| \|w\|.$$

Außerdem gilt die Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis. Falls $v = 0$ ist der Satz klar, da beide Seiten null sind. Wir dürfen damit annehmen, dass $v \neq 0$. Sind v und w linear abhängig, so existiert $\lambda \in \mathbb{K}$ mit $w = \lambda v$, und es gilt:

$$|\langle v, w \rangle| = |\lambda \langle v, v \rangle| = |\lambda| \|v\|^2 = \|v\| \|w\|,$$

wie gewünscht. Für $v \in V \setminus \{0\}$ und $w \in V \setminus \mathbb{K}v$ setzen wir jetzt

$$\lambda := \frac{\langle v, w \rangle}{\|v\|^2}.$$

Da $w \notin \mathbb{K}v$ gilt $\lambda v - w \neq 0$. Man berechnet:

$$\begin{aligned} 0 &< \langle \lambda v - w, \lambda v - w \rangle \\ &= \lambda \bar{\lambda} \langle v, v \rangle - \bar{\lambda} \langle v, w \rangle - \lambda \langle w, v \rangle + \langle w, w \rangle \\ &= \frac{|\langle v, w \rangle|^2}{\|v\|^4} \|v\|^2 - 2 \frac{|\langle v, w \rangle|^2}{\|v\|^2} + \|w\|^2 \\ &= \|w\|^2 - \frac{|\langle v, w \rangle|^2}{\|v\|^2}. \end{aligned}$$

Damit ist $\|v\|^2 \|w\|^2 > |\langle v, w \rangle|^2$. Die gewünschte Ungleichung folgt, da die Wurzelfunktion $\sqrt{-}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ streng monoton wachsend ist. □

Beispiel 7.2.12. Für das Skalarprodukt auf $C^0([a, b], \mathbb{C})$ aus Beispiel 7.2.6 erhalten wir die Ungleichung:

$$\left| \int_a^b \overline{f(x)} g(x) \, dx \right| \leq \sqrt{\int_a^b |f(x)|^2 \, dx} \sqrt{\int_a^b |g(x)|^2 \, dx}.$$

Proposition 7.2.13 (Eigenschaften der Norm)**.** *Sei $(V, \langle -, - \rangle)$ ein euklidischer/unitärer Vektorraum. Dann hat die induzierte Norm $\|-\|\colon V \to \mathbb{R}_{\geq 0}$ die folgenden Eigenschaften:*

(i) (Definitheit) *Ist $v \in V \setminus \{0\}$, so ist $\|v\| > 0$.*

(ii) (Homogenität) *Für alle $\lambda \in \mathbb{K}$ und alle $v \in V$ gilt: $\|\lambda \cdot v\| = |\lambda| \|v\|$.*

(iii) (Dreiecksungleichung) *Für alle $v, w \in V$ gilt: $\|v + w\| \leq \|v\| + \|w\|$.*

Beweis. Zu (i). Ist $v \in V \setminus \{0\}$, so ist $\langle v, v \rangle > 0$ nach positiv Definitheit, und somit ist auch $\|v\| = \sqrt{\langle v, v \rangle} > 0$.

Zu (ii). Nach Definition der Norm ist

$$\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v \rangle} = \sqrt{|\lambda|^2 \langle v, v \rangle} = |\lambda| \|v\|.$$

Zu (iii). Es genügt zu zeigen, dass $\|v + w\|^2 \leq \big(\|v\| + \|w\|\big)^2$. Auf der einer Seite gilt:

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \|v\|^2 + 2 \operatorname{Re} \langle v, w \rangle + \|w\|^2.$$

Auf der anderen Seite gilt:

$$\big(\|v\| + \|w\|\big)^2 = \|v\|^2 + 2\|v\|\|w\| + \|w\|^2.$$

Die gewünschte Ungleichung folgt jetzt aus der Ungleichung $\operatorname{Re} z \leq |z|$ für $z \in \mathbb{C}$ und der Ungleichung von Cauchy-Schwarz $|\langle v, w \rangle| \leq \|v\| \|w\|$ (Satz 7.2.11). □

Bemerkung 7.2.14. Ein $\mathbb{K}$-Vektorraum V mit einer Abbildung $\|-\|\colon V \to \mathbb{R}_{\geq 0}$, die die Bedingungen (i)–(iii) aus Proposition 7.2.13 erfüllt, heißt ein *normierter Vektorraum*. Jeder euklidischer/unitärer Vektorraum ist insbesondere mit seiner Norm ein normierter Vektorraum, aber es gibt auch normierte Vektorräume, deren Norm von keinem Skalarprodukt induziert wird. In der Analysis spielen beide euklidische/unitäre und normierte Vektorräume eine wichtige Rolle.

Beispiel 7.2.15. Sei $p \in \mathbb{R}$ eine reelle Zahl mit $p \geq 1$. Dann ist die Abbildung

$$\begin{aligned} \|-\|_p\colon C^0([a,b],\mathbb{C}) &\to \mathbb{R}_{\geq 0}, \\ f &\mapsto \left(\int_a^b |f(x)|^p\,dx\right)^{1/p}, \end{aligned}$$

eine Norm auf dem $\mathbb{C}$-Vektorraum $C^0([a,b],\mathbb{C})$ (d.h., die Bedingungen (i)–(iii) der Proposition 7.2.13 sind erfüllt). Die Norm $\|-\|_2$ wird von dem Skalarprodukt aus Beispiel 7.2.6 induziert. Man kann aber zeigen, dass die Norm $\|-\|_p$ von keinem Skalarprodukt induziert wird, wenn $p \neq 2$.

Definition 7.2.16 (normierter Vektor, Normierung)**.** Sei V ein euklidischer/unitärer Vektorraum (oder allgemeiner ein normierter Vektorraum).

- Ein Vektor $v \in V$ heißt *normiert*, falls $\|v\| = 1$.
- Ist $v \in V \setminus \{0\}$, so heißt der Vektor $\frac{1}{\|v\|}v$ die *Normierung* von v.

Nach der Homogenität der Norm ist die Normierung von v ein normierter Vektor, und zwar der einzige normierte Vektor, der ein positives Skalarvielfaches von v ist.

Beispiel 7.2.17. Im euklidischen Vektorraum $C^0([0,2\pi],\mathbb{R})$ aus Beispiel 7.2.6 gilt:

$$\|\cos\|^2 = \int_0^{2\pi} \cos(x)^2\,dx = \pi.$$

Deswegen ist die Funktion $\frac{1}{\sqrt{\pi}}\cos$ normiert. Auf ähnliche Weise ist die Funktion $\frac{1}{\sqrt{\pi}}\sin$ normiert.

Bemerkung 7.2.18 (Abstand, Winkel)**.** In einem euklidischen/unitären Vektorraum V kann mann den Abstand und den Winkel zwischen zwei Vektoren definieren. Der *Abstand* zwischen $v, w \in V$ ist $\|v - w\| \in \mathbb{R}_{\geq 0}$. Aus Proposition 7.2.13 folgt leicht, dass die Funktion

$$V \times V \to \mathbb{R}_{\geq 0}, \quad (v,w) \mapsto \|v - w\|,$$

eine Metrik auf V ist. Der *Winkel* zwischen $v, w \in V \setminus \{0\}$ ist die eindeutige reelle Zahl $\alpha \in [0,\pi]$ mit

$$\cos(\alpha) = \frac{\mathrm{Re}\langle v, w\rangle}{\|v\|\|w\|}.$$

Dies ist wohldefiniert, da die rechte Seite in $[-1,1]$ liegt (nach der Ungleichung von Cauchy-Schwarz) und die Abbildung $\cos\colon [0,\pi] \to [-1,1]$ bijektiv ist. Wenn V entweder $\mathbb{R}^n$ oder $\mathbb{C}^n$ mit dem Standardskalarprodukt ist, dann erhalten wir die „gewöhnlichen“ Begriffe von Abstand und Winkel.

Ein Skalarprodukt ist eigentlich durch seine induzierte Norm eindeutig bestimmt:

Proposition 7.2.19 (Polarisierung)**.** *Sei $(V, \langle -,-\rangle)$ ein euklidischer/unitärer $\mathbb{K}$-Vektorraum und seien $v, w \in V$.*

(i) *Falls* $\mathbb{K} = \mathbb{R}$ *gilt:*

$$\langle v, w\rangle = \frac{1}{2}\big(\|v+w\|^2 - \|v\|^2 - \|w\|^2\big).$$

(ii) *Falls* $\mathbb{K} = \mathbb{C}$ *gilt:*

$$\langle v, w\rangle = \frac{1}{4}\big(\|v+w\|^2 - \|v-w\|^2\big) - \frac{i}{4}\big(\|v+iw\|^2 - \|v-iw\|^2\big).$$

Beweis. Beide Gleichungen lassen sich leicht prüfen, indem man die rechte Seite entwickelt und die Sesquilinearität von $\langle -, -\rangle$ verwendet. Wir beweisen stellvertretend die erste Gleichung:

$$\begin{aligned}\|v+w\|^2 - \|v\|^2 - \|w\|^2 &= \langle v+w, v+w\rangle - \langle v, v\rangle - \langle w, w\rangle \\ &= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle - \langle v, v\rangle - \langle w, w\rangle = 2\langle v, w\rangle.\end{aligned}$$ □

Bemerkung 7.2.20 (Skalarprodukte über anderen Körpern)**.** Der Begriff von Skalarprodukt haben wir nur für $\mathbb{R}$- und $\mathbb{C}$-Vektorräume definiert. Warum nicht über einem beliebigen Körper mit Involution (K, σ)? Um die positiv Definitheit einer hermiteschen Form auf V zu erfordern, ist es nötig, dass der Fixkörper K^σ *angeordnet* ist, d.h., es muss auf K^σ eine totale Ordnung $\leq$ gegeben werden, die auf geeignete Weise mit den Körperverknüpfungen verträglich ist (vgl. die Anordnungsaxiome für die reellen Zahlen). Insbesondere muss die Charakteristik von K null sein. Außerdem muss jeder positive Skalar in K^σ eine Quadratwurzel besitzen, um die Norm $\|-\|\colon V \to (K^\sigma)_{\geq 0}$ zu definieren. Ist K^σ ein angeordneter Körper mit dieser weiteren Eigenschaft, dann gelten die meisten Sätze in diesem Abschnitt (z.B. die Ungleichung von Cauchy-Schwarz oder das Orthonormalisierungsverfahren von Gram-Schmidt) für unitäre K-Vektorräume. Für den Spektralsatz und seine Konsequenzen braucht man außerdem, dass der Körper K algebraisch abgeschlossen ist. Ein Beispiel ist der Körper $\bar{\mathbb{Q}}$ der algebraischen Zahlen, d.h., der Teilkörper von $\mathbb{C}$ bestehend aus allen Nullstellen von Polynomen über $\mathbb{Q}$, mit der komplexen Konjugation. Es gibt auch solche Körper, die „größer“ als $\mathbb{C}$ sind.

7.2.1 Orthogonalität und Orthonormalität

Definition 7.2.21 (orthogonal, Orthogonalraum)**.** Sei V ein euklidischer/unitärer Vektorraum.

- Seien $v, w \in V$. Man sagt, dass v *orthogonal* zu w ist, und man schreibt $v \perp w$, wenn $\langle v, w\rangle = 0$.
- Sei $A \subset V$ eine Teilmenge. Der *Orthogonalraum* zu A ist der Untervektorraum

$$A^\perp := \{v \in V \mid \text{für alle } u \in A \text{ gilt } v \perp u\} \subset V.$$

Beispiel 7.2.22. In $\mathbb{K}^n$ mit dem Standardskalarprodukt sind alle Standardeinheitsvektoren zueinander orthogonal, d.h.: $e_i \perp e_j$ für alle $i \neq j$. Für eine Teilmenge $I \subset \{1, \dots, n\}$ gilt:

$$\mathrm{Span}_\mathbb{K}\{e_i \mid i \in I\}^\perp = \mathrm{Span}_\mathbb{K}\{e_j \mid j \in \{1, \dots, n\} \setminus I\}.$$

Beispiel 7.2.23. In $\mathbb{R}^2$ sind $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} -x_2 \\ x_1 \end{smallmatrix}\right)$ orthogonal. In $\mathbb{C}^2$ sind $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} -\bar{x}_2 \\ \bar{x}_1 \end{smallmatrix}\right)$ orthogonal.

Bemerkung 7.2.24. Aus der Sesquilinearität von $\langle -, -\rangle$ folgt $A^\perp = \mathrm{Span}_\mathbb{K}(A)^\perp$.

Bemerkung 7.2.25. Im euklidischen Fall sind zwei Vektoren $v, w \in V \setminus \{0\}$ genau dann orthogonal, wenn der Winkel zwischen denen gleich $\frac{\pi}{2}$ ist (siehe Bemerkung 7.2.18). Dies gilt aber nicht im unitären Fall: Der Winkel zwischen 1 und i in $\mathbb{C}$ ist $\frac{\pi}{2}$, aber $\langle 1, i\rangle = i \neq 0$.

Lemma 7.2.26. *Sei V ein euklidischer/unitärer Vektorraum und sei $A \subset V$ eine Teilmenge. Dann ist $A \cap A^\perp = \{0\}$.*

Beweis. Ein Vektor $v \in A \cap A^\perp$ muss insbesondere zu sich selbst orthogonal sein, d.h., $\langle v, v\rangle = 0$. Nach der positiv Definitheit von $\langle -, -\rangle$ ist das nur möglich, wenn $v = 0$. □

Definition 7.2.27 (Orthonormalsystem, Orthonormalbasis)**.** Sei V ein euklidischer/unitärer Vektorraum.

- Eine Familie $(v_i)_{i\in I}$ in V heißt *Orthonormalsystem*, wenn für alle $i, j \in I$ gilt:

  $$\langle v_i, v_j\rangle = \delta_{ij}.$$

  Dabei ist δ_{ij} das Kronecker-Delta (Notation 4.2.5). Anders gesagt: Jedes v_i ist normiert und orthogonal zu allen anderen v_j.

- Eine *Orthonormalbasis* von V ist eine Basis, die auch ein Orthonormalsystem ist.

Beispiel 7.2.28. Die Standardbasis $(e_1, \dots, e_n)$ von $\mathbb{K}^n$ ist ein Orthonormalsystem (bezügich dem Standardskalarprodukt auf $\mathbb{K}^n$), und somit eine Orthonormalbasis.

Beispiel 7.2.29. Im euklidischen Vektorraum $C^0([0, 2\pi], \mathbb{R})$ aus Beispiel 7.2.6 sind die Funktionen cos und sin orthogonal, denn:

$$\int_0^{2\pi} \cos(x)\sin(x)\,dx = 0.$$

Nach Beispiel 7.2.17 ist $\left(\frac{1}{\sqrt{\pi}}\cos, \frac{1}{\sqrt{\pi}}\sin\right)$ ein Orthonormalsystem in $C^0([0, 2\pi], \mathbb{R})$.

Proposition 7.2.30. *Jedes Orthonormalsystem ist linear unabhängig. Insbesondere ist jedes erzeugende Orthonormalsystem eine Orthonormalbasis.*

Beweis. Sei $(v_i)_{i\in I}$ ein Orthonormalsystem und sei $(\lambda_i)_{i\in I}$ ein Element von $\mathbb{K}^{(I)}$, so dass $\sum_{i\in I} \lambda_i v_i = 0$. Für jedes $j \in I$ gilt dann:

$$\lambda_j = \sum_{i\in I} \lambda_i \delta_{ij} = \sum_{i\in I} \lambda_i \langle v_j, v_i\rangle = \left\langle v_j, \sum_{i\in I} \lambda_i v_i \right\rangle = 0. \qquad \square$$

Man kann Koordinatenvektoren bzgl. einer Orthonormalbasis B durch das Skalarprodukt berechnen:

Proposition 7.2.31 (Koordinatenvektor bzgl. einer Orthonormalbasis)**.** *Sei V ein euklidischer/unitärer Vektorraum mit einer Orthonormalbasis $B = (w_i)_{i\in I}$, und sei $v \in V$. Dann gilt:*

$$[v]_B = (\langle w_i, v\rangle)_{i\in I}.$$

Beweis. Die Aussage ist die Gleichung $v = \sum_{i\in I} \langle w_i, v\rangle w_i$. Da $\langle -, -\rangle$ nicht ausgeartet ist, genügt es zu zeigen, dass für alle $w \in V$ gilt $\langle w, v\rangle = \left\langle w, \sum_{i\in I} \langle w_i, v\rangle w_i\right\rangle$. Da B erzeugend ist, können wir $w = w_j$ nehmen. Da B ein Orthonormalsystem ist, gilt nun:

$$\left\langle w_j, \sum_{i\in I} \langle w_i, v\rangle w_i \right\rangle = \sum_{i\in I} \langle w_i, v\rangle \delta_{ij} = \langle w_j, v\rangle,$$

wie gewünscht. □

Satz 7.2.32 (Orthonormalisierungsverfahren von Gram-Schmidt)**.** *Sei V ein euklidischer/ unitärer $\mathbb{K}$-Vektorraum. Sei $n \in \mathbb{N}$ und sei $(v_1, \dots, v_n)$ eine linear unabhängige Familie in V. Man definiert die Familie $(w_1, \dots, w_n)$ rekursiv durch:*

$$w_k = \frac{1}{\|\tilde{w}_k\|}\tilde{w}_k, \quad \textit{wobei} \quad \tilde{w}_k = v_k - \sum_{i=1}^{k-1} \langle w_i, v_k \rangle w_i.$$

Dann ist $(w_1, \dots, w_n)$ ein Orthonormalsystem in V, und es gilt

$$\mathrm{Span}_{\mathbb{K}}\{v_1, \dots, v_k\} = \mathrm{Span}_{\mathbb{K}}\{w_1, \dots, w_k\}$$

für alle $k \in \{0, \dots, n\}$.

Beweis. Wir beweisen durch Induktion über $k \in \{0, \dots, n\}$, dass $(w_1, \dots, w_k)$ ein Orthonormalsystem mit demselben Spann wie $(v_1, \dots, v_k)$ ist. Für $k \geq 1$ braucht man auch zu zeigen, dass $\tilde{w}_k \neq 0$, so dass w_k wohldefiniert ist. Im Fall $k = 0$ gibt es nichts zu zeigen. Sei also $k \in \{1, \dots, n\}$. Nach Induktionsvoraussetzung ist $(w_1, \dots, w_{k-1})$ ein Orthonormalsystem mit

$$\mathrm{Span}_{\mathbb{K}}\{v_1, \dots, v_{k-1}\} = \mathrm{Span}_{\mathbb{K}}\{w_1, \dots, w_{k-1}\}.$$

- *Es gilt $\tilde{w}_k \neq 0$.* Nach Definition gilt

  $$v_k - \tilde{w}_k \in \mathrm{Span}_{\mathbb{K}}\{w_1, \dots, w_{k-1}\} = \mathrm{Span}_{\mathbb{K}}\{v_1, \dots, v_{k-1}\}.$$

  Insbesondere ist $\tilde{w}_k$ eine nicht-triviale Linearkombination der Familie $(v_1, \dots, v_k)$. Da diese Familie linear unabhängig ist, gilt $\tilde{w}_k \neq 0$.

- *Es gilt* $\mathrm{Span}_{\mathbb{K}}\{v_1, \dots, v_k\} = \mathrm{Span}_{\mathbb{K}}\{w_1, \dots, w_k\}$. Nach Definition ist w_k eine Linearkombination von $w_1, \dots, w_{k-1}, v_k$, und liegt deswegen in $\mathrm{Span}_{\mathbb{K}}\{v_1, \dots, v_k\}$. Umgekehrt ist v_k eine Linearkombination von $w_1, \dots, w_{k-1}, w_k$, und liegt deswegen in $\mathrm{Span}_{\mathbb{K}}\{w_1, \dots, w_k\}$.

- $(w_1, \dots, w_k)$ *ist ein Orthonormalsystem.* Es bleibt zu zeigen, dass

  $$\langle w_j, w_k \rangle = \delta_{jk} \quad \text{für alle } j \in \{1, \dots, k\}.$$

  Für $j < k$ gilt:

  $$\langle w_j, \tilde{w}_k \rangle = \langle w_j, v_k \rangle - \sum_{i=1}^{k-1} \langle w_i, v_k \rangle \underbrace{\langle w_j, w_i \rangle}_{=\delta_{ij}} = \langle w_j, v_k \rangle - \langle w_j, v_k \rangle = 0,$$

  und somit auch $\langle w_j, w_k \rangle = 0$. Für $j = k$ gilt $\langle w_k, w_k \rangle = \|w_k\|^2 = 1$ nach Konstruktion. □

Beispiel 7.2.33. Wir führen das Orthonormalisierungsverfahren von Gram-Schmidt mit den Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad v_3 = \begin{pmatrix} 0 \\ -3 \\ 0 \\ 1 \end{pmatrix}$$

aus $\mathbb{R}^4$ durch, um ein Orthonormalsystem (w_1, w_2, w_3) mit

$$\mathrm{Span}_{\mathbb{R}}\{v_1, v_2, v_3\} = \mathrm{Span}_{\mathbb{R}}\{w_1, w_2, w_3\}$$

zu erhalten:

$$
\begin{aligned}
w_1 &= \frac{1}{\|v_1\|} v_1 = \frac{1}{3} v_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ -2/3 \\ 0 \end{pmatrix}, \\
\tilde{w}_2 &= v_2 - \langle w_1, v_2 \rangle w_1 = v_2 - w_1 = \begin{pmatrix} 2/3 \\ 1/3 \\ 2/3 \\ 0 \end{pmatrix}, \\
w_2 &= \frac{1}{\|\tilde{w}_2\|} \tilde{w}_2 = \tilde{w}_2 = \begin{pmatrix} 2/3 \\ 1/3 \\ 2/3 \\ 0 \end{pmatrix}, \\
\tilde{w}_3 &= v_3 - \langle w_1, v_3 \rangle w_1 - \langle w_2, v_3 \rangle w_2 = v_3 + 2w_1 + w_2 = \begin{pmatrix} 4/3 \\ -4/3 \\ -2/3 \\ 1 \end{pmatrix}, \\
w_3 &= \frac{1}{\|\tilde{w}_3\|} \tilde{w}_3 = \frac{1}{\sqrt{5}} \tilde{w}_3 = \frac{1}{\sqrt{5}} \begin{pmatrix} 4/3 \\ -4/3 \\ -2/3 \\ 1 \end{pmatrix}.
\end{aligned}
$$

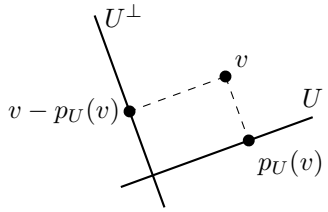
Korollar 7.2.34 (Existenz von Orthonormalbasen)**.** *Jeder endlich-dimensionale euklidische/unitäre Vektorraum V besitzt eine Orthonormalbasis. Allgemeiner: Jedes Orthonormalsystem in V kann zu einer Orthonormalbasis ergänzt werden.*

Beweis. Sei $(w_1, \ldots, w_k)$ ein Orthonormalsystem in V, das nach Proposition 7.2.30 linear unabhängig ist. Man ergänzt es zu einer Basis $(w_1, \ldots, w_k, v_{k+1}, \ldots, v_n)$ von V und führt das Orthonormalisierungsverfahren von Gram-Schmidt durch, um eine Orthonormalbasis zu erhalten. Man beachte dabei, dass sich die ersten k Vektoren nicht verändern. □

Proposition 7.2.35 (Existenz von orthogonalen Projektionen)**.** *Sei V ein euklidischer/unitärer Vektorraum und sei $U \subset V$ ein* endlich-dimensionaler *Untervektorraum. Dann existiert genau ein Endomorphismus $p_U \colon V \to V$ mit folgenden Eigenschaften:*

- $\operatorname{im} p_U \subset U$.
- $\operatorname{im}(\mathrm{id}_V - p_U) \subset U^\perp$.

Außerdem gilt $p_U(u) = u$ für alle $u \in U$ und $p_U(w) = 0$ für alle $w \in U^\perp$. Man bezeichnet p_U als orthogonale Projektion *von V auf U.*



Beweis. Sei p_U eine lineare Abbildung mit den gegebenen Eigenschaften. Jedes $v \in V$ lässt sich schreiben als

$$v = p_U(v) + (v - p_U(v)),$$

mit $p_U(v) \in U$ und $v - p_U(v) \in U^\perp$. Ist $v \in U$, so ist $v - p_U(v) \in U \cap U^\perp = \{0\}$ (Lemma 7.2.26), und somit ist $p_U(v) = v$. Ist $v \in U^\perp$, so ist $p_U(v) \in U \cap U^\perp = \{0\}$.

Zur Eindeutigkeit. Seien p_U und q_U zwei solche Endomorphismen und sei $v \in V$. Dann ist $p_U(v) - q_U(v) \in U$ und

$$p_U(v) - q_U(v) = (v - q_U(v)) - (v - p_U(v)) \in U^\perp.$$

Nach Lemma 7.2.26 ist $U \cap U^\perp = \{0\}$, und somit ist $p_U(v) = q_U(v)$.

Zur Existenz. Nach Korollar 7.2.34 besitzt U eine Orthonormalbasis $(u_1, \ldots, u_n)$. Man definiert eine lineare Abbildung $p_U \colon V \to V$ durch:

$$p_U(v) := \sum_{i=1}^{n} \langle u_i, v \rangle u_i.$$

Nach Konstruktion ist $\operatorname{im} p_U \subset U$. Sei $v \in V$ und $j \in \{1, \ldots, n\}$. Aus $\langle u_j, u_i \rangle = \delta_{ij}$ folgt:

$$\langle u_j, v - p_U(v) \rangle = \langle u_j, v \rangle - \sum_{i=1}^{n} \langle u_i, v \rangle \langle u_j, u_i \rangle = \langle u_j, v \rangle - \langle u_j, v \rangle = 0.$$

Da dies für beliebiges j gilt, liegt $v - p_U(v)$ in $U^\perp$, wie gewünscht. □

Korollar 7.2.36. *Sei V ein euklidischer/unitärer Vektorraum und sei $U \subset V$ ein* endlich-dimensionaler *Untervektorraum. Dann:*

(i) *$U^\perp$ ist komplementär zu U in V.*

(ii) *Es gilt $(U^\perp)^\perp = U$.*

Beweis. Zu (i). Nach Lemma 7.2.26 genügt es zu zeigen, dass $U + U^\perp = V$. Nach Proposition 7.2.35 existiert die orthogonale Projektion $p_U \colon V \to V$ auf U. Ist $v \in V$, so liegen insbesondere $p_U(v)$ in U und $v - p_U(v)$ im Orthogonalraum $U^\perp$. Da $v = p_U(v) + (v - p_U(v))$, liegt v in der Summe $U + U^\perp$.

Zu (ii). Ist $u \in U$ und $w \in U^\perp$, so gilt $\langle u, w \rangle = \overline{\langle w, u \rangle} = 0$ nach Definition von $U^\perp$. Damit ist $u \in (U^\perp)^\perp$. Sei umgekehrt $v \in (U^\perp)^\perp$. Nach (i) kann man schreiben $v = u + w$ mit $u \in U$ und $w \in U^\perp$. Aus $v, u \in (U^\perp)^\perp$ folgt $w \in (U^\perp)^\perp$. Aus Lemma 7.2.26, angewendet mit $A = U^\perp$, folgt $w = 0$ und somit $v \in U$. □

Bemerkung 7.2.37. Das Korollar 7.2.36 gilt im Allgemeinen nicht, wenn $\dim_{\mathbb{K}} U = \infty$.

Bemerkung 7.2.38 (Hilberträume)**.** Ein euklidischer/unitärer $\mathbb{K}$-Vektorraum V heißt *Hilbertraum*, wenn er als metrischer Raum *vollständig* ist, d.h., wenn jede Cauchy-Folge in V konvergiert. Die Vollständigkeit von V ist automatisch, wenn V endlich-dimensional ist (nach Beispiel 7.2.41); sie ist also nur eine weitere Bedingung im unendlich-dimensionalen Fall. Wenn V ein Hilbertraum ist, gilt Korollar 7.2.36 für alle Unterhilberträume $U \subset V$, selbst wenn $\dim_{\mathbb{K}} U = \infty$. Hilberträume spielen eine wichtige Rolle in der Analysis (z.B. in der Maßtheorie) und in der Quantenmechanik (wobei $\mathbb{K} = \mathbb{C}$).

7.2.2 Orthogonale und unitäre Gruppen

Definition 7.2.39 (lineare Isometrie)**.** Seien V und W euklidische bzw. unitäre Vektorräume. Eine lineare Abbildung $f \colon V \to W$ heißt *lineare Isometrie*, wenn sie ein Isomorphismus von $(V, \langle -, - \rangle)$ nach $(W, \langle -, - \rangle)$ ist, im Sinne der Definition 7.1.43, d.h., wenn die folgenden Bedingungen erfüllt sind:

(i) f ist ein Isomorphismus von $\mathbb{K}$-Vektorräumen.

(ii) Für alle $v, v' \in V$ gilt:
$$\langle v, v' \rangle = \langle f(v), f(v') \rangle.$$

Bemerkung 7.2.40. Eine lineare Abbildung $f \colon V \to W$, die nur die Bedingung (ii) der Definition 7.2.39 erfüllt, heißt *lineare isometrische Einbettung*. Wie dieser Name vermuten lässt ist eine solche Abbildung automatisch injektiv (denn jeder Vektor v im Kern muss $\langle v, v \rangle = 0$ erfüllen). Eine solche Abbildung ist insbesondere *normerhaltend*, d.h., sie erfüllt die weitere Bedingung:

(iii) Für alle $v \in V$ gilt: $\|v\| = \|f(v)\|$.

Aus Proposition 7.2.19 folgt eigentlich, dass Bedingungen (ii) und (iii) äquivalent sind. Lineare isometrische Einbettung erhalten auch Abstände und Winkel zwischen Vektoren (siehe Bemerkung 7.2.18).

Beispiel 7.2.41. Sei V ein euklidischer/unitärer $\mathbb{K}$-Vektorraum der endlichen Dimension n mit einer Basis B. Dann ist die Abbildung

$$\varphi_B \colon \mathbb{K}^n \xrightarrow{\sim} V$$

genau dann eine lineare Isometrie von $\mathbb{K}^n$ (mit dem Standardskalarprodukt) nach V, wenn B eine Orthonormalbasis ist. Nach Korollar 7.2.34 ist insbesondere jeder endlich-dimensionale euklidische/unitäre $\mathbb{K}$-Vektorraum zu $(\mathbb{K}^n, \langle -, - \rangle)$ isomorph.

Definition 7.2.42 (orthogonale/unitäre Gruppen und Matrizen)**.** Sei $n \in \mathbb{N}$.

- Die *orthogonale Gruppe* ist die Menge

  $$\mathrm{O}(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid L_A \colon \mathbb{R}^n \to \mathbb{R}^n \text{ ist eine lineare Isometrie}\}$$

  versehen mit der Matrixmultiplikation. Elemente von $\mathrm{O}(n)$ heißen *orthogonale Matrizen*.

- Die *unitäre Gruppe* ist die Menge

  $$\mathrm{U}(n) = \{A \in \mathrm{GL}_n(\mathbb{C}) \mid L_A \colon \mathbb{C}^n \to \mathbb{C}^n \text{ ist eine lineare Isometrie}\}$$

  versehen mit der Matrixmultiplikation. Elemente von $\mathrm{U}(n)$ heißen *unitäre Matrizen*.

Proposition 7.2.43 (Charakterisierung orthogonaler/unitärer Matrizen)**.** *Für eine Matrix $A \in \mathrm{GL}_n(\mathbb{K})$ sind die folgenden Aussagen äquivalent:*

(i) *Die Abbildung $L_A \colon \mathbb{K}^n \to \mathbb{K}^n$ ist eine lineare Isometrie.*

(ii) *Es gilt $A^{-1} = A^{\mathsf{H}}$.*

(iii) *Die Spalten von A bilden ein Orthonormalsystem.*

(iv) *Die Zeilen von A bilden ein Orthonormalsystem.*

Beweis. Zu (i) $\Leftrightarrow$ (ii). Der Isomorphismus L_A ist genau dann eine lineare Isometrie, wenn für alle $x, y \in \mathbb{K}^n$ gilt $L_A(x)^{\mathsf{H}} \cdot L_A(y) = x^{\mathsf{H}} \cdot y$, d.h.:

$$x^{\mathsf{H}} A^{\mathsf{H}} A y = x^{\mathsf{H}} y.$$

Dies gilt natürlich, falls $A^{-1} = A^{\mathsf{H}}$. Nimmt man umgekehrt Standardeinheitsvektoren für x und y in der obigen Gleichung, kann man daraus schließen, dass $A^{\mathsf{H}} A = I_n$, d.h., $A^{-1} = A^{\mathsf{H}}$.

Zu (i) $\Leftrightarrow$ (iii). Die Spalten von A sind die Vektoren $L_A(e_i)$. Nach Definition bilden sie genau dann ein Orthonormalsystem, wenn

$$\langle L_A(e_i), L_A(e_j) \rangle = \delta_{ij} = \langle e_i, e_j \rangle.$$

Da jeder Vektor aus $\mathbb{K}^n$ eine Linearkombination der Standardeinheitsvektoren ist, ist das Letztere äquivalent zu: Für alle $x, y \in \mathbb{K}^n$ gilt

$$\langle L_A(x), L_A(y) \rangle = \langle x, y \rangle,$$

was genau die Definition einer linearen Isometrie ist.

Zu (i) $\Rightarrow$ (iv). Nach der bereits bewiesenen Äquivalenz zwischen (i) und (ii) ist $L_{A^{\mathsf{H}}} = L_A^{-1}$ wieder eine lineare Isometrie. Aus der Implikation (i) $\Rightarrow$ (iii) folgt, dass die Spalten von A^{H}

ein Orthonormalsystem bilden. Die Spalten von A^{H} sind aber die Zeilen von $\bar{A}$. Für Vektoren $x, y \in \mathbb{K}^n$ gilt

$$\langle \bar{x}, \bar{y} \rangle = \overline{\langle x, y \rangle},$$

und daraus schließen wir, dass die Zeilen von A auch ein Orthonormalsystem bilden.

Zu (iv) $\Rightarrow$ (ii). Die Spalten von A^{H} bilden ein Orthonormalsystem, da sie zu den Zeilen von A konjugiert sind. Aus der Implikation (iii) $\Rightarrow$ (ii) folgt, dass $(A^{\mathsf{H}})^{-1} = (A^{\mathsf{H}})^{\mathsf{H}} = A$, und damit dass $A^{\mathsf{H}} = A^{-1}$. □

Bemerkung 7.2.44. Eine quadratische Matrix über $\mathbb{R}$ kann man auch als Matrix über $\mathbb{C}$ betrachten, und sie ist genau dann orthogonal, wenn sie unitär ist. Dies folgt zum Beispiel aus der Aussage (ii) der Proposition 7.2.43. Anders gesagt gilt:

$$\mathrm{O}(n) = \mathrm{U}(n) \cap \mathrm{GL}_n(\mathbb{R}).$$

Beispiel 7.2.45. Nach Proposition 7.2.43(iii) sind folgende Matrizen unitär:

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ i & 0 \end{pmatrix}, \quad \frac{1}{3} \begin{pmatrix} 2 & 1 & 2 \\ -2 & 2 & 1 \\ 1 & 2 & -2 \end{pmatrix}.$$

Korollar 7.2.46. *Sei $n \in \mathbb{N}$.*

(i) *Ist $A \in \mathrm{O}(n)$, so gilt $\det(A) \in \{\pm 1\}$.*

(ii) *Ist $A \in \mathrm{U}(n)$, so gilt $|\det(A)| = 1$, d.h., es gibt ein $\vartheta \in [0, 2\pi)$ mit $\det(A) = \exp(i\vartheta)$.*

Beweis. Die erste Aussage ist ein Sonderfall der zweiten, denn ± 1 sind die einzigen reellen Zahlen mit Betrag 1. Ist $A \in \mathrm{U}(n)$, so folgt $A^{\mathsf{H}} A = I_n$, und somit

$$1 = \det(A^{\mathsf{H}} A) = \det(\bar{A}^{\mathsf{T}}) \det(A) = \det(\bar{A}) \det(A) = \overline{\det(A)} \det(A) = |\det(A)|^2,$$

wie gewünscht. Die benutzte Gleichheit $\det(\bar{A}) = \overline{\det(A)}$ folgt unmittelbar aus der Leibniz-Formel für die Determinante. □

Proposition 7.2.47 (Orthonormalbasiswechsel)**.** *Sei V ein endlich-dimensionaler euklidischer/unitärer Vektorraum mit einer Orthonormalbasis B und einer beliebigen Basis B'. Dann ist die Basis B' genau dann orthonormal, wenn die Basiswechselmatrix $T_B^{B'}$ orthogonal/unitär ist.*

Beweis. Sei $B' = (v'_1, \ldots, v'_n)$. Die Spalten von $T_B^{B'}$ sind dann die Koordinatenvektoren $[v'_i]_B = \varphi_B^{-1}(v'_i)$. Da B eine Orthonormalbasis ist, ist $\varphi_B \colon \mathbb{K}^n \xrightarrow{\sim} V$ eine Isometrie (siehe Beispiel 7.2.41). Damit gilt:

$$\langle v'_i, v'_j \rangle = \langle [v'_i]_B, [v'_j]_B \rangle.$$

Also ist die Basis B' genau dann orthonormal, wenn die Spalten von $T_B^{B'}$ ein Orthonormalsystem bilden, wie gewünscht. □

Beispiel 7.2.48 (lineare Isometrien von $\mathbb{R}$)**.** Die Gruppe $\mathrm{O}(1) \subset \mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^\times$ hat genau zwei Elemente, 1 und -1, denn: Ein beliebiger Automorphismus f von $\mathbb{R}$ hat die Form $f(x) = \lambda \cdot x$ mit einem $\lambda \in \mathbb{R}^\times$. Dann gilt $\langle f(x), f(y) \rangle = \lambda^2 \langle x, y \rangle$, und damit ist f genau dann eine Isometrie, wenn $\lambda^2 = 1$, d.h., wenn $\lambda = \pm 1$.

Beispiel 7.2.49 (lineare Isometrien von $\mathbb{R}^2$)**.** Normierte Vektoren in $\mathbb{R}^2$ bilden den *Einheitskreis* (die eindimensionale Sphäre)

$$S^1 = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\} = \left\{ \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} \;\middle|\; \alpha \in [0, 2\pi) \right\}.$$

Sei $A \in \mathrm{O}(2)$. Dann liegen beide Spalten von A in S^1, d.h., es gibt $\alpha, \beta \in [0, 2\pi)$, so dass

$$A = \begin{pmatrix} \cos\alpha & \cos\beta \\ \sin\alpha & \sin\beta \end{pmatrix}.$$

Auf der anderen Seite müssen die Zeilen von A auch in S^1 liegen. Es gibt dann genau zwei Möglichkeiten, $\cos\beta = \pm\sin\alpha$ und entsprechend $\sin\beta = \mp\cos\alpha$, und beide liefern eine orthogonale Matrix. Deswegen besteht die Gruppe $\mathrm{O}(2)$ genau aus den Matrizen

$$D(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{und} \quad S(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}$$

mit $\alpha \in [0, 2\pi)$. Die Matrix $D(\alpha)$ hat Determinante 1 und entspricht der Drehung um den Winkel α (siehe Beispiel 4.2.33). Die Matrix $S(\alpha)$ hat Determinante -1 und ist gleich dem Produkt $D(\alpha)\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, d.h., sie entspricht der Komposition der Spiegelung an der Geraden $\mathbb{R} \cdot e_1$ mit der Drehung um den Winkel α, welche gleich der Spiegelung an der Geraden $\mathbb{R} \cdot D(\frac{\alpha}{2})e_1$ ist. Diese Analyse zeigt, dass die linearen Isometrien von $\mathbb{R}^2$ genau die Drehungen und die Spiegelungen sind.

Beispiel 7.2.50 (lineare Isometrien von $\mathbb{C}$)**.** Unitäre 1×1-Matrizen sind einfach komplexe Zahlen mit Betrag 1:

$$\mathrm{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\} \subset \mathbb{C}^\times.$$

Unter der gewöhnlichen Identifikation von $\mathbb{C}$ mit $\mathbb{R}^2$, können wir $\mathrm{U}(1)$ mit dem Einheitskreis $S^1 \subset \mathbb{R}^2$ aud Beispiel 7.2.49 identifizieren.

Definition 7.2.51 (spezielle lineare Gruppen)**.** Sei $n \in \mathbb{N}$.

- Die *spezielle lineare Gruppe* über einem Körper K ist

  $$\mathrm{SL}_n(K) = \{A \in \mathrm{GL}_n(K) \mid \det(A) = 1\}.$$

  Sie ist eine Untergruppe von $\mathrm{GL}_n(K)$, da die Determinante $\det\colon \mathrm{GL}_n(K) \to K^\times$ ein Gruppenhomomorphismus ist (Bemerkung 5.3.28).

- Die *spezielle orthogonale Gruppe* ist

  $$\mathrm{SO}(n) = \mathrm{O}(n) \cap \mathrm{SL}_n(\mathbb{R}).$$

- Die *spezielle unitäre Gruppe* ist

  $$\mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}_n(\mathbb{C}).$$

Beispiel 7.2.52. Nach Beispiel 7.2.49 ist $\mathrm{SO}(2) = \{D(\alpha) \,|\, \alpha \in [0, 2\pi)\}$. Nach Beispiel 7.2.50 ist die Abbildung

$$\mathrm{SO}(2) \to \mathrm{U}(1), \quad D(\alpha) \mapsto \exp(i\alpha),$$

bijektiv, und zwar ein Gruppenisomorphismus.

Bemerkung 7.2.53 (allgemeine Isometrien)**.** Eine *Isometrie* zwischen metrischen Räumen ist eine abstandserhaltende bijektive Abbildung. Eine Isometrie zwischen euklidischen/unitären Vektorräumen ist nicht unbedingt linear: Ist $v_0 \in V \setminus \{0\}$, so ist die Verschiebung

$$V \to V, \quad v \mapsto v + v_0,$$

eine Isometrie, die nicht linear ist. Bei *euklidischen* Vektorräumen kann man leicht zeigen, dass jede Isometrie, die Null auf Null abbildet, automatisch linear ist. Anders gesagt ist jede Isometrie die Komposition einer *linearen* Isometrie mit einer Verschiebung. Deswegen genügt es in diesem Fall, lineare Isometrien zu verstehen, um allgemeine Isometrien zu verstehen. Im unitären Fall gilt die Aussage aber nicht: Zum Beispiel ist die komplexe Konjugation auf $\mathbb{C}$ eine Isometrie, die nicht $\mathbb{C}$-linear ist.

7.3 Der Spektralsatz

Der Spektralsatz ist ein Diagonalisierbarkeitskriterium für besondere Endomorphismen von endlich-dimensionalen $\mathbb{K}$-Vektorräumen. Im reellen Fall sagt er, dass jede symmetrische Matrix über $\mathbb{R}$ diagonalisierbar ist, und zudem dass man eine Orthonormalbasis von Eigenvektoren finden kann.

7.3.1 Selbstadjungierte Endomorphismen

Symmetrische Matrizen rühren von symmetrischen Bilinearformen her (Proposition 7.1.16), aber es ist nicht offensichtlich, was die Symmetrie einer Matrix für die zugehörige lineare Abbildung bedeutet (deshalb hatten wir den Begriff von symmetrischer Matrix bis jetzt nicht betrachtet). Sei $A \in M_n(\mathbb{R})$. Der Zusammenhang zwischen der Bilinearform b_A und der linearen Abbildung L_A ist folgender: Für alle $x, y \in \mathbb{R}^n$ gilt

$$b_A(x, y) = x^\intercal Ay = x^\intercal L_A(y) = \langle x, L_A(y)\rangle,$$

wobei $\langle -, -\rangle$ das Standardskalarprodukt ist. Deswegen ist die Matrix A genau dann symmetrisch, wenn für alle $x, y \in \mathbb{R}^n$ gilt:

$$\langle L_A(x), y\rangle = \langle x, L_A(y)\rangle.$$

Dies führt zum Begriff von *selbstadjungiertem Endomorphismus* eines euklidischen/unitären Vektorraums:

Definition 7.3.1 (selbstadjungierter Endomorphismus)**.** Sei V ein euklidischer/unitärer $\mathbb{K}$-Vektorraum. Ein Endomorphismus $f \in \mathrm{End}_{\mathbb{K}}(V)$ heißt *selbstadjungiert*, wenn für alle $v, w \in V$ gilt:

$$\langle f(v), w\rangle = \langle v, f(w)\rangle.$$

Bemerkung 7.3.2. Sei V ein euklidischer/unitärer $\mathbb{K}$-Vektorraum und sei $f \in \mathrm{End}_{\mathbb{K}}(V)$. Da das Skalarprodukt auf V nicht ausgeartet ist (Bemerkung 7.2.2), ist die induzierte lineare Abbildung

$$\begin{aligned} V &\to V^\dagger, \\ v &\mapsto \langle -, v\rangle, \end{aligned}$$

injektiv. Die adjungierte Abbildung

$$\begin{aligned} f^\dagger \colon V^\dagger &\to V^\dagger, \\ \alpha &\mapsto \alpha \circ f, \end{aligned}$$

bildet $\langle -, v\rangle$ auf $\langle f(-), v\rangle$ ab. Es folgt daraus, dass f genau dann selbstadjungiert ist, wenn folgendes Quadrat kommutiert:

$$\begin{array}{ccc} V & \hookrightarrow & V^\dagger \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f^\dagger} \\ V & \hookrightarrow & V^\dagger. \end{array}$$

Ist V endlich-dimensional, so ist die Abbildung $V \hookrightarrow V^\dagger$ sogar ein Isomorphismus, damit man f mit seiner adjungierten Abbildung $f^\dagger$ identifizieren kann. Das ist der Grund für den Namen „selbstadjungiert“.

Proposition 7.3.3 (Kriterium für selbstadjungierte Endomorphismen)**.** *Sei V ein endlich-dimensionaler euklidischer/unitärer $\mathbb{K}$-Vektorraum und sei B eine* Orthonormalbasis *von V. Ein Endomorphismus $f \in \mathrm{End}_{\mathbb{K}}(V)$ ist genau dann selbstadjungiert, wenn die Matrix $[f]_B^B$ hermitesch ist.*

Beweis. Sei $B = (v_1, \dots, v_n)$ und sei $[f]_B^B = (a_{ij})_{i,j}$. Nach Definition der Darstellungsmatrix gilt:

$$f(v_i) = \sum_{k=1}^{n} a_{ki} v_k.$$

Da B ein Erzeugendensystem ist, ist der Endomorphismus f genau dann selbstadjungiert, wenn $\langle f(v_i), v_j \rangle = \langle v_i, f(v_j) \rangle$ für alle $i, j \in \{1, \dots, n\}$. Wegen der Orthonormalität von B gilt:

$$\langle f(v_i), v_j \rangle = \sum_{k=1}^{n} \overline{a_{ki}} \langle v_k, v_j \rangle = \overline{a_{ji}},$$
$$\langle v_i, f(v_j) \rangle = \sum_{k=1}^{n} a_{kj} \langle v_i, v_k \rangle = a_{ij}.$$

Deswegen ist f genau dann selbstadjungiert, wenn $a_{ij} = \overline{a_{ji}}$ für alle i, j, d.h., wenn $[f]_B^B$ hermitesch ist. □

Bemerkung 7.3.4. Eine quadratische Matrix $A \in M_n(\mathbb{K})$ ist genau dann hermitesch, wenn der Endomorphismus $L_A \colon \mathbb{K}^n \to \mathbb{K}^n$ selbstadjungiert ist, d.h., wenn für alle $x, y \in \mathbb{K}^n$ gilt

$$\langle Ax, y \rangle = \langle x, Ay \rangle.$$

Das ist ein Sonderfall der Proposition 7.3.3 (mit der Standardbasis von $\mathbb{K}^n$), und kann auch leicht direkt nachgeprüft werden.

Lemma 7.3.5 (Eigenwerte sind reell)**.** *Sei V ein unitärer Vektorraum und $f \in \mathrm{End}_{\mathbb{C}}(V)$ ein selbstadjungierter Endomorphismus. Dann sind alle Eigenwerte von f reell.*

Beweis. Sei v ein Eigenvektor von f zum Eigenwert $\lambda \in \mathbb{C}$. Dann gilt:

$$\lambda \langle v, v \rangle = \langle v, f(v) \rangle = \langle f(v), v \rangle = \bar{\lambda} \langle v, v \rangle.$$

Da $v \neq 0$ ist $\langle v, v \rangle \neq 0$ und somit $\lambda = \bar{\lambda}$, d.h., $\lambda \in \mathbb{R}$. □

Korollar 7.3.6 (Orthogonalität von Eigenvektoren)**.** *Sei V ein endlich-dimensionaler euklidischer/unitärer $\mathbb{K}$-Vektorraum und sei $f \in \mathrm{End}_{\mathbb{K}}(V)$ ein selbstadjungierter Endomorphismus. Sind $v, w \in V$ Eigenvektoren von f zu verschiedenen Eigenwerten λ, μ, so gilt $v \perp w$.*

Beweis. Es gilt:

$$\bar{\lambda} \langle v, w \rangle = \langle \lambda v, w \rangle = \langle f(v), w \rangle = \langle v, f(w) \rangle = \langle v, \mu w \rangle = \mu \langle v, w \rangle.$$

Nach Lemma 7.3.5 ist $\lambda \in \mathbb{R}$, und somit $\bar{\lambda} = \lambda \neq \mu$. Die obige Gleichung ist dann nur möglich, wenn $\langle v, w \rangle = 0$. □

Satz 7.3.7 (Spektralsatz)**.** *Sei V ein endlich-dimensionaler euklidischer/unitärer $\mathbb{K}$-Vektorraum und sei $f \in \mathrm{End}_{\mathbb{K}}(V)$ ein selbstadjungierter Endomorphismus. Dann existiert eine Orthonormalbasis von V bestehend aus Eigenvektoren von f. Insbesondere ist f diagonalisierbar.*

Beweis. Wir verwenden Induktion über die Dimension $n = \dim_{\mathbb{K}} V$. Falls $n = 0$ leistet die leere Basis das Gewünschte. Sei also $n \geq 1$.

Behauptung. f besitzt mindestens einen Eigenvektor.

Falls $\mathbb{K} = \mathbb{C}$ folgt die Behauptung aus dem Fundamentalsatz der Algebra, denn das charakteristische Polynom $\chi_f \in \mathbb{C}[T]$ besitzt mindestens eine Nullstelle. Um den Fall $\mathbb{K} = \mathbb{R}$ zu behandeln, nehmen wir eine Orthonormalbasis B von V und betrachten wir die Darstellungsmatrix $A = [f]_B^B \in M_n(\mathbb{R})$. Es gilt $\chi_f = \chi_A$ nach Definition. Da f selbstadjungiert ist, ist die Matrix A symmetrisch (Proposition 7.3.3). Wir betrachten jetzt A als hermitesche Matrix über $\mathbb{C}$, die einen selbstadjungierten Endomorphismus L_A von $\mathbb{C}^n$ induziert. Nach Lemma 7.3.5 sind alle Eigenwerte von A reell, d.h., die Nullstellen von $\chi_f = \chi_A$ in $\mathbb{C}$ sind reell. Insbesondere besitzt χ_f eine reelle Nullstelle, d.h., f besizt einen Eigenvektor, wie behauptet.

Sei also $v \in V$ ein Eigenvektor von f zum Eigenwert $\lambda \in \mathbb{K}$, und sei $U = (\mathbb{K}v)^\perp$ der Orthogonalraum der Geraden $\mathbb{K}v$. Dann ist U f-invariant, d.h., $f(U) \subset U$, denn für alle $u \in U$ gilt

$$\langle f(u), v\rangle = \langle u, f(v)\rangle = \lambda\langle u, v\rangle = 0.$$

Sei $f_U \in \mathrm{End}_{\mathbb{K}}(U)$ die Einschränkung von f auf U. Bezüglich des auf U eingeschränkten Skalarproduktes ist dann f_U ein selbstadjungierter Endomorphismus von U. Da U zu $\mathbb{K}v$ komplementär ist (Korollar 7.2.36(i)), hat U die Dimension $n-1$. Wir dürfen also die Induktionsvoraussetzung anwenden: Es gibt eine Orthonormalbasis $(v_1, \ldots, v_{n-1})$ von U, in der jedes v_i ein Eigenvektor von f ist. Setzt man

$$v_n := \frac{1}{\|v\|} v,$$

so ist die Familie $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht. □

Bemerkung 7.3.8. In unserem Beweis des Spektralsatzes haben wir den Fundamentalsatz der Algebra (dass $\mathbb{C}$ algebraisch abgeschlossen ist) verwendet, den wir noch nicht bewiesen haben. Dieser Satz wird erstmal in der Vorlesung *Analysis III* mithilfe von der Funktionentheorie bewiesen werden.

Korollar 7.3.9 (Diagonalisierbarkeit von hermiteschen bzw. symmetrischen Matrizen)**.** *Sei $n \in \mathbb{N}$.*

(i) *Sei $A \in M_n(\mathbb{C})$ eine hermitesche Matrix. Dann gibt es eine unitäre Matrix $S \in \mathrm{U}(n)$, so dass $S^{-1}AS = S^{\mathsf{H}}AS$ eine Diagonalmatrix ist.*

(ii) *Sei $A \in M_n(\mathbb{R})$ eine symmetrische Matrix. Dann gibt es eine orthogonale Matrix $S \in \mathrm{O}(n)$, so dass $S^{-1}AS = S^{\mathsf{T}}AS$ eine Diagonalmatrix ist.*

Beweis. Sei $A \in M_n(\mathbb{K})$ hermitesch. Nach Bemerkung 7.3.4 ist der Endomorphismus L_A von $\mathbb{K}^n$ selbstadjungiert. Nach dem Spektralsatz existiert eine Orthonormalbasis B von $\mathbb{K}^n$, so dass $[L_A]_B^B$ eine Diagonalmatrix ist. Ist $S = T_E^B$ die Basiswechselmatrix, so gilt $[L_A]_B^B = S^{-1}AS$ nach der Basiswechselformel für lineare Abbildungen (Proposition 4.2.43). Die Spalten von S sind die Vektoren aus B, die ein Orthonormalsystem bilden. Nach Proposition 7.2.43 ist also S eine orthogonale (falls $\mathbb{K} = \mathbb{R}$) oder unitäre (falls $\mathbb{K} = \mathbb{C}$) Matrix. □

Rezept 7.3.10 (Diagonalisierung durch eine orthogonale/unitäre Matrix)**.** Gegeben sei eine hermitesche Matrix $A \in M_n(\mathbb{K})$ und ihre Eigenwerte $\lambda_1, \ldots, \lambda_k$. Gesucht ist eine orthogonale (falls $\mathbb{K} = \mathbb{R}$) oder unitäre (falls $\mathbb{K} = \mathbb{C}$) Matrix S, so dass $S^{-1}AS$ eine Diagonalmatrix ist. Man findet zunächst Basen der Eigenräume $\mathrm{Eig}_{\lambda_i}(A)$ mit Rezept 6.2.19, und man führt danach das Orthonormalisierungsverfahren von Gram-Schmidt durch, um Orthonormalbasen der Eigenräume zu erhalten. Setzt man diese Basen zusammen, so erhält man eine Orthonormalbasis $B = (v_1, \ldots, v_n)$ von $\mathbb{K}^n$ (nach Korollar 7.3.6) bestehend aus Eigenvektoren von A. Die Basiswechselmatrix $S = T_E^B = \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}$ hat die gewünschten Eigenschaften.

Beispiel 7.3.11.

(i) Sei A die symmetrische Matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} \in M_3(\mathbb{R}).$$

Ihr charakteristisches Polynom $\chi_A \in \mathbb{R}[T]$ ist $(T-1)(T^2-1) = (T-1)^2(T+1)$. Die Eigenwerte von A sind also 1 und -1 mit geometrischen Vielfachheiten 2 und 1 (da A diagonalisierbar sein muss). Man kann hier gleich bemerken, dass e_1 und $e_2 - e_3$ Eigenvektoren zu 1 sind, und dass $e_2 + e_3$ ein Eigenvektor zu -1 ist. Die Vektoren e_1 und $e_2 - e_3$ sind bereits orthogonal, und damit ist $\left(e_1, \frac{1}{\sqrt{2}}(e_2 - e_3)\right)$ eine Orthonormalbasis von $\mathrm{Eig}_1(A)$. Der Vektor $\frac{1}{\sqrt{2}}(e_2 + e_3)$ bildet eine Orthonormalbasis von $\mathrm{Eig}_{-1}(A)$. Setzt man

$$S = \begin{pmatrix} e_1 & \frac{1}{\sqrt{2}}(e_2 - e_3) & \frac{1}{\sqrt{2}}(e_2 + e_3) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

so ist S eine orthogonale Matrix, so dass $S^{-1}AS = \mathrm{diag}(1, 1, -1)$.

(ii) Sei A die hermitesche Matrix

$$A = \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 1+i \\ -i & 1-i & -1 \end{pmatrix} \in M_3(\mathbb{C}).$$

Ihr charakteristisches Polynom ist

$$\begin{aligned} \chi_A = \det \begin{pmatrix} T-1 & 0 & -i \\ 0 & T-1 & -1-i \\ i & -1+i & T+1 \end{pmatrix} &\overset{1.\mathrm{S}}{=} (T-1)(T^2-1-2) + i(i(T-1)) \\ &= (T-1)(T^2-4) = (T-1)(T-2)(T+2). \end{aligned}$$

Die Eigenwerte von A sind also 1, 2 und -2. Mit dem Gaußschen Eliminationsverfahren findet man zugehörige Eigenvektoren

$$v_1 = \begin{pmatrix} 1+i \\ -1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} i \\ 1+i \\ 1 \end{pmatrix} \quad \text{und} \quad v_3 = \begin{pmatrix} i \\ 1+i \\ -3 \end{pmatrix},$$

die automatisch zueinander orthogonal sind (nach Korollar 7.3.6). Ihre Normierungen bilden eine unitäre Matrix

$$S = \begin{pmatrix} \frac{1}{\sqrt{3}}v_1 & \frac{1}{2}v_2 & \frac{1}{2\sqrt{3}}v_3 \end{pmatrix},$$

für die gilt $S^{-1}AS = \mathrm{diag}(1, 2, -2)$.

Korollar 7.3.12 (Diagonalisierbarkeit von hermiteschen Formen)**.** *Sei V ein endlich-dimensionaler euklidischer/unitärer $\mathbb{K}$-Vektorraum und sei s eine hermitesche Form auf V. Dann existiert eine Orthonormalbasis B von V, so dass die Matrix $[s]_{B,B}$ diagonal ist.*

Beweis. Sei B' eine beliebige Orthonormalbasis von V (Korollar 7.2.34). Nach Proposition 7.1.40 ist die Matrix $A = [s]_{B',B'}$ hermitesch. Nach Korollar 7.3.9 gibt es eine orthogonale/unitäre Matrix S, so dass $S^{\mathsf{H}}AS$ diagonal ist. Sei B die Basis von V, so dass $T^B_{B'} = S$ (d.h., die Spalten von S sind die Koordinatenvektoren bzgl. B' der Vektoren aus B). Nach der Basiswechselformel für Sesquilinearformen (Proposition 7.1.37) ist $[s]_{B,B} = S^{\mathsf{H}}AS$. Außerdem ist B wieder eine Orthonormalbasis nach Proposition 7.2.47. □

Bemerkung 7.3.13. Das Korollar 7.3.12 impliziert insbesondere, dass jede hermitesche Form s auf einem endlich-dimensionalen $\mathbb{K}$-Vektorraum V „diagonalisierbar“ ist, d.h., dass eine Basis B von V existiert, so dass die Darstellungsmatrix $[s]_{B,B}$ diagonal ist (denn man kann eine euklidische/unitäre Struktur auf V mithilfe eines Isomorphismus $V \xrightarrow{\sim} \mathbb{K}^n$ definieren). Diese Aussage ist aber elementarer und gilt eigentlich für einen beliebigen Grundkörper mit Involution (K, σ), sofern $\operatorname{char} K \neq 2$.

7.3.2 Definitheitskriteria

Proposition 7.3.14 (Definitheit und Eigenwerte)**.** *Sei $n \in \mathbb{N}$, sei $A \in M_n(\mathbb{C})$ eine hermitesche Matrix (z.B. eine symmetrische Matrix über $\mathbb{R}$) und sei $\sigma(A) \subset \mathbb{R}$ das Spektrum von A (d.h., die Menge aller Eigenwerte von A).*

(i) *A ist genau dann positiv semidefinit, wenn $\sigma(A) \subset \mathbb{R}_{\geq 0}$.*

(ii) *A ist genau dann negativ semidefinit, wenn $\sigma(A) \subset \mathbb{R}_{\leq 0}$.*

(iii) *A ist genau dann positiv definit, wenn $\sigma(A) \subset \mathbb{R}_{>0}$.*

(iv) *A ist genau dann negativ definit, wenn $\sigma(A) \subset \mathbb{R}_{<0}$.*

Beweis. Nach dem Korollar 7.3.9 gibt es eine unitäre Matrix $S \in \mathrm{U}(n)$ mit $S^{-1}AS = S^{\mathsf{H}}AS = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, wobei $\{\lambda_1, \ldots, \lambda_n\} = \sigma(A)$. Für $x \in \mathbb{K}^n$ gilt:

$$s_A(x,x) = x^{\mathsf{H}}Ax = (S^{-1}x)^{\mathsf{H}}(S^{\mathsf{H}}AS)(S^{-1}x) = \sum_{i=1}^{n} \lambda_i |(S^{-1}x)_i|^2.$$

Insbesondere ist

$$s_A(Se_i, Se_i) = \lambda_i.$$

Aus diesen zwei Gleichungen schließen wir die gewünschten Äquivalenzen. □

Sei A eine $n \times n$-Matrix. Für eine Teilmenge $I \subset \{1, \ldots, n\}$ mit k Elemente schreibt man $A(I)$ für die $k \times k$-Untermatrix von A bestehend aus den Einträgen mit Indizes in $I \times I$. Wenn $I = \{1, \ldots, k\}$ schreibt man auch $A(k)$ anstelle von $A(I)$. Ist A symmetrisch oder hermitesch, so ist auch jedes $A(I)$. Die Determinanten der Matrizen $A(I)$ heißen die *Hauptminoren* von A, und folgendes Kriterium wird auch *Hauptminorenkriterium* genannt.

Proposition 7.3.15 (Kriterium von Sylvester)**.** *Sei $n \in \mathbb{N}$ und sei $A \in M_n(\mathbb{C})$ eine hermitesche Matrix.*

(i) *A ist genau dann positiv definit, wenn $\det A(k) \in \mathbb{R}_{>0}$ für alle $k \in \{1, \ldots, n\}$. Außerdem gilt dann $\det A(I) \in \mathbb{R}_{>0}$ für alle $I \subset \{1, \ldots, n\}$.*

(ii) *A ist genau dann negativ definit, wenn $(-1)^k \det A(k) \in \mathbb{R}_{>0}$ für alle $k \in \{1, \ldots, n\}$. Außerdem gilt dann $(-1)^{|I|} \det A(I) \in \mathbb{R}_{>0}$ für alle $I \subset \{1, \ldots, n\}$.*

Beweis. Die zweite Aussage ist äquivalent zu der ersten mit der Matrix $-A$.

Sei A positiv definit und sei $I \subset \{1, \ldots, n\}$. Man identifiziert den Vektorraum $\mathbb{C}^I$ mit dem Untervektorraum von $\mathbb{C}^n$ bestehend aus Vektoren x mit $x_j = 0$ für alle $j \notin I$. Die Sesquilinearform $s_{A(I)}$ auf $\mathbb{C}^I$ ist dann die Einschränkung von s_A und ist damit positiv definit. Nach Proposition 7.3.14 sind alle Eigenwerte von $A(I)$ positiv, und aus Proposition 6.2.36 folgt, dass $\det A(I) \in \mathbb{R}_{>0}$.

Die Umkehrung beweisen wir durch Induktion über n. Falls $n = 0$ gibt es nichts zu zeigen. Sei also $n \geq 1$ und sei $\det A(k) \in \mathbb{R}_{>0}$ für alle $k \in \{1, \ldots, n\}$. Zu zeigen ist, dass $A = (a_{ij})_{i,j}$ positiv definit ist. Es gilt insbesondere $a_{11} \in \mathbb{R}_{>0}$. Durch elementare Zeilenumformungen

und Spaltenumformungen können wir alle Koeffizienten a_{i1} mit $i > 1$ und a_{1j} mit $j > 1$ zu Null machen. Genauer sei

$$S = A_{21}(-a_{21}/a_{11}) \cdot \ldots \cdot A_{n1}(-a_{n1}/a_{11}) \in \mathrm{GL}_n(\mathbb{C}).$$

Dann hat die Matrix $A' = SAS^{\mathsf{H}}$ die Form

$$A' = \begin{pmatrix} a_{11} & 0 \\ 0 & C \end{pmatrix}.$$

Außerdem gilt $\det(A'(k)) = \det(A(k))$ für alle $k \in \{1, \ldots, n\}$, denn diese Zeilen- und Spaltenumformungen ändern die Determinanten dieser Untermatrizen nicht. Für $l \in \{1, \ldots, n-1\}$ ist $\det(C(l)) = a_{11}^{-1} \det(A'(l+1)) \in \mathbb{R}_{>0}$ nach dem Laplaceschen Entwicklungssatz. Nach Induktionsvoraussetzung ist die Matrix C positiv definit. Sei $x = \binom{x_1}{y} \in \mathbb{C}^n \setminus \{0\}$ mit $x_1 \in \mathbb{C}$ und $y \in \mathbb{C}^{n-1}$. Dann ist

$$s_{A'}(x, x) = x^{\mathsf{H}} A' x = a_{11}|x_1|^2 + s_C(y, y) > 0.$$

Also ist A' positiv definit, d.h., die hermitesche Form $s_{A'}$ ist positiv definit. Nach Proposition 7.1.49 sind aber die Paare $(\mathbb{C}^n, s_A)$ und $(\mathbb{C}^n, s_{A'})$ isomorph, da A und A' kongruent sind. Damit ist auch s_A positiv definit, wie gewünscht. □

Beispiel 7.3.16. Die symmetrische Matrix

$$A = \begin{pmatrix} -1 & 2 & 0 \\ 2 & -5 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

ist negativ definit, denn $\det A(1) = -1$, $\det A(2) = 1$ und $\det A = -1$.

Bemerkung 7.3.17. Man kann auch zeigen, dass eine hermitesche Matrix A genau dann positiv semidefinit ist, wenn $\det A(I) \in \mathbb{R}_{\geq 0}$ für alle $I \subset \{1, \ldots, n\}$. In diesem Fall ist es aber *nicht* hinreichend, dass $\det A(k) \in \mathbb{R}_{\geq 0}$ für alle $k \in \{1, \ldots, n\}$. Zum Beispiel ist die Matrix

$$A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$$

nicht postiv semidefinit, doch sie erfüllt $\det A(1) = \det A(2) = 0$.

Bemerkung 7.3.18 (symmetrische Bilinearformen in der Analysis)**.** Symmetrische multilinearformen sind wichtig in der Analysis mehrerer Veränderlicher. Sei $f \colon U \to \mathbb{R}$ eine k-mal stetig differenzierbare Funktion, wobei $U \subset \mathbb{R}^n$ eine offene Teilmenge ist, und sei $a \in U$. Die k-te Ableitung $D^k f(a)$ von f in a ist dann eine symmetrische k-lineare Form auf dem Tangentialraum $T_a U$, den man mit $\mathbb{R}^n$ identifizieren kann. Man kann insbesondere $D^2 f(a)$ als symmetrische $n \times n$-Matrix über $\mathbb{R}$ auffassen, nämlich die *Hesse-Matrix*

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{i,j}.$$

Falls a ein kritischer Punkt von f ist (d.h., die erste Ableitung $Df(a)$ verschwindet), kann man in den meisten Fällen mithilfe der Hesse-Matrix bestimmen, ob a ein lokales Maximum oder Minimum ist, und zwar: Ist die Form $D^2 f(a)$ positiv bzw. negativ definit, so ist a ein lokales Minimum bzw. Maximum von f auf U. Dabei ist das Kriterium von Sylvester besonders hilfreich.

7.3.3 Der Trägheitssatz von Sylvester

Der Trägheitssatz von Sylvester ist eine Klassifikation von symmetrischen Bilinearformen auf endlich-dimensionalen $\mathbb{R}$-Vektorräumen.

Satz 7.3.19 (Trägheitssatz von Sylvester)**.** *Sei V ein $\mathbb{R}$-Vektorraum der endlichen Dimension n und sei $b\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf V. Dann gibt es eine Basis B von V und natürliche Zahlen $n_+, n_-, n_0 \in \mathbb{N}$ mit $n_+ + n_- + n_0 = n$, so dass*

$$[b]_{B,B} = \begin{pmatrix} I_{n_+} & 0 & 0 \\ 0 & -I_{n_-} & 0 \\ 0 & 0 & 0_{n_0} \end{pmatrix} \in M_n(\mathbb{R}).$$

Außerdem ist das Tripel (n_+, n_-, n_0) eindeutig durch b bestimmt.

Definition 7.3.20 (Signatur)**.** Sei V ein $\mathbb{R}$-Vektorraum der endlichen Dimension n und sei $b\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf V. Das Tripel (n_+, n_-, n_0) vom Satz 7.3.19 heißt die *Signatur* von b.

Bemerkung 7.3.21. Sei $b\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf einem endlich-dimensionalen $\mathbb{R}$-Vektorraum V mit Signatur (n_+, n_-, n_0). Dann:

- b ist genau dann nicht ausgeartet, wenn $n_0 = 0$.
- b ist genau dann positiv bzw. negativ semidefinit, wenn $n_- = 0$ bzw. $n_+ = 0$.
- b ist genau dann positiv bzw. negativ definit, wenn $n_0 = n_- = 0$ bzw. $n_0 = n_+ = 0$.

Beweis vom Satz 7.3.19. Nach Korollar 7.3.9 gibt es eine Basis $C = (w_1, \dots, w_n)$ von V, so dass

$$[b]_{C,C} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

mit $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Sei $n_+/n_-/n_0$ die Anzahl der Indizes $i \in \{1, \dots, n\}$, so dass λ_i positiv/negativ/null ist. Durch Permutation der Basisvektoren können wir annehmen, dass die ersten n_+ λ_i positiv sind, die nächsten n_- negativ sind, und die letzten n_0 null sind. Sei nun $B = (v_1, \dots, v_n)$, wobei

$$v_i = \begin{cases} \frac{1}{\sqrt{|\lambda_i|}} w_i, & \text{falls } i \leq n_+ + n_-, \\ w_i, & \text{falls } i > n_+ + n_-. \end{cases}$$

Dann hat die Darstellungsmatrix $[b]_{B,B}$ die gewünschte Form.

Es bleibt zu zeigen, dass das Tripel (n_+, n_-, n_0) eindeutig bestimmt ist. Hierzu betrachten wir die natürlichen Zahlen

$$m_+ = \max\{\dim_{\mathbb{R}} U \mid U \subset V \text{ ist ein Untervektorraum, so dass } b|_{U\times U} \text{ positiv definit ist}\},$$
$$m_- = \max\{\dim_{\mathbb{R}} U \mid U \subset V \text{ ist ein Untervektorraum, so dass } b|_{U\times U} \text{ negativ definit ist}\},$$

die offensichtich nur von b abhängen. Wir behaupten, dass $n_+ = m_+$ und $n_- = m_-$. Da die Einschränkung von b auf $\operatorname{Span}_{\mathbb{R}}\{v_1, \dots, v_{n_+}\}$ positiv definit ist, gilt $n_+ \leq m_+$. Sei umgekehrt $U \subset V$ ein Untervektorraum, so dass $b|_{U\times U}$ positiv definit ist. Nach Wahl der Basis B ist die Einschränkung von b auf $\operatorname{Span}_{\mathbb{R}}\{v_{n_+ +1}, \dots, v_n\}$ negativ semidefinit. Auf dem Untervektorraum $U \cap \operatorname{Span}_{\mathbb{R}}\{v_{n_+ +1}, \dots, v_n\}$ ist daher b gleichzeitig positiv definit und negativ semidefinit. Das ist aber nur möglich, wenn dieser Durchschnitt trivial ist. Aus der Dimensionsformel für Untervektorräume folgt jetzt, dass $\dim_{\mathbb{R}} U \leq n_+$. Also gilt auch $m_+ \leq n_+$. Der Beweis von $n_- = m_-$ geht analog, und schließlich ist $n_0 = n - m_+ - m_-$. □

Beispiel 7.3.22 (Lorentzsche Formen)**.** Die Abbildung

$$\begin{aligned} b\colon \mathbb{R}^4 \times \mathbb{R}^4 &\to \mathbb{R}, \\ (x, y) &\mapsto x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4 \end{aligned}$$

ist eine nicht ausgeartete symmetrische Bilinearform auf $\mathbb{R}^4$, mit Signatur $(3, 1, 0)$. Solche Bilinearformen (und allgemeiner Bilinearformen mit Signatur $(n, 1, 0)$) heißen *Lorentzsche Formen*. Diese Form ist wichtig in der Relativitätstheorie, indem man $\mathbb{R}^4$ als die Raumzeit auffasst. Vektoren $v, w \in \mathbb{R}^4$ mit $b(v - w, v - w) \geq 0$ entsprechen Ereignisse, die kausal zusammenhängend sind.

Bemerkung 7.3.23 (Berechnung der Signatur)**.** Sei $A \in M_n(\mathbb{R})$ eine symmetrische Matrix und sei (n_+, n_-, n_0) die Signatur der Bilinearform b_A. Dann ist n_+ bzw. n_- die Anzahl der positiven bzw. negativen Eigenwerte von A mit algebraischer Vielfachheit gezählt. Denn nach dem Beweis des Trägheitssatz von Sylvester ist n_+ bzw. n_- die Anzahl der positiven bzw. negativen Einträge in einer diagonalen Darstellungsmatrix von b_A, und nach Korollar 7.3.12 gibt es eine Orthonormalbasis B, so dass die Matrix $[b_A]_{B,B}$ diagonal ist. Die Basiswechselmatrix $S = T_E^B$ ist dann orthogonal, so dass $[b_A]_{B,B} = S^\intercal AS = S^{-1}AS$, und deswegen sind die Diagonalkoeffizienten von $[b_A]_{B,B}$ die Eigenwerte von A.

Kapitel 8

Moduln über Hauptidealringen

Kommutative Ringe sind eine Abschwächung von Körpern, in denen es keine multiplikative Inverse geben muss (siehe Bemerkung 2.3.5). Das prototypische Beispiel ist der kommutative Ring $\mathbb{Z}$ der ganzen Zahlen, in dem nur 1 und -1 multiplikative Inverse besitzen.

Ringe sind eine weitere Abschwächung von kommutativen Ringen, in denen die Multiplikation nicht kommutativ sein muss. Ein typisches Beispiel ist der Ring $M_n(K)$ von $n \times n$-Matrizen, der nicht kommutativ ist wenn $n \geq 2$ (siehe Bemerkung 4.2.17). Manchmal verzichtet man auch auf die Assoziativität der Multiplikation oder auf die Existenz des neutralen Elements 1, aber man spricht dann eher von nicht-assoziativen oder nicht-unitären Ringen.

Man kann nun bemerken, dass die Definition von Vektorraum über einem Körper K (Definition 3.2.1) sinnvoll bleibt, wenn der Körper K durch einen beliebigen Ring R ersetzt wird (das heißt, die Existenz von multiplikativen Inversen und die Kommutativität der Multiplikation spielen keine Rollen in dieser Definition). Man spricht dann von einem *Modul* über R. Ein Modul über einem Körper K ist dann genau ein K-Vektorraum. Moduln über ein paar anderen Ringen sind uns auch schon bekannt: Moduln über $\mathbb{Z}$ sind genau abelsche Gruppen, und Moduln über dem Polynomring $K[T]$ sind Paare (V, f) bestehend aus einem K-Vektorraum und einem Endomorphismus f von V.

Viele grundlegende Begriffe der Theorie der Vektorräume lassen sich auf Moduln übertragen (Erzeugendensystem, Basis, lineare Abbildung, usw.). Aber die Modultheorie über beliebigen Ringen ist bestimmt komplizierter als die von Vektorräumen, und viele Aussagen über Vektorräume gelten nicht für Moduln. Zum Beispiel existieren Moduln die keine Basis besitzen, und damit können lineare Abbildungen zwischen Moduln nicht mit Matrizen dargestellt werden.

Ein weiteres neues Phänomen bei Ringen ist, dass die *Teilbarkeitsrelation* interessant ist (der Teilbarkeitsrelation auf $\mathbb{Z}$ und auf $K[T]$ haben wir schon begegnet, siehe Abschnitt 1.1.1 und Definition 6.3.11). Dies führt unter anderem zum Begriff von *Primelement*, der für Körper inhaltslos ist:

Ring R	Moduln über R	Primelemente von R
Körper K	K-Vektorräume	keine
die ganzen Zahlen $\mathbb{Z}$	abelsche Gruppen	$\pm$Primzahlen
Polynomring $K[T]$	Endomorphismen von K-Vektorräumen	irreduzible Polynome

Alle Ringe in dieser Tabelle sind *Hauptidealringen*. Körper sind im gewissen Sinne die einfachsten Ringe (z.B. unter dem Gesichtpunkt ihrer Modultheorie), und Hauptidealringe bilden die nächste einfachste Klasse von Ringen. Der Hauptsatz dieses Kapitels ist eine vollständige Klassifikation von *endlich erzeugten Moduln* über Hauptidealringen (Satz 8.3.30).

Insbesondere werden wir endlich erzeugte abelsche Gruppen klassifizieren. Für den Hauptidealring $K[T]$ von Polynomen über einem Körper K enthält diese Klassifikation insbesondere eine Klassifikation von Endomorphismen von endlich-dimensionalen K-Vektorräumen bis auf Isomorphie (d.h., von quadratischen Matrizen über K bis auf Ähnlichkeit, siehe Abschnitt 6.1.3). Diese Anwendung werden wir im Kapitel 9 näher untersuchen.

8.1 Ringe und Moduln

8.1.1 Ringe

Definition 8.1.1 (Ring, kommutativer Ring)**.** Ein *Ring* ist ein Tripel $(R, +, \cdot)$ bestehend aus einer Menge R und zwei Verknüpfungen

$$+\colon R \times R \to R, \quad \cdot\colon R \times R \to R,$$

die als *Addition* und *Multiplikation* bezeichnet werden, mit folgenden Eigenschaften:

(i) $(R, +)$ ist eine abelsche Gruppe. Das neutrale Element bzgl. $+$ wird mit 0 bezeichnet.

(ii) Die Multiplikation ist assoziativ und besitzt ein neutrales Element 1, d.h., für alle $x, y, z \in R$ gilt:
$$x \cdot (y \cdot z) = (x \cdot y) \cdot z \quad \text{und} \quad 1 \cdot x = x = x \cdot 1.$$

(iii) Es gilt das Distributivgesetz, d.h., für alle $x, y, z \in R$ gilt:

$$x \cdot (y + z) = x \cdot y + x \cdot z \quad \text{und} \quad (x + y) \cdot z = x \cdot z + y \cdot z.$$

Das Tripel $(R, +, \cdot)$ heißt *kommutativer Ring*, wenn folgende zusätzliche Eigenschaft gilt:

(iv) Die Multiplikation ist kommutativ, d.h., für alle $x, y \in R$ gilt:

$$x \cdot y = y \cdot x.$$

Beispiel 8.1.2.

(i) Jeder Körper ist ein kommutativer Ring. Ein kommutativer Ring $(R, +, \cdot)$ ist genau dann ein Körper, wenn $0 \neq 1$ und jedes $x \in R \setminus \{0\}$ ein inverses Element x^{-1} bzgl. $\cdot$ besitzt.

(ii) Der *Nullring* ist die Menge $\{0\}$ mit den einzig möglichen Verknüpfungen $+$ und $\cdot$. Er ist auf triviale Weise ein kommutativer Ring, in dem $0 = 1$. Ist umgekehrt R ein Ring, in dem $0 = 1$, so ist $R = \{0\}$, denn es gilt $x = 1 \cdot x = 0 \cdot x = 0$ für alle $x \in R$ (nach Proposition 8.1.5(i)).

(iii) Die ganzen Zahlen $\mathbb{Z}$ bilden einen kommutativen Ring mit der gewöhnlichen Addition bzw. Multiplikation.

(iv) Der Polynomring $K[T]$ über einem Körper K ist ein kommutativer Ring (Proposition 6.3.4).

(v) Sei K ein Körper und $n \in \mathbb{N}$. Dann ist $M_n(K)$ ein Ring mit der Addition und der Multiplikation von Matrizen (Proposition 4.2.16). Wenn $n \geq 2$ ist dieser Ring nicht kommutativ (siehe Bemerkung 4.2.17).

(vi) Ist K ein Körper und V ein K-Vektorraum, so ist $(\mathrm{End}_K(V), +, \circ)$ ein Ring. Wenn $\dim_K V \geq 2$ ist er nicht kommutativ. Wenn $V = K^n$ können wir diesen Ring mit $M_n(K)$ identifizieren.

(vii) Eine *Gaußsche Zahl* ist eine komplexe Zahl $a + bi$ mit $a, b \in \mathbb{Z}$; man bezeichnet mit $\mathbb{Z}[i] \subset \mathbb{C}$ die Menge aller Gaußschen Zahlen. Für alle $z, w \in \mathbb{Z}[i]$ sind $z + w$, $-z$ und $z \cdot w$ wieder Gaußsche Zahlen, wie man leicht nachrechnen kann. Es folgt daraus, dass $(\mathbb{Z}[i], +, \cdot)$ ein kommutativer Ring ist (und zwar ein Unterring von $\mathbb{C}$, siehe Proposition 8.1.8).

(viii) Für alle $n \in \mathbb{N} \setminus \{0\}$ ist die Menge $\mathbb{Z}/n\mathbb{Z}$ der Restklassen modulo n ein kommutativer Ring mit n Elementen (siehe Abschnitt 2.4.4). Er ist genau dann ein Körper, wenn n eine Primzahl ist.

(ix) Das Tripel $(\mathbb{N}, +, \cdot)$ ist kein Ring: Es erfüllt alle Axiome für einen kommutativen Ring, außer der Existenz von Inversen bzgl. $+$ (d.h., $(\mathbb{N}, +)$ ist nur eine abelsche Monoid und keine Gruppe).

Beispiel 8.1.3 (Polynome über einem Ring)**.** Sei R ein Ring. Ein *Polynom* über R in einer Variablen T ist ein Ausdruck der Gestalt $\sum_{i=0}^{n} a_i T^i$ mit $n \in \mathbb{N}$ und $a_i \in R$. Addition und Multiplikation von Polynomen über R können wie im Abschnitt 6.3.1 definiert werden und liefern einen Ring $R[T]$. Ist R kommutativ, so ist auch $R[T]$ kommutativ. Zum Beispiel gibt es den Ring $\mathbb{Z}[T]$ von Polynomen mit ganzen Koeffizienten. Diese Konstruktion kann man auch iterieren: Man definiert rekursiv

$$R[T_1, \ldots, T_n] = R[T_1, \ldots, T_{n-1}][T_n],$$

und man bezeichnet Elemente von $R[T_1, \ldots, T_n]$ als *Polynome in n Variablen* über R.

Beispiel 8.1.4 (Matrizen über einem Ring)**.** Die Addition und die Multiplikation von Matrizen bleiben sinnvoll für Matrizen mit Einträgen in einem beliebigen Ring R. Für jedes $n \in \mathbb{N}$ erhalten wir insbesondere den Matrizenring $(M_n(R), +, \cdot)$ über R. Der Ring $M_0(R)$ ist der Nullring, und der Ring $M_1(R)$ kann mit R selbst identifiziert werden.

Proposition 8.1.5 (Rechnen in Ringen)**.** *Sei $(R, +, \cdot)$ ein Ring.*

(i) *Für alle $x \in R$ gilt $0 \cdot x = x \cdot 0 = 0$.*

(ii) *Für alle $x \in R$ gilt $(-1) \cdot x = x \cdot (-1) = -x$. Insbesondere ist $(-1) \cdot (-1) = 1$.*

Beweis. Siehe Proposition 2.3.8. □

Bemerkung 8.1.6. Im Gegensatz zu Körpern (siehe Proposition 2.3.8(iii)) sind Ringe im Allgemeinen nicht nullteilerfrei: Es kann sein, dass $xy = 0$, ohne dass x oder y null sind. Zum Beispiel:

- Im Matrizenring $M_2(K)$ ist die Matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ nicht null, dennoch ist $A^2 = 0$.
- Im Ring $\mathbb{Z}/6\mathbb{Z}$ der Restklassen modulo 6 gilt $[2] \cdot [3] = [6] = 0$.

Definition 8.1.7 (Unterring)**.** Sei R ein Ring. Eine Teilmenge $S \subset R$ heißt *Unterring* von R, wenn $1 \in S$ und beide Verknüpfungen $+$ und $\cdot$ sich zu Verknüpfungen $S \times S \to S$ einschränken, so dass das Tripel $(S, +, \cdot)$ wieder ein Ring ist.

Proposition 8.1.8 (Kriterium für Unterringe)**.** *Sei R ein Ring. Eine Teilmenge $S \subset R$ ist genau dann ein Unterring, wenn folgende drei Bedingungen erfüllt sind:*

(i) *$1 \in S$ und $-1 \in S$.*

(ii) *Für alle $r, s \in S$ gilt $r + s \in S$.*

(iii) *Für alle $r, s \in S$ gilt $r \cdot s \in S$.*

Außerdem gilt in diesem Fall:

(iv) $0 \in S$.

(v) *Für alle $r \in S$ gilt $-r \in S$.*

Beweis. Sei S ein Unterring. Nach Definition gelten $1 \in S$, (ii) und (iii). Da $(S, +, \cdot)$ ein Ring ist, gibt es ein neutrales Element $0_S \in S$ bzgl. $+$. Dann gilt insbesondere $0_S + 1 = 1$ und somit $0_S = 0$, d.h., es gilt (iv). Ist $r \in S$, so gibt es ein Element $r^* \in S$ mit $r^* + r = 0$. Nach Eindeutigkeit des inversen Elements in der Gruppe $(R, +)$ (Proposition 2.1.3) ist $r^* = -r$. Also gilt (v), und insbeonsere $-1 \in S$.

Seien umgekehrt die Bedingungen (i)–(iii) erfüllt. Da die Verknüpfung $+$ auf R assoziativ und kommutativ ist, ist die eingeschränkte Verknüpfung $+$ auf S auch assoziativ und kommutativ. Ebenso ist die Verknüpfung $\cdot$ auf S assoziativ mit dem neutralen Element 1, und es gilt das Distributivgesetz in S. Nach (i,ii) gilt $0 = 1 + (-1) \in S$, und nach (i,iii) und Proposition 8.1.5(ii) gilt $-r \in S$ für alle $r \in S$. Damit ist $(S, +)$ eine abelsche Gruppe. Dies zeigt, dass das Tripel $(S, +, \cdot)$ ein Ring ist. □

Beispiel 8.1.9.

(i) Die folgenden Teilmengen von $\mathbb{C}$ sind Unterringe:

$$\begin{array}{ccccc} \mathbb{Z}[i] & \subset & \mathbb{Q}(i) & \subset & \mathbb{C} \\ \cup & & \cup & & \cup \\ \mathbb{Z} & \subset & \mathbb{Q} & \subset & \mathbb{R} \end{array}$$

(siehe Beispiele 2.4.4 und 8.1.2(vii)).

(ii) Ist $z \in \mathbb{C}$ eine komplexe Zahl mit $z^2 \in \mathbb{Z}$ (d.h., $z = \pm\sqrt{d}$ oder $z = \pm i\sqrt{d}$ mit einem $d \in \mathbb{N}$), so ist

$$\mathbb{Z}[z] = \{a + bz \mid a, b \in \mathbb{Z}\}$$

ein Unterring von $\mathbb{C}$. Dies kann man leicht mit dem Kriterium 8.1.8 nachrechnen. Man schreibt üblicherweise $\mathbb{Z}[\sqrt{-d}]$ anstelle von $\mathbb{Z}[i\sqrt{d}]$.

(iii) Sei $\omega = \exp(2\pi i/3) = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \in \mathbb{C}$. Eine *Eisenstein-Zahl* ist eine komplexe Zahl der Gestalt $a + b\omega$ mit $a, b \in \mathbb{Z}$; man bezeichnet mit $\mathbb{Z}[\omega] \subset \mathbb{C}$ die Menge aller Eisenstein-Zahlen. Aus $\omega^2 = -1 - \omega$ folgt leicht mit dem Kriterum 8.1.8, dass $\mathbb{Z}[\omega]$ ein Unterring von $\mathbb{C}$ ist.

Bemerkung 8.1.10. Ein Teilkörper eines Körpers K ist insbesondere ein Unterring von K, aber nicht jeder Unterring eines Körpers ist wieder ein Körper (z.B.: $\mathbb{Z} \subset \mathbb{Q}$).

Bemerkung 8.1.11. Ist $(S_i)_{i \in I}$ eine Familie von Unterringen von R, so ist der Durchschnitt $\bigcap_{i \in I} S_i$ wieder ein Unterring (dies folgt unmittelbar aus Proposition 8.1.8). Für eine beliebige Teilmenge $E \subset R$ gibt es deswegen ein *kleinster* Unterring von R, der E enthält, nämlich der Durchschnitt aller solchen Unterringe. Er heißt der *von E erzeugter Unterring* von R.

Definition 8.1.12 (Einheit)**.** Sei R ein Ring. Ein Element $x \in R$ heißt *Einheit* von R, wenn es ein inverses Element bzgl. $\cdot$ besitzt, d.h., wenn ein $y \in R$ existiert mit $xy = 1 = yx$. Die Menge aller Einheiten von R wird mit $R^\times$ oder R^* bezeichnet.

Bemerkung 8.1.13 (Einheitengruppe)**.** Die Multiplikation auf R schränkt sich zu einer Verknüpfung $\cdot\colon R^\times \times R^\times \to R^\times$, denn: Sind $x, y \in R^\times$, so gilt

$$(xy)(y^{-1}x^{-1}) = 1 = (y^{-1}x^{-1})(xy),$$

und somit ist $xy \in R^\times$. Das Paar $(R^\times, \cdot)$ ist dann eine Gruppe: Ihr neutrales Element ist 1 und das Inverse von x ist x^{-1}. Sie heißt die *Einheitengruppe* oder die *multiplikative Gruppe* von R, im Gegensatz zu der *additiven Gruppe* $(R, +)$.

Beispiel 8.1.14.

(i) Es gilt $\mathbb{Z}^\times = \{\pm 1\}$.

(ii) Ist K ein Körper und ist $n \in \mathbb{N}$, so gilt $M_n(K)^\times = \mathrm{GL}_n(K)$.

(iii) Für einen Körper K gilt $K[T]^\times = K^\times$, d.h., die Einheiten in $K[T]$ sind genau die Polynome vom Grad 0. Dies folgt aus Proposition 6.3.7(ii).

(iv) Für den Nullring $\{0\}$ gilt $\{0\}^\times = \{0\}$. Ist umgekehrt R ein Ring mit $0 \in R^\times$, so gilt $R = \{0\}$, denn $1 = 0 \cdot 0^{-1} = 0$ nach Proposition 8.1.5(i).

(v) Ein kommutativer Ring R ist genau dann ein Körper, wenn $R^\times = R \setminus \{0\}$. Ein (nicht unbedingt kommutativer) Ring R mit $R^\times = R \setminus \{0\}$ heißt *Schiefkörper*.

Die *Quaternionen* sind eine Verallgemeinerung der komplexen Zahlen, die einen nichtkommutativen Schiefkörper $\mathbb{H}$ bilden: Elemente von $\mathbb{H}$ sind Ausdrücke $a + bi + cj + dk$ mit $a, b, c, d \in \mathbb{R}$, wobei die Symbole i, j, k erfüllen

$$i^2 = j^2 = k^2 = -1 \quad \text{und} \quad ij = -ji = k.$$

Die Multiplikation von beliebigen Quaternionen ist dann durch die Ringaxiome bestimmt (z.B.: $ik = i^2 j = -j$). Als $\mathbb{R}$-Vektorraum ist $\mathbb{H}$ vierdimensional mit Basis $(1, i, j, k)$. Der *Satz von Frobenius* sagt, dass $\mathbb{R}$, $\mathbb{C}$ und $\mathbb{H}$ die *einzigen* endlichdimensionalen Schiefkörpererweiterungen von $\mathbb{R}$ sind, bis auf Isomorphie.

Beispiel 8.1.15 (invertierbare Matrizen)**.** Man schreibt $\mathrm{GL}_n(R)$ für die Einheitengruppe des Matrizenringes $M_n(R)$, deren Elemente als *invertierbare Matrizen* bezeichnet werden. Falls R kommutativ ist kann man die Determinante $\det(A)$ und die adjunkte Matrix $\mathrm{adj}(A)$ einer Matrix $A \in M_n(R)$ definieren (mit der Leibniz-Formel (5.3.24) und Definition 5.3.37), so dass es gilt

$$\det(A \cdot B) = \det(A) \cdot \det(B) \quad \text{und} \quad A \cdot \mathrm{adj}(A) = \mathrm{adj}(A) \cdot A = \det(A) \cdot I_n$$

(in den Beweisen von Proposition 5.3.26(ii) und Korollar 5.3.39 wurden die Kommutativität der Multiplikation verwendet, aber die Existenz von multiplikativen Inversen nicht). Diese zwei Gleichungen haben wir tatsächlich schon für den Polynomring $K[T]$ im Abschnitt 6.3.2 benutzt. Es folgt daraus, genau wie in der Proposition 5.3.40, dass eine Matrix $A \in M_n(R)$ genau dann invertierbar ist, wenn ihre Determinante $\det(A) \in R$ eine Einheit ist:

$$\mathrm{GL}_n(R) = M_n(R)^\times = \{A \in M_n(R) \mid \det(A) \in R^\times\}.$$

Zum Beispiel besteht $\mathrm{GL}_n(\mathbb{Z})$ aus ganzen Matrizen der Determinante ± 1.

Definition 8.1.16 (Nullteiler, Integritätsring)**.**

- Sei R ein Ring. Ein Element $x \in R$ heißt *Nullteiler*, wenn ein Element $y \in R \setminus \{0\}$ existiert, so dass $xy = 0$ oder $yx = 0$.
- Ein *Integritätsring* ist ein kommutativer Ring, in dem 0 der einzige Nullteiler ist.

Bemerkung 8.1.17.

(i) Nach Proposition 8.1.5(i) ist $0 \in R$ ein Nullteiler, sofern $R \neq \{0\}$. Im Nullring ist aber 0 kein Nullteiler, damit ist der Nullring kein Integritätsring.

(ii) Jeder Unterring eines Integritätsringes ist wieder ein Integritätsring.

(iii) Falls $x|y$ in einem Integritätsring R und $x \neq 0$, dann gibt es *genau ein* Element $t \in R$ mit $tx = y$. Denn aus $tx = t'x$ folgt $(t - t')x = 0$, und damit $t - t' = 0$. Man darf also schreiben: $t = y/x$.

Beispiel 8.1.18.

(i) Jeder Körper ist ein Integritätsring, nach Proposition 2.3.8(iii).

(ii) Der Ring $\mathbb{Z}$ ist ein Integritätsring, denn er ist ein Unterring des Körpers $\mathbb{Q}$.

(iii) Der Polynomring $K[T]$ über einem Körper K ist ein Integritätsring. Dies folgt aus Proposition 6.3.7(i,ii): Ist $fg = 0$ in $K[T]$, so ist $-\infty = \deg(fg) = \deg(f) + \deg(g)$ und somit muss f oder g null sein.

(iv) Allgemeiner kann man zeigen, mit demselben Beweis wie in (iii), dass der Polynomring $R[T]$ ein Integritätsring ist, sofern R ein Integritätsring ist. Zum Beispiel sind $\mathbb{Z}[T]$ und $K[T_1, \ldots, T_n]$ Integritätsringe.

(v) Wenn $n \in \mathbb{N} \setminus \{0\}$ keine Primzahl ist, ist $\mathbb{Z}/n\mathbb{Z}$ kein Integritätsring. Wenn $n = 1$ ist sogar $\mathbb{Z}/n\mathbb{Z} = \{0\}$. Sonst kann man schreiben $n = rs$ mit $r, s \in \{2, \ldots, n-2\}$, so dass im Ring $\mathbb{Z}/n\mathbb{Z}$ gilt $[r] \neq 0$, $[s] \neq 0$ und $[r] \cdot [s] = [n] = 0$.

Definition 8.1.19 (Ringhomomorphismus)**.** Seien R und S Ringe. Eine Abbildung $f\colon R \to S$ heißt *Ringhomomorphismus*, wenn folgende Eigenschaften erfüllt sind:

(i) Für alle $r, r' \in R$ gilt
$$f(r + r') = f(r) + f(r').$$

(ii) Für alle $r, r' \in R$ gilt
$$f(r \cdot r') = f(r) \cdot f(r').$$

(iii) Es gilt $f(1) = 1$.

Ein Ringhomomorphismus zwischen Körpern heißt auch *Körperhomomorphismus*.

Bemerkung 8.1.20. Aus (i) folgt $f(0) = 0$ (denn: $f(0) + f(0) = f(0 + 0) = f(0)$ und man kann $f(0)$ von beiden Seiten subtrahieren). Die Bedingung (iii) ist aber nicht automatisch: Die konstante Abbildung $R \to S$, $r \mapsto 0$, erfüllt (i) und (ii), aber wenn $S \neq \{0\}$ ist sie kein Ringhomorphismus.

Beispiel 8.1.21.

(i) Ist $S \subset R$ ein Unterring, so ist die Inklusionsabbildung $S \hookrightarrow R$ ein Ringhomomorphismus.

(ii) Ist R ein beliebiger Ring, so gibt es *genau einen* Ringhomomorphismus $\mathbb{Z} \to R$. Denn 0 muss auf 0 und 1 auf 1 gehen, daher muss $n \in \mathbb{N}$ auf die n-fache Summe $1 + \cdots + 1$ in R abgebildet werden, und $-n$ auf das additive Inverse davon. Durch diese Abbildung können wir jede ganze Zahl als Element von R auffassen.

(iii) Sei R ein *kommutativer* Ring und sei $r \in R$. Dann ist der *Einsetzungshomomorphismus*
$$\varepsilon_r\colon R[T] \to R, \quad p \mapsto p(r),$$
ein Ringhomomorphismus. Sind zum Beispiel $p = \sum_{i=0}^{\infty} a_i T^i$ und $q = \sum_{j=0}^{\infty} b_j T^j$, so gilt
$$(p \cdot q)(r) = \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i b_j \right) r^k \quad \text{und}$$
$$p(r) \cdot q(r) = \left(\sum_{i=0}^{\infty} a_i r^i \right) \cdot \left(\sum_{j=0}^{\infty} b_j r^j \right).$$
Die sind gleich wegen der Kommutativität der Multiplikation: $r^i b_j = b_j r^i$. Ist R nicht kommutativ, so ist ε_r genau dann ein Ringhomomorphismus, wenn $rs = sr$ für alle $s \in R$ (man sagt dann, dass r ein *zentrales* Element von R ist).

(iv) Die Quotientenabbildung $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, $x \mapsto [x]$, ist ein Ringhomomorphismus nach Definition der Verknüpfungen $+$ und $\cdot$ auf $\mathbb{Z}/n\mathbb{Z}$ (siehe Abschnitt 2.4.4).

(v) Es gibt keinen Ringhomomorphismus $\mathbb{Q} \to \mathbb{Z}$, denn: $\frac{1}{2}$ müsste auf ein Element $x \in \mathbb{Z}$ mit $x + x = 1$ gehen, aber es gibt kein solches Element in $\mathbb{Z}$.

(vi) Ist $n \in \mathbb{N}\setminus\{0\}$ und gibt es einen Ringhomomorphismus $\mathbb{Z}/n\mathbb{Z} \to R$, so muss die n-fache Summe $1+\cdots+1$ in R gleich 0 sein. Insbesondere gibt es keinen Ringhomomorphismus von $\mathbb{Z}/n\mathbb{Z}$ nach $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$.

Bemerkung 8.1.22 (Bild und Kern von Ringhomomorphismen)**.** Sei $f\colon R \to S$ ein Ringhomomorphismus. Mit dem Kriterum 8.1.8 kann man leicht nachprüfen, dass das Bild $\operatorname{im} f = f(R)$ ein Unterring von S ist. Auf der anderen Seite ist der Kern $\ker f = f^{-1}(\{0\})$ nur ein Unterring von R, wenn $S = \{0\}$. Denn aus $1 \in \ker f$ folgt $1 = f(1) = 0$ in S.

Bemerkung 8.1.23. Sei $f\colon R \to S$ ein Ringhomomorphismus. Nach Definition ist f insbesondere ein Gruppenhomomorphismus zwischen den additiven Gruppen $(R, +)$ und $(S, +)$. Ist $x \in R$ eine Einheit, so ist $f(x) \in S$ wieder eine Einheit mit $f(x)^{-1} = f(x^{-1})$, denn:

$$f(x)f(x^{-1}) = f(xx^{-1}) = f(1) = 1,$$
$$f(x^{-1})f(x) = f(x^{-1}x) = f(1) = 1.$$

Deswegen schränkt sich f zu einem Gruppenhomomorphismus zwischen den multiplikativen Gruppen $(R^\times, \cdot)$ und $(S^\times, \cdot)$.

Definition 8.1.24 (Ringisomorphismus)**.** Ein Ringhomomorphismus $f\colon R \to S$ heißt *Ringisomorphismus* oder einfach *Isomorphismus*, wenn ein Ringhomomorphismus $g\colon S \to R$ mit $g \circ f = \mathrm{id}_R$ und $f \circ g = \mathrm{id}_S$ existiert.

Proposition 8.1.25. *Ein Ringhomomorphismus $f\colon R \to S$ ist genau dann ein Ringisomorphismus, wenn er bijektiv ist.*

Beweis. Ist f bijektiv, so existiert die Umkehrabbildung $f^{-1}\colon S \to R$. Es genügt zu zeigen, dass f^{-1} ein Ringhomomorphismus ist. Es gilt $f^{-1}(1) = 1$, denn $1 = f(1)$. Für $* \in \{+, \cdot\}$ gilt $f^{-1}(s * s') = f^{-1}(s) * f^{-1}(s')$, denn f schickt beide Seiten auf $s * s'$ (und f ist injektiv). □

Eine besondere Eigenschaft von Körperhomomorphismen ist, dass sie automatisch injektiv sind:

Proposition 8.1.26. *Sei K ein Schiefkörper und sei R ein Ring mit $R \neq \{0\}$. Dann ist jeder Ringhomomorphismus $f\colon K \to R$ injektiv. Insbesondere ist jeder Körperhomomorphismus injektiv.*

Beweis. Nach Bemerkung 8.1.23 gilt

$$f(K \setminus \{0\}) = f(K^\times) \subset R^\times \subset R \setminus \{0\},$$

deswegen ist der Kern von f null, und somit ist f injektiv. □

8.1.2 Moduln

Die Definition eines Moduls über einem Ring ist identisch mit der Definition eines Vektorraums über einem Körper (vgl. Definition 3.2.1):

Definition 8.1.27 (Modul)**.** Sei R ein Ring. Ein *Modul* über R, oder *R-Modul*, ist ein Tripel $(M, +, \cdot)$ bestehend aus einer Menge M und Verknüpfungen

$$+\colon M \times M \to M, \qquad \cdot\colon R \times M \to M,$$
$$(x, y) \mapsto x + y, \qquad (r, x) \mapsto r \cdot x,$$

die als *Addition* und *Skalarmultiplikation* bezeichnet werden, mit den folgenden Eigenschaften:

(i) $(M, +)$ ist eine abelsche Gruppe.

(ii) Für alle $r, s \in R$ und $x \in M$ gilt

$$r \cdot (s \cdot x) = (r \cdot s) \cdot x.$$

(iii) Für alle $x \in M$ gilt

$$1 \cdot x = x.$$

(iv) Für alle $r, s \in K$ und $x, y \in M$ gilt

$$r \cdot (x + y) = r \cdot x + r \cdot y,$$
$$(r + s) \cdot x = r \cdot x + s \cdot x.$$

Beispiel 8.1.28.

(i) Ein Modul über einem Körper K ist nach Definition genau ein K-Vektorraum.

(ii) Sei R ein Ring und $n \in \mathbb{N}$. Dann ist R^n ein R-Modul mit der komponentenweisen Addition bzw. Skalarmultiplikation (siehe Definition 3.1.2). Insbesondere ist R selbst ein R-Modul.

(iii) Sei $f\colon R \to S$ ein Ringhomomorphismus. Dann ist S ein R-Modul mit seiner Addition und mit der Skalarmultiplikation

$$R \times S \to S, \quad (r, s) \mapsto f(r) \cdot s.$$

(iv) Sei R ein Ring und $n \in \mathbb{N}$. Dann ist R^n ein $M_n(R)$-Modul mit der Skalarmultiplikation

$$M_n(R) \times R^n \to R^n, \quad (A, x) \mapsto A \cdot x,$$

wobei wir x als Spaltenvektor auffassen. Die Modulaxiome (ii)–(iv) folgen aus Proposition 4.2.16(i)–(iii).

Bemerkung 8.1.29 (Links- und Rechtsmoduln)**.** Wenn R nicht kommutativ ist, Moduln über R wie in der obigen Definition heißen genauer *Linksmoduln*. Ein *R-Rechtsmodul* M ist fast das Gleiche, aber die Skalarmultiplikation wird als $\cdot\colon M \times R \to M$ geschrieben und die Bedingung (ii) lautet: $(x \cdot s) \cdot r = x \cdot (s \cdot r)$. Das heißt: In einem Linksmodul ist Skalarmultiplikation mit s und dann mit r dasselbe wie Skalarmultiplikation mit $r \cdot s$, aber in einem Rechtsmodul ist es wie Skalarmultiplikation mit $s \cdot r$. Wenn der Ring R kommutativ ist, gibt es also keinen Unterschied zwischen Links- und Rechtsmoduln über R. Im Allgemeinen ist ein R-Rechtsmodul dasselbe wie ein R^{op}-Linksmodul, wobei R^{op} der gleiche Ring wie R ist, aber mit umgekehrter Multiplikation $(r, s) \mapsto s \cdot r$.

Definition 8.1.30 (Modulhomomorphismus)**.** Sei R ein Ring und seien M, N Moduln über R. Eine *(R-)lineare Abbildung* oder *(R-)Modulhomomorphismus* von M nach N ist eine Abbildung $f\colon M \to N$ mit folgenden Eigenschaften:

(i) f ist ein Gruppenhomorphismus, d.h., für alle $x, y \in M$ gilt

$$f(x + y) = f(x) + f(y).$$

(ii) Für alle $x \in M$ und $r \in R$ gilt

$$f(r \cdot x) = r \cdot f(x).$$

Notation 8.1.31. Wie bei Vektorräumen schreiben wir $\mathrm{Hom}_R(M, N) \subset \mathrm{Abb}(M, N)$ für die Teilmenge von R-linearen Abbildungen, und $\mathrm{End}_R(M) = \mathrm{Hom}_R(M, M)$.

Viele Begriffe und Resultate über Vektorräume und lineare Abbildungen, die in Kapiteln 3 und 4 behandelt wurden, lassen sich unmittelbar auf Moduln und Modulhomomorphismen übertragen:

- Die Rechenregeln (i)–(iii) der Proposition 3.2.6 gelten auch für Moduln; die Aussage (iv) aber nicht.
- Ein *Untermodul* eines Moduls wird genau wie ein Untervektorraum definiert (Definition 3.2.7). Das Kriterium 3.2.8 für Untervektorräume kann man auch mit Moduln anwenden: Eine Teilmenge $N \subset M$ ist genau dann ein Untermodul, wenn sie nicht leer ist und sie unter Addition und Skalarmultiplikation abgeschlossen ist.
- Ist $(N_i)_{i\in I}$ eine Familie von Untermoduln von M, so ist ihre Durchschnitt $\bigcap_{i\in I} N_i$ wieder ein Untermodul von M (Proposirion 3.2.16).
- Ist M ein R-Modul und ist $E \subset M$ eine beliebige Teilmenge, so gibt es einen kleinsten Untermodul $\mathrm{Span}_R(E)$ von M, der E enthält. Man bezeichnet diesen Untermodul als den *von E erzeugten Untermodul* von M, und es gilt

  $$\mathrm{Span}_R(E) = \left\{ \sum_{i=1}^{n} r_i \cdot x_i \;\middle|\; n \in \mathbb{N},\ r_i \in R,\ x_i \in E \right\}$$

  (Proposition 3.2.18). Die *Summe* $\sum_{i\in I} N_i$ einer Familie von Untermoduln von M ist der Untermodul $\mathrm{Span}_R\left(\bigcup_{i\in I} N_i\right)$.
- Eine Teilmenge $E \subset M$ heißt *Erzeugendensystem* von M, wenn $\mathrm{Span}_R(E) = M$, und M heißt *endlich erzeugt*, wenn ein endliches Erzeugendensystem existiert.
- Ein R-Modulhomomorphismus $f\colon M \to N$ ist genau dann ein Isomorphismus (d.h., besitzt eine R-lineare Umkehrabbildung), wenn er bijektiv ist (Proposition 4.1.17).
- Ist $(M_i)_{i\in I}$ eine Familie von R-Moduln, so hat die Produktmenge $\prod_{i\in I} M_i$ eine Struktur von R-Modul mit der punktweisen Addition bzw. Skalarmultiplikation. Dieser R-Modul heißt das *Produkt* der Familie $(M_i)_{i\in I}$. Die *direkte Summe* der Familie ist der Untermodul

  $$\bigoplus_{i\in I} M_i \subset \prod_{i\in I} M_i$$

  bestehend aus den Familien $(x_i)_{i\in I}$, die außerhalb einer endlichen Teilmenge von I null sind. Die universellen Eigenschaften aus Proposition 6.1.3 gelten. Wenn I endlich ist gibt es keinen Unterschied zwischen Produkt und direkte Summe. Insbesondere haben wir das Produkt/die direkte Summe $M \times N = M \oplus N$ von zwei R-Moduln.
- Als Sonderfall der letzten Konstruktion definieren wir die R-Moduln

  $$R^I = \mathrm{Abb}(I, R) = \prod_{i\in I} R \quad \text{und} \quad R^{(I)} = \bigoplus_{i\in I} R.$$

- Ist $F = (x_i)_{i\in I}$ eine Familie von Elementen in einem R-Modul M, so gibt es genau einen Modulhomomorphismus

  $$\varphi_F\colon R^{(I)} \to M,$$

  der e_i auf x_i abbildet. Die Familie F heißt *linear unabhängig* bzw. *erzeugend*, wenn φ_F injektiv bzw. surjektiv ist, und sie ist eine *Basis* von M, wenn φ_F bijektiv ist. Insbesondere ist $(e_1, \ldots, e_n)$ eine Basis von R^n. Die universelle Eigenschaft von Basen gilt (Satz 4.1.22).
- Der *Kern* $\ker f \subset M$ und das *Bild* $\mathrm{im}\, f \subset N$ eines Modulhomomorphismus $f\colon M \to N$ sind Untermoduln. Außerdem ist f genau dann injektiv, wenn $\ker f = \{0\}$.

- Sei M ein R-Modul und $N \subset M$ ein Untermodul. Dann können wir der *Quotientenmodul* von M nach N wie folgt definieren:

$$M/N := M/\sim_N, \quad \text{wobei} \quad x \sim_N y \iff x - y \in N$$

(Proposition 3.2.23). Die Äquivalenzklasse von $x \in M$ bezüglich $\sim_N$ ist $x + N$. Die Quotientenabbildung $M \twoheadrightarrow M/N$, $x \mapsto x + N$, ist dann R-linear und hat dieselbe universelle Eigenschaft wie bei Vektorräumen (Proposition 4.1.33).

- Es gilt den *Homomorphiesatz* (Satz 4.1.35): Ist $f\colon M \to N$ eine R-lineare Abbildung, so gibt es einen induzierten Isomorphismus von R-Moduln

$$\bar{f}\colon M/\ker f \xrightarrow{\sim} \operatorname{im} f, \quad x + \ker f \mapsto f(x).$$

- Ist R kommutativ und sind M und N Moduln über R, so ist die Menge $\operatorname{Hom}_R(M, N)$ wieder ein R-Modul mit der punktweisen Addition bzw. Skalarmultiplikation. Insbesondere definieren wir den *Dualmodul* $M^* = \operatorname{Hom}_R(M, R)$. Dabei ist die Kommutativität von R notwendig, damit die punktweise Skalarmultiplikation auf $\operatorname{Abb}(M, N)$ R-Modulhomomorphismen erhält: Für $f \in \operatorname{Hom}_R(M, N)$ und $r \in R$ muss $r \cdot f$ wieder R-linear sein, insbesondere müssen

$$\begin{aligned} &(r \cdot f)(s \cdot x) = r \cdot f(s \cdot x) = r \cdot s \cdot f(x) \\ \text{und} \quad &s \cdot (r \cdot f)(x) = s \cdot r \cdot f(x) \end{aligned}$$

für alle $s \in R$ und $x \in M$ gleich sein.

- Sei R ein kommutativer Ring. Dann definiert jede Matrix $A \in M_{m\times n}(R)$ eine R-lineare Abbildung

$$L_A\colon R^n \to R^m, \quad x \mapsto A \cdot x,$$

wobei wir x als Spaltenvektor auffassen. Die Kommutativität der Multiplikation ist hier notwendig, denn man braucht $r \cdot (A \cdot x) = A \cdot (r \cdot x)$ für alle $r \in R$ (im Allgemeinen ist L_A trotzdem R-linear, wenn wir R^n und R^m als R-*Rechts*moduln betrachten, siehe Bemerkung 8.1.29). Zudem ist die Abbildung

$$M_{m\times n}(R) \to \operatorname{Hom}_R(R^n, R^m), \quad A \mapsto L_A,$$

R-linear und *bijektiv*, d.h., man kann R-lineare Abbildungen $R^n \to R^m$ mit Matrizen identifizieren.

Ein wichtiger Unterschied zwischen Vektorräumen und Moduln über beliebigen Ringen ist folgender: Es gibt Moduln, die keine Basis besitzen. Anders gesagt gibt es R-Moduln, die nicht zu einem R-Modul der Gestalt $R^{(I)}$ isomorph sind. Deswegen ist die Dimension eines beliebigen Moduls nicht definiert. Selbst wenn Basen existieren, können sich Moduln auf ungewöhnliche Weise verhalten. Zum Beispiel ist es möglich, dass ein Untermodul U von R^n mit $U \neq R^n$ trotzdem zu R^n isomorph ist (z.B.: $R = \mathbb{Z}$, $n = 1$ und $U = 2\mathbb{Z}$).

Definition 8.1.32 (freier Modul)**.** Sei R ein Ring. Ein R-Modul M heißt *frei*, wenn er eine Basis besitzt, d.h., wenn eine Menge I existiert, so dass $M \cong R^{(I)}$. Ist $n \in \mathbb{N}$, so heißt M *frei vom Rang* n, wenn $M \cong R^n$.

Beispiel 8.1.33 (nicht-freie Moduln)**.**

(i) Eine endliche abelsche Gruppe $A \neq \{0\}$ (z.B.: $\mathbb{Z}/n\mathbb{Z}$ mit $n \geq 2$) ist ein $\mathbb{Z}$-Modul (nach Proposition 8.1.36 unten), der nicht frei ist. Denn nicht-triviale freie $\mathbb{Z}$-Moduln müssen unendlich sein.

(ii) Sei K ein Körper und sei $\varepsilon_0 \colon K[T] \to K$ der Einsetzungshomomorphismus $\varepsilon_0(p) = p(0)$ (Beispiel 8.1.21(iii)). Nach Beispiel 8.1.28(iii) definiert ε_0 eine $K[T]$-Modulstruktur auf K. Dieser $K[T]$-Modul ist dann nicht frei, und zwar: Kein einzelnes Element $a \in K$ bildet eine linear unabhängige Familie, denn es gilt $T \cdot a = \varepsilon_0(T) \cdot a = 0 \cdot a = 0$.

(iii) Sei K ein Körper und $n \geq 2$. Dann ist der $M_n(K)$-Modul K^n aus Beispiel 8.1.28(iv) nicht frei, denn zu jedem $x \in K^n$ gibt es eine Matrix $A \in M_n(K) \setminus \{0\}$ mit $A \cdot x = 0$.

(iv) Der $\mathbb{Z}$-Modul $\mathbb{Q}$ ist nicht frei, denn: In $\mathbb{Q} \setminus \{0\}$ gibt es eine Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n = 2 \cdot x_{n+1}$, aber in $\mathbb{Z}^{(I)} \setminus \{0\}$ gibt es keine solche Folge.

(v) Es zeigt sich, dass der $\mathbb{Z}$-Modul $\mathbb{Z}^{\mathbb{N}}$ nicht frei ist, aber das ist nicht offensichtlich.

Bemerkung 8.1.34. Ist $R \neq \{0\}$ und ist I unendlich, so ist $R^{(I)}$ nicht endlich erzeugt. Denn zu jeder endlichen Teilmenge $E \subset R^{(I)}$ gibt es eine endliche Teilmenge $J \subset I$ mit $\mathrm{Span}_R(E) \subset \mathrm{Span}_R\{e_j \mid j \in J\}$. Ist $i \in I \setminus J$, so folgt aus $1 \neq 0$, dass $e_i \notin \mathrm{Span}_R\{e_j \mid j \in J\}$.

Bemerkung 8.1.35. Im Gegensatz zu Vektorräumen (Satz 3.3.29) gibt es Ringe $R \neq \{0\}$, so dass $R^{(I)}$ und $R^{(J)}$ isomorph sind, selbst wenn I und J nicht gleichmächtig sind (tatsächlich können R und R^2 isomorph sein). Das kann aber nur mit gewissen nicht-kommutativen Ringen geschehen (und auf triviale Weise, wenn $R = \{0\}$). Wenn R kommutativ und nicht null ist, kann man zeigen, dass alle Basen eines freien R-Moduls dieselbe Länge haben.

Die nächste Proposition zeigt, dass es im Wesentlichen keinen Unterschied zwischen abelschen Gruppen und $\mathbb{Z}$-Moduln gibt:

Proposition 8.1.36 (abelsche Gruppen sind $\mathbb{Z}$-Moduln)**.**

(i) *Sei $(A,+)$ eine abelsche Gruppe. Dann gibt es* genau eine *Abbildung $\cdot \colon \mathbb{Z} \times A \to A$, so dass das Tripel $(A,+,\cdot)$ ein $\mathbb{Z}$-Modul ist.*

(ii) *Seien $(A,+)$ und $(B,+)$ abelsche Gruppen. Eine Abbildung $f \colon A \to B$ ist genau dann ein Gruppenhomorphismus von $(A,+)$ nach $(B,+)$, wenn sie ein Modulhomomorphismus zwischen den entsprechenden $\mathbb{Z}$-Moduln ist.*

Beweis. Zu (i). Für $n \in \mathbb{Z}$ und $a \in A$ definiert man $n \cdot a$ wie in Notation 2.1.11, d.h.:

$$n \cdot a = \begin{cases} 0, & \text{falls } n = 0, \\ \underbrace{a + \cdots + a}_{n \text{ mal}}, & \text{falls } n > 0, \\ -((-n) \cdot a), & \text{falls } n < 0. \end{cases}$$

Das ist die einzige Möglichkeit, denn es muss $1 \cdot a = a$ gelten nach Axiom (iii), $0 \cdot a = 0$ und $(-1) \cdot a = -a$ nach Proposition 3.2.6(iii), und alle anderen Fälle sind durch Axiom (iv) bestimmt. Um zu zeigen, dass dies tatsächlich eine $\mathbb{Z}$-Modulstruktur auf A definiert, muss man noch nachprüfen:

$$\begin{aligned} n \cdot (m \cdot a) &= (n \cdot m) \cdot a, \\ n \cdot (a + b) &= (n \cdot a) + (n \cdot b), \\ (n + m) \cdot a &= (n \cdot a) + (m \cdot a). \end{aligned}$$

Die erste Aussage ist genau Proposition 2.1.7(ii) (in additiver Notation), die dritte ist Proposition 2.1.7(i) und die zweite folgt aus Bemerkung 2.1.8.

Zu (ii). Nach Definition ist jeder Modulhomomorphismus insbesondere ein Gruppenhomomorphismus. Sei umgekehrt $f \colon A \to B$ ein Gruppenhomomorphismus. Aus der obigen Definition von $n \cdot a$ folgt unmittelbar, dass $f(n \cdot a) = n \cdot f(a)$ gilt, für alle $n \in \mathbb{Z}$ und $a \in A$. Das heißt, f ist ein $\mathbb{Z}$-Modulhomomorphismus. □

Sei R ein Ring und M ein Modul über dem Polynomring $R[T]$. Da R ein Unterring von $R[T]$ ist, ist M insbesondere ein R-Modul, indem man die Skalarmultiplikation $R[T] \times M \to M$ auf $R \times M$ einschränkt. Die Skalarmultiplikation mit T definiert dann einen R-linearen Endomorphismus

$$M \to M, \quad x \mapsto T \cdot x,$$

denn für alle $r \in R$ gilt $r \cdot (T \cdot x) = (r \cdot T) \cdot x = (T \cdot r) \cdot x = T \cdot (r \cdot x)$. Die nächste Proposition sagt, dass alle R-linearen Endomorphismen aus dieser Konstruktion entstehen:

Proposition 8.1.37 (Endomorphismen sind Moduln über dem Polynomring)**.** *Sei R ein Ring.*

(i) *Sei M ein R-Modul. Zu jedem Endomorphismus $f \in \mathrm{End}_R(M)$ gibt es* genau eine *$R[T]$-Modulstruktur auf M, die die gegebene R-Modulstruktur fortsetzt, so dass für alle $x \in M$ gilt $T \cdot x = f(x)$.*

(ii) *Seien M und N Moduln über R mit Endomorphismen $f \in \mathrm{End}_R(M)$ und $g \in \mathrm{End}_R(N)$, die $R[T]$-Modulstrukturen auf M und N definieren. Dann ist eine R-lineare Abbildung $\varphi\colon M \to N$ genau dann $R[T]$-linear, wenn $\varphi \circ f = g \circ \varphi$.*

Beweis. Zu (i). Man definiert die Skalarmultiplikation $R[T] \times M \to M$ durch

$$\left(\sum_{i=0}^{n} a_i T^i\right) \cdot x = \sum_{i=0}^{n} a_i \cdot f^i(x),$$

was offenbar die einzige Möglichkeit ist. Man kann dann leicht nachrechnen, dass dies tatsächlich eine $R[T]$-Modulstruktur auf M definiert. Seien zum Beispiel $p = \sum_{i=0}^{n} a_i T^i$ und $q = \sum_{j=0}^{m} b_j T^j$ Polynome über R. Dann gilt:

$$p \cdot (q \cdot x) = \sum_{i=0}^{n} a_i f^i \left(\sum_{j=0}^{m} b_j f^j(x)\right) = \sum_{i=0}^{n} \sum_{j=0}^{m} a_i b_j f^{i+j}(x) = (p \cdot q) \cdot x,$$

wobei in der zweiten Gleichung die R-Linearität von f^i benutzt wurde.

Zu (ii). Sei $\varphi\colon M \to N$ R-linear. Dann ist φ genau dann $R[T]$-linear, wenn für alle $x \in M$ und $p \in R[T]$ gilt $\varphi(p \cdot x) = p \cdot \varphi(x)$. Für $p = T$ ist das genau die Aussage $\varphi \circ f = g \circ \varphi$. Sei umgekehrt $\varphi \circ f = g \circ \varphi$. Dann gilt auch $\varphi \circ f^i = g^i \circ \varphi$ für alle $i \in \mathbb{N}$, das heißt, $\varphi(T^i \cdot x) = T^i \cdot \varphi(x)$. Da φ R-linear vorausgesetzt wurde, schließen wir daraus, dass $\varphi(p \cdot x) = p \cdot \varphi(x)$ für alle $p \in R[T]$. □

Notation 8.1.38. Sei M ein R-Modul mit einem Endomorphismus $f \in \mathrm{End}_R(M)$. Nach Proposition 8.1.37(i) können wir M auf eindeutige Weise als $R[T]$-Modul auffassen, so dass f gleich der Skalarmultiplikation mit T ist. Man schreibt dann $M[f]$ für diesen $R[T]$-Modul.

Bemerkung 8.1.39. Seien M und N Moduln über R mit Endomorphismen $f \in \mathrm{End}_R(M)$ und $g \in \mathrm{End}_R(N)$. Aus Proposition 8.1.37(ii) folgt, dass die Paare (M, f) und (N, g) genau dann isomorph im Sinne der Definition 6.1.13 sind, wenn die entsprechenden $R[T]$-Moduln $M[f]$ und $N[g]$ isomorph sind.

Bemerkung 8.1.40. Sei M ein R-Modul und $f \in \mathrm{End}_R(M)$ ein Endomorphismus. Ein Untermodul $N \subset M$ ist genau dann f-invariant im Sinne der Definition 6.1.8, wenn N ein Untermodul von $M[f]$ ist.

8.1.3 Algebren

Eine *Algebra* ist ein Ring, der gleichzeitig ein Modul über einem kommutativen Grundring (oft einem Körper) ist:

Definition 8.1.41 (Algebra, Algebrenhomomorphismus)**.** Sei R ein kommutativer Ring. Eine *R-Algebra* ist ein Quadrupel $(A, +, \cdot, *)$ mit folgenden Eigenschaften:

(i) $(A, +, \cdot)$ ist ein Ring.

(ii) $(A, +, *)$ ist ein R-Modul.

(iii) Für alle $r \in R$ und alle $a, b \in A$ gilt:

$$r * (a \cdot b) = (r * a) \cdot b = a \cdot (r * b).$$

Sie heißt *kommutativ*, wenn der Ring $(A, +, \cdot)$ kommutativ ist. Ein *Algebrenhomomorphismus* zwischen R-Algebren ist ein R-linearer Ringhomomorphismus, d.h., eine Abbildung, die gleichzeitig ein Ringhomomorphismus und ein R-Modulhomomorphismus ist.

Bemerkung 8.1.42. Die Skalarmultiplikation $*$ in einer R-Algebra wird üblicherweise auch mit $\cdot$ geschrieben. Die Schreibweise $*$ verwenden wir hier nur aus Gründen der Klarheit.

Bemerkung 8.1.43. Das dritte Axiom kann man wie folgt verstehen. Das Distributivgesetz in einem Ring A ist äquivalent zu der Aussage, dass die Multiplikation $\cdot\colon A \times A \to A$ eine $\mathbb{Z}$-bilineare Abbildung ist. Das Axiom (iii) in der Definition einer R-Algebra ist genau die zusätzliche Aussage, dass die Multiplikation R-bilinear ist.

Beispiel 8.1.44. Sei R ein kommutativer Ring.

(i) Der Polynomring $R[T]$ ist in natürlicher Weise eine kommutative R-Algebra (siehe Bemerkung 6.3.5).

(ii) Der Matrizenring $M_n(R)$ ist eine R-Algebra mit der Skalarmultiplikation aus Definition 4.2.4.

(iii) Ist M ein R-Modul, so ist $(\mathrm{End}_R(M), +, \circ, \cdot)$ eine R-Algebra.

(iv) Da abelsche Gruppen auf eindeutige Weise $\mathbb{Z}$-Moduln sind, sind auch Ringe auf eindeutige Weise $\mathbb{Z}$-Algebren (das Axiom (iii) gilt automatisch nach Bemerkung 8.1.43).

Der Polynomring $R[T]$ hat eine nützliche universelle Eigenschaft als R-Algebra:

Proposition 8.1.45 (universelle Eigenschaft des Polynomringes)**.** *Sei R ein kommutativer Ring. Zu jeder R-Algebra A und jedem Element $a \in A$ gibt es* genau einen *R-Algebrenhomomorphismus $\varepsilon_a\colon R[T] \to A$ mit $\varepsilon_a(T) = a$.*

Die Abbildung ε_a heißt der *Einsetzungshomomorphismus* zu a.

Beweis. Ein solches ε_a muss $\varepsilon_a(T^i) = a^i$ für alle $i \in \mathbb{N}$ erfüllen. Als R-Modul hat $R[T]$ die Basis $(T^i)_{i\in\mathbb{N}}$. Nach der universellen Eigenschaft von Basen gibt es genau eine R-lineare Abbildung $\varepsilon_a\colon R[T] \to A$ mit $\varepsilon_a(T^i) = a^i$, nämlich:

$$\varepsilon_a\left(\sum_{i=0}^{n} r_i T^i\right) = \sum_{i=0}^{n} r_i a^i \in A.$$

Das ist also die einzige Möglichkeit für ε_a, und es bleibt zu zeigen, dass dieses ε_a ein Ringhomomorphismus ist, d.h.:

$$\left(\sum_{i=0}^{n} r_i a^i\right) \cdot \left(\sum_{j=0}^{m} s_j a^j\right) = \sum_{k=0}^{n+m} \left(\sum_{i+j=k} r_i s_j\right) a^k.$$

Dies folgt aus dem Axiom (iii) für die R-Algebra A. □

Beispiel 8.1.46. Sei R ein kommutativer Ring, M ein R-Modul und $f \in \mathrm{End}_R(M)$ ein Endomorphismus. Da $\mathrm{End}_R(M)$ eine R-Algebra ist, gibt es genau einen R-Algebrenhomomorphismus

$$\varepsilon_f \colon R[T] \to \mathrm{End}_R(M) \quad \text{mit} \quad \varepsilon_f(T) = f.$$

Für ein beliebiges Polynom $p \in R[T]$ gilt dann $\varepsilon_f(p) = p(f)$, wobei $p(f)$ in Definition 6.3.37 definiert wurde. Für jedes $x \in M$ gilt deswegen $\varepsilon_f(p)(x) = p \cdot x$, wobei die rechte Seite die Skalarmultiplikation im $R[T]$-Modul $M[f]$ ist (siehe Notation 8.1.38).

Definition 8.1.47 (zentrales Element)**.** Sei A ein Ring. Ein Element $a \in A$ heißt *zentral*, wenn für alle $b \in A$ gilt $ab = ba$.

Bemerkung 8.1.48. Ein Ring A ist genau dann kommutativ, wenn alle seine Elemente zentral sind. Im Allgemeinen bilden die zentralen Elemente von A einen kommutativen Unterring von A (nach Proposition 8.1.8), das *Zentrum* von A.

Beispiel 8.1.49. Sei A ein beliebiger Ring. Nach Definition der Multiplikation von Polynomen ist T ein zentrales Element im Polynomring $A[T]$.

Proposition 8.1.50 (Algebren als Ringhomomorphismen)**.** *Sei R ein kommutativer Ring und sei $(A, +, \cdot)$ ein Ring. Dann gibt es eine kanonische Bijektion zwischen:*

(i) *Abbildungen $* \colon R \times A \to A$, so dass $(A, +, \cdot, *)$ eine R-Algebra ist.*

(ii) *Ringhomomorphismen $f \colon R \to A$, so dass jedes Element von $f(R)$ zentral in A ist.*

Beweis. Gegeben sei $*$ wie in (i). Man definiert

$$f \colon R \to A, \quad f(r) = r * 1.$$

Mit Axiomen (ii) und (iii) für R-Algebren berechnen wir

$$\begin{aligned}(r * 1) + (s * 1) &= (r + s) * 1,\\ (r * 1) \cdot (s * 1) &= r * (1 \cdot (s * 1)) = r * (s * 1) = (r \cdot s) * 1,\\ 1 * 1 &= 1,\end{aligned}$$

d.h., f ist ein Ringhomomorphismus. Nach Axiom (iii) ist zudem $r * 1$ ein zentrales Element in A:

$$(r * 1) \cdot a = 1 \cdot (r * a) = (r * a) \cdot 1 = a \cdot (r * 1).$$

Sei umgekehrt f wie in (ii). Man definiert

$$* \colon R \times A \to A, \quad r * a = f(r) \cdot a.$$

Nach Beispiel 8.1.28(iii) ist dann $(A, +, *)$ ein R-Modul. Das Axiom (iii) für eine R-Algebra lautet

$$f(r) \cdot (a \cdot b) = (f(r) \cdot a) \cdot b = a \cdot (f(r) \cdot b),$$

und es folgt aus der Assoziativität von $\cdot$ und der Zentralität von $f(r)$.

Man kann leicht nachprüfen, dass diese Konstruktionen zueinander invers sind, was die gewünschte Bijektion liefert. □

Beispiel 8.1.51. Für die Matrizenalgebra $M_n(R)$ über einem kommutativen Ring R ist der entsprechende Ringhomomorphismus folgender:

$$R \to M_n(R), \quad r \mapsto rI_n = \begin{pmatrix} r & & & 0 \\ & r & & \\ & & \ddots & \\ 0 & & & r \end{pmatrix}.$$

8.1.4 Ideale

Definition 8.1.52 (Ideal)**.** Sei R ein Ring. Ein *Ideal* in R ist ein Untermodul von R, d.h., eine nicht-leere Teilmenge $I \subset R$ mit folgenden Eigenschaften:

(i) Für alle $x, y \in I$ gilt $x + y \in I$.

(ii) Für alle $x \in I$ und $r \in R$ gilt $r \cdot x \in I$.

Notation 8.1.53. Sei $E \subset R$ eine Teilmenge. Man schreibt (E) für das von E erzeugtes Ideal $\mathrm{Span}_R(E)$ in R. Für Elemente $x_1, \dots, x_n \in R$ schreibt man auch

$$(x_1, \dots, x_n) := \mathrm{Span}_R\{x_1, \dots, x_n\} = \sum_{i=1}^{n} Rx_i = \left\{ \sum_{i=1}^{n} r_i x_i \;\middle|\; r_1, \dots, r_n \in R \right\}.$$

Beispiel 8.1.54.

(i) $\{0\}$ und R sind stets Ideale eines Ringes R. Es gilt $\{0\} = (0)$ und $R = (1)$. Deswegen heißt $\{0\}$ das *Nullideal* und R das *Einsideal* von R. In einem Körper gibt es keine anderen Ideale.

(ii) Ist $f\colon R \to S$ ein Ringhomomorphismus, so ist $\ker f$ ein Ideal von R. Denn f ist auch ein R-Modulhomomorphismus, wenn S mit der R-Modulstruktur aus Beispiel 8.1.28(iii) versehen wird.

(iii) Für jedes $n \in \mathbb{N}$ ist die Teilmenge $n\mathbb{Z} \subset \mathbb{Z}$ ein Ideal in $\mathbb{Z}$, nämlich das von n erzeugte Ideal. Wir werden später beweisen, dass $\mathbb{Z}$ keine anderen Ideale hat (Korollar 8.2.18).

(iv) Die Menge aller Polynome über einem Ring R mit Absolutglied Null ist ein Ideal im Polynomring $R[T]$, nämlich das von T erzeugte Ideal.

Bemerkung 8.1.55 (Links- und Rechtsideal, zweiseitiges Ideal)**.** Wenn R nicht kommutativ ist gibt es Links- und Rechtsmoduln über R (Bemerkung 8.1.29). Der Ring R is gleichzeitig ein Links- sowie ein Rechtsmodul über sich selbst. Dementsprechend können wir Unterlinksmoduln und Unterrechtsmoduln von R betrachten, die als *Linksideale* und *Rechtsideale* bezeichnet werden. Ideale wie in der Definition 8.1.52 sind genauer Linksideale. Die Definition eines Rechtsideals erhält man, indem man die Bedingung (ii) durch die folgende ersetzt:

(ii′) Für alle $x \in I$ und $r \in R$ gilt $x \cdot r \in I$.

Ein *zweiseitiges Ideal* in R ist eine Teilmenge von R, die gleichzeitig ein Linksideal und ein Rechtsideal ist. Der Kern eines Ringhomomorphismus $f\colon R \to S$ ist stets ein zweiseitiges Ideal in R, denn für $x \in \ker f$ und $r \in R$ beliebig gelten $f(rx) = f(r)f(x) = f(r) \cdot 0 = 0$ und $f(xr) = f(x)f(r) = 0 \cdot f(r) = 0$.

Definition 8.1.56 (zyklischer Modul)**.** Sei R ein Ring. Ein R-Modul M heißt *zyklisch*, wenn ein Element $x \in M$ existiert, so dass $M = \mathrm{Span}_R\{x\}$.

Beispiel 8.1.57.

(i) Ein Vektorraum V über einem Körper K ist genau dann zyklisch, wenn $V = \{0\}$ oder $V \cong K$.

(ii) Für jedes $n \in \mathbb{N}$ ist $\mathbb{Z}/n\mathbb{Z}$ ein zyklischer $\mathbb{Z}$-Modul, denn er ist von $[1]$ erzeugt.

(iii) Sei M ein R-Modul und $f \in \mathrm{End}_R(M)$. Der $R[T]$-Modul $M[f]$ ist genau dann zyklisch, wenn ein Element $x \in M$ existiert, so dass

$$M = \mathrm{Span}_R\{x, f(x), f^2(x), \dots\}.$$

Sei zum Beispiel $f_\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ die Drehung um den Winkel α. Dann ist der $\mathbb{R}[T]$-Modul $\mathbb{R}^2[f_\alpha]$ genau dann zyklisch, wenn α kein ganzzahliges Vielfaches von π ist. In diesem Fall ist $\mathbb{R}^2 = \mathrm{Span}_{\mathbb{R}}\{e_1, f_\alpha(e_1)\}$.

Proposition 8.1.58. *Seien R ein Ring und M ein R-Modul. Die folgenden Aussagen sind äquivalent:*

(i) *M ist zyklisch.*

(ii) *Es gibt ein Ideal $I \subset R$, so dass $M \cong R/I$.*

Beweis. Zu (i) $\Rightarrow$ (ii). Sei $M = \mathrm{Span}_R\{x\}$. Es gibt genau eine R-lineare Abbildung $\varphi_x \colon R \to M$ mit $\varphi_x(1) = x$. Da M von x erzeugt ist, ist φ_x surjektiv. Nach dem Homomorphiesatz (Satz 4.1.35) erhalten wir einen Isomorphismus $\bar{\varphi}_x \colon R/\ker\varphi_x \xrightarrow{\sim} M$, wie gewünscht.

Zu (ii) $\Rightarrow$ (i). Sei $f \colon R/I \xrightarrow{\sim} M$ ein Isomorphismus. Dann ist M von dem Element $f(1+I)$ erzeugt, denn ein beliebiges Element hat die Form $f(r+I) = f(r\cdot(1+I)) = r\cdot f(1+I)$. □

Definition 8.1.59 (Hauptideal)**.** Sei R ein Ring. Ein *Hauptideal* in R ist ein zyklischer Untermodul von R, d.h., ein Ideal der Gestalt

$$(x) = Rx = \{rx \mid r \in R\}$$

mit einem $x \in R$.

Definition 8.1.60 (Hauptidealring)**.** Ein *Hauptidealring* ist ein Integritätsring, in dem jedes Ideal ein Hauptideal ist.

Beispiel 8.1.61.

(i) Ein Körper K ist ein Hauptidealring, denn $\{0\} = (0)$ und $K = (1)$ sind die einzigen Ideale in K.

(ii) Im nächsten Abschnitt beweisen wir, dass $\mathbb{Z}$ und der Polynomring $K[T]$ über einem Körper K Hauptidealringe sind (Korollar 8.2.18).

(iii) Der Polynomring $\mathbb{Z}[T]$ ist kein Hauptidealring, denn das Ideal $(2, T)$ bestehend aus Polynomen mit geradem Absolutglied ist kein Hauptideal.

(iv) Der Polynomring $K[T_1, \ldots, T_n]$ in n Variablen über einem Körper K ist kein Hauptidealring wenn $n \geq 2$: Das Ideal (T_1, T_2) ist kein Hauptideal.

8.2 Teilbarkeit

In diesem Abschnitt untersuchen wir den Begriff der Teilbarkeit in einem Ring. Der Einfachheit halber werden wir ab jetzt nur kommutative Ringe behandeln.

Definition 8.2.1 (teilbar, assoziiert)**.** Sei R ein kommutativer Ring und seien $x, y \in R$.

- Man sagt, dass y *durch x teilbar* ist oder dass x *y teilt*, und man schreibt $x|y$, wenn ein Element $t \in R$ mit $tx = y$ existiert. Man sagt dann auch, dass x ein *Teiler* von y ist und dass y ein *Vielfaches* von x ist.
- Man sagt, dass x und y *assoziiert* sind, wenn sie durcheinander teilbar sind.

Bemerkung 8.2.2. Das Hauptideal $(x) \subset R$ ist nach Definition die Menge aller Elemente, die durch x teilbar sind. Das heißt:

$$x|y \iff y \in (x) \iff (y) \subset (x).$$

(die zweite Äquivalenz folgt aus der Definition des erzeugten Untermoduls). Insbesondere: x und y sind genau dann assoziiert, wenn $(x) = (y)$.

Es folgt unmittelbar aus diesen Bemerkungen, dass Teilbarkeit eine reflexive und transitive Relation auf R ist, und dass Assoziiertheit eine Äquivalenzrelation auf R ist. Insbesondere

können wir die Quotientenmenge $R/\text{Assoziiertheit}$ bilden. Um sie zu verstehen betrachten wir die surjektive Abbildung

$$\begin{aligned} R &\twoheadrightarrow \{\text{Hauptideale in } R\}, \\ x &\mapsto (x). \end{aligned}$$

Nach der universellen Eigenschaft der Quotientenmenge (Satz 1.4.10), gibt es eine induzierte bijektive Abbildung

$$\begin{aligned} R/\text{Assoziiertheit} &\xrightarrow{\sim} \{\text{Hauptideale in } R\}, \\ [x] &\mapsto (x). \end{aligned}$$

Allgemeine Ideale in R können wir dann als „verallgemeinerte Elemente“ von R (bis auf Assoziiertheit) auffassen. Historisch gesehen wurden solche verallgemeinerten Elemente zuerst in sogenannten Zahlringen (wie den Ringen $\mathbb{Z}[z]$ aus Beispiel 8.1.9) betrachtet und wurden als *ideale Zahlen* bezeichnet, was später zu *Ideal* abgekürzt wurde.

Bemerkung 8.2.3. Sei R ein kommutativer Ring.

(i) Ein Element $x \in R$ ist genau dann eine Einheit, wenn es 1 teilt. Das heißt:

$$x \in R^\times \iff x|1 \iff 1 \in (x) \iff (x) = R.$$

(ii) Alle Elemente $x \in R$ teilen 0, und 0 teilt nur 0. Anders gesagt ist 0 das eindeutige *größte* Element von R bezüglich der Teilbarkeitsrelation.

In einem Integritätsring können wir die Assoziiertheitsrelation konkreter beschreiben:

Proposition 8.2.4. *Sei R ein Integritätsring. Zwei Elemente $x, y \in R$ sind genau dann assoziiert, wenn eine Einheit $r \in R^\times$ existiert, so dass $rx = y$.*

Beweis. Existiert eine solche Einheit r, so gilt auch $r^{-1}y = x$, d.h., x und y sind durcheinander teilbar. Seien umgekehrt x und y assoziiert. Nach Definition gibt es $r, s \in R$ mit $rx = y$ und $sy = x$. Falls x oder y gleich Null ist, dann gilt $x = y$ und man kann $r = 1$ nehmen. Sonst folgt aus $(rs - 1)y = 0$ und der Integrität von R, dass $rs - 1 = 0$ gilt, d.h., $rs = 1$. Auf ähnliche Weise folgt aus $(sr - 1)x = 0$, dass $sr = 1$. Damit ist r eine Einheit mit $r^{-1} = s$. □

Beispiel 8.2.5.

(i) Da $\mathbb{Z}^\times = \{\pm 1\}$ ist jede ganze Zahl zu genau einer natürlichen Zahl assoziiert.

(ii) Sei K ein Körper. Da $K[T]^\times = K^\times$ (Beispiel 8.1.14(iii)) ist jedes $p \in K[T] \setminus \{0\}$ zu genau einem monischen Polynom assoziiert.

8.2.1 Primelemente

Definition 8.2.6 (prim, irreduzibel)**.** Sei R ein kommutativer Ring und sei $x \in R$.

- x heißt *prim* oder ein *Primelement*, wenn $x \notin \{0\} \cup R^\times$ und für alle $r, s \in R$ gilt:

$$x|rs \implies (x|r \text{ oder } x|s).$$

- x heißt *irreduzibel*, wenn $x \notin \{0\} \cup R^\times$ und für alle $r, s \in R$ gilt:

$$x = rs \implies (r \in R^\times \text{ oder } s \in R^\times).$$

Bemerkung 8.2.7. Die gewöhnliche Definition einer *Primzahl* ist ein positives irreduzibles Element von $\mathbb{Z}$. Ein beliebiges irreduzibles Element in $\mathbb{Z}$ ist dann $\pm$ eine Primzahl. Aber wir werden sehen, dass es im Ring $\mathbb{Z}$ keinen Unterschied zwischen Primelementen und irreduziblen Elementen gibt (siehe Proposition 8.2.9 und Korollar 8.2.18).

Bemerkung 8.2.8. Die Implikation in der Definition eines Primelements kann auch folgendermaße geschrieben werden:

$$rs \in (x) \implies (r \in (x) \text{ oder } s \in (x)).$$

Insbesondere: Ob x ein Primelement ist, hängt nur von dem Ideal (x) ab. Anders gesagt, sind x und y assoziiert, so ist x genau dann prim, wenn y prim ist.

Proposition 8.2.9 (prim vs. irreduzibel)**.**

(i) *In einem Integritätsring ist jedes Primelement irreduzibel.*

(ii) *In einem Hauptidealring ist jedes irreduzible Element prim.*

Beweis. Zu (i). Sei R ein Integritätsring und $p \in R$ ein Primelement. Nach Definition ist $p \notin \{0\} \cup R^\times$. Sei $p = rs$ mit $r, s \in R$. Insbesondere gilt $p|rs$. Da p prim ist, gilt $p|r$ oder $p|s$. Falls $p|r$ sind p und r assoziiert, und nach Proposition 8.2.4 gibt es ein $u \in R^\times$ mit $p = ru$. Aus $ru = rs$ und der Integrität von R folgt $u = s$, und damit ist s eine Einheit. Falls $p|s$ schließen wir ebenso, dass r eine Einheit ist.

Zu (ii). Sei R ein Hauptidealring und sei $p \in R$ ein irreduzibles Element. Nach Definition ist $p \notin \{0\} \cup R^\times$. Sei $p|rs$ mit $r, s \in R$. Zu zeigen ist, dass $p|r$ oder $p|s$. Da R ein Hauptidealring ist, gibt es ein $t \in R$ mit $(p, r) = (t)$. Insbesondere ist p durch t teilbar: $p = tu$ mit einem $u \in R$. Da p irreduzibel ist, gilt $t \in R^\times$ oder $u \in R^\times$. Wir betrachten die beiden Möglichkeiten:

- Falls $t \in R^\times$, dann ist $(p, r) = (t) = R$. Insbesondere ist $1 \in (p, r)$, d.h., es existiert $a, b \in R$ mit $1 = ap + br$. Dann ist $s = aps + brs$. Da p beide Summanden auf der rechten Seite teilt, teilt p auch s.
- Falls $u \in R^\times$, dann sind p und t assoziiert, so dass $(p) = (t) = (p, r)$. Insbesondere ist $r \in (p)$, d.h., p teilt r. □

Beispiel 8.2.10.

(i) Im Ring $\mathbb{Z}[i]$ ist 2 nicht irreduzibel und damit nicht prim, denn $2 = (1 + i)(1 - i)$ und weder $1 + i$ noch $1 - i$ ist eine Einheit in $\mathbb{Z}[i]$ (da die komplexen Zahlen $(1 \pm i)^{-1}$ keine Gaußschen Zahlen sind).

(ii) Sei K ein Körper. Nach Beispiel 8.1.14 besteht $K[T]^\times$ genau aus den Polynomen vom Grad 0. Damit ist ein Polynom vom Grad ≥ 1 genau dann irreduzibel, wenn es kein Produkt von zwei Polynomen vom Grad ≥ 1 ist. Nach Proposition 6.3.7(ii) sind insbesondere alle Polynome vom Grad 1 irreduzibel, und nach Proposition 6.3.15 ist ein Polynom vom Grad 2 oder 3 genau dann irreduzibel, wenn es keine Nullstellen hat. Zum Beispiel: Für alle $a \in \mathbb{R}_{>0}$ ist $T^2 + a$ irreduzibel in $\mathbb{R}[T]$, und $T^3 + T + 1$ ist irreduzibel in $\mathbb{F}_2[T]$. Da $K[T]$ ein Hauptidealring ist (Korollar 8.2.18), gibt es hier keinen Unterschied zwischen irreduziblen Elementen und Primelementen.

(iii) Sei K ein *algebraisch abgeschlossener* Körper. Nach Proposition 6.3.15 ist dann jedes Polynom von Grad ≥ 2 über K ein Produkt von zwei Polynomen vom Grad ≥ 1, und somit nicht irreduzibel. In diesem Fall sind also die irreduziblen Polynome genau die vom Grad 1.

Bemerkung 8.2.11. Die Umkehrung von Proposition 8.2.9(i) gilt im Allgemeinen nicht. Zum Beispiel: Im Integritätsring $\mathbb{Z}[\sqrt{-5}]$ kann man zeigen, dass 3 irreduzibel ist, aber 3 ist kein Primelement von $\mathbb{Z}[\sqrt{-5}]$, denn 3 teilt $9 = (2 + i\sqrt{5})(2 - i\sqrt{5})$, ohne einen der beiden Faktoren zu teilen. Aus dem Beweis von Proposition 8.2.9(ii) schließen wir, dass das Ideal $(3, 2 + i\sqrt{5})$ in $\mathbb{Z}[\sqrt{-5}]$ kein Hauptideal ist.

8.2.2 Euklidische Ringe

Euklidische Ringe sind grob gesagt Ringe, in denen Division mit Rest möglich ist. Zum Beispiel sind bekanntlich $\mathbb{Z}$ und der Polynomring $K[T]$ über einem Körper K euklidische Ringe. In diesem Abschnitt beweisen wir, dass jeder euklidische Ring ein Hauptidealring ist. Damit erhalten wir viele Beispiele von Hauptidealringen.

Definition 8.2.12 (euklidische Gradfunktion, euklidischer Ring)**.**

- Sei R ein kommutativer Ring. Eine *euklidische Gradfunktion* auf R ist eine Abbildung $\delta\colon R \setminus \{0\} \to \mathbb{N}$ mit folgender Eigenschaft: Für alle $a, b \in R$ mit $b \neq 0$ gibt es $q, r \in R$ (den „Quotient“ und den „Rest“) mit
$$a = qb + r \quad \text{und} \quad (r = 0 \text{ oder } \delta(r) < \delta(b)).$$
- Ein *euklidischer Ring* ist ein Integritätsring, auf dem eine euklidische Gradfunktion existiert.

Beispiel 8.2.13 (ganze Zahlen)**.** Die Abbildung
$$\begin{aligned}\delta\colon \mathbb{Z} \setminus \{0\} &\to \mathbb{N},\\ x &\mapsto |x|,\end{aligned}$$
ist eine euklidische Gradfunktion auf $\mathbb{Z}$, denn: Seien $a, b \in \mathbb{Z}$ mit $b \neq 0$. Wir suchen q und r wie in der Definition einer euklidischen Gradfunktion. Falls $a = 0$ können wir $q = r = 0$ nehmen. Ohne Beschränkung der Allgemeinheit (indem wir die Vorzeichen von q und r anpassen) können wir annehmen, dass a und b positiv sind. Die Menge $A = \{n \in \mathbb{N} \mid nb \leq a\}$ ist nicht leer (sie enthält 0) und endlich (da $b > 0$). Sei q das Maximum von A und sei $r = a - qb$. Es gilt dann $(q + 1)b > a$, d.h., $r < b$, wie gewünscht.

Beispiel 8.2.14 (Polynome über einem Körper)**.** Sei K ein Körper. Dann ist die Abbildung
$$\begin{aligned}\delta\colon K[T] \setminus \{0\} &\to \mathbb{N},\\ p &\mapsto \deg(p),\end{aligned}$$
eine euklidische Gradfunktion auf $K[T]$. Dies folgt aus Satz 6.3.12.

Beispiel 8.2.15 (Gaußsche Zahlen)**.** Die Abbildung
$$\begin{aligned}\delta\colon \mathbb{Z}[i] \setminus \{0\} &\to \mathbb{N},\\ x &\mapsto |x|^2,\end{aligned}$$
ist eine euklidische Gradfunktion auf $\mathbb{Z}[i]$, denn: Seien $a, b \in \mathbb{Z}[i]$ mit $b \neq 0$. Um q und r zu finden, betrachten wir zunächst den Quotient $c = b^{-1}a \in \mathbb{C}$. Nach elementarer Geometrie ($\mathbb{Z}[i]$ ist ein quadratisches Gitter in $\mathbb{C}$) gibt es dann ein $q \in \mathbb{Z}[i]$ mit $|c - q| \leq \frac{1}{\sqrt{2}}$. Sei $r = a - qb$. Dann gilt
$$|r|^2 = |a - qb|^2 = |cb - qb|^2 \leq \frac{1}{2}|b|^2 < |b|^2,$$
wie gewünscht.

Beispiel 8.2.16 (Eisenstein-Zahlen)**.** Die Abbildung
$$\begin{aligned}\delta\colon \mathbb{Z}[\omega] \setminus \{0\} &\to \mathbb{N},\\ x &\mapsto |x|^2,\end{aligned}$$
ist eine euklidische Gradfunktion auf $\mathbb{Z}[\omega]$. Der Beweis ist ähnlich wie bei den Gaußschen Zahlen, aber $\mathbb{Z}[\omega] \subset \mathbb{C}$ ist jetzt ein regelmäßiges dreieckiges Gitter. Ist $c = b^{-1}a \in \mathbb{C}$, so gibt es eine nächste Eisenstein-Zahl $q \in \mathbb{Z}[\omega]$ mit $|c - q| \leq \frac{1}{\sqrt{3}}$. Für $r = a - qb$ gilt dann
$$|r|^2 = |a - qb|^2 = |cb - qb|^2 \leq \frac{1}{3}|b|^2 < |b|^2.$$

Proposition 8.2.17. *Jeder euklidische Ring ist ein Hauptidealring.*

Beweis. Sei R ein euklidischer Ring und sei $\delta\colon R \setminus \{0\} \to \mathbb{N}$ eine euklidische Gradfunktion auf R. Sei $I \subset R$ ein beliebiges Ideal mit $I \neq \{0\}$. Da $\mathbb{N}$ wohlgeordnet ist, hat die Teilmenge $\delta(I \setminus \{0\}) \subset \mathbb{N}$ ein kleinstes Element. Das heißt, es gibt ein $a \in I \setminus \{0\}$, so dass für alle $x \in I \setminus \{0\}$ gilt $\delta(a) \leq \delta(x)$.

Behauptung. Es gilt $I = (a)$.

Die Inklusion $(a) \subset I$ ist klar. Sei $x \in I$ beliebig. Dann existieren $q, r \in R$ mit $x = qa + r$ und entweder $r = 0$ oder $\delta(r) < \delta(a)$. Da I beide x und qa enthält, liegt auch $r = x - qa$ in I. Falls $r \neq 0$ gilt dann $\delta(r) \geq \delta(a)$ nach der Wahl von a. Also muss $r = 0$ sein, und damit ist x durch a teilbar, d.h., $x \in (a)$. □

Korollar 8.2.18. *Die folgenden Ringe sind Hauptidealringe:* $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$ *und* $K[T]$ *für einen Körper* K.

Beweis. Folgt aus Proposition 8.2.17 und den Beispielen 8.2.13, 8.2.14, 8.2.15 und 8.2.16. □

Bemerkung 8.2.19. Es gibt Hauptidealringe, die nicht euklidisch sind, aber das ist nicht trivial. Das einfachste Beispiel ist der von $\frac{1}{2}(1 + i\sqrt{19})$ erzeugte Unterring von $\mathbb{C}$.

***Beispiel 8.2.20** (formale Potenzreihen)**.** Sei R ein kommutativer Ring. Eine *formale Potenzreihe* über R in einer Variablen T ist ein Ausdruck der Gestalt $\sum_{i=0}^{\infty} a_i T^i$ mit $a_i \in R$. Formale Potenzreihen bilden eine kommutative R-Algebra $R[[T]]$: der unterliegende R-Modul ist $R^{\mathbb{N}}$, und die Multiplikation wird genau wie bei Polynomen definiert, d.h.:

$$\left(\sum_{i=0}^{\infty} a_i T^i\right) \cdot \left(\sum_{j=0}^{\infty} b_j T^j\right) = \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i b_j\right) T^k.$$

Diese Formel zeigt, dass jede formale Potenzreihe $\sum_{i=0}^{\infty} a_i T^i$ mit $a_0 \in R^\times$ eine Einheit in $R[[T]]$ ist: Die inverse Potenzreihe $\sum_{j=0}^{\infty} b_j T^j$ hat $b_0 = a_0^{-1}$, und b_j für $j \geq 1$ wird rekursiv durch

$$b_j = -a_0^{-1} \sum_{i=1}^{j} a_i b_{j-i}$$

bestimmt. Daraus folgt:

$$R[[T]]^\times = \left\{ \sum_{i=0}^{\infty} a_i T^i \in R[[T]] \;\middle|\; a_0 \in R^\times \right\}.$$

Ist K ein Körper, so ist die Abbildung

$$\begin{aligned} \delta\colon K[[T]] \setminus \{0\} &\to \mathbb{N}, \\ \sum_{i=0}^{\infty} a_i T^i &\mapsto \min\{i \in \mathbb{N} \mid a_i \neq 0\}, \end{aligned}$$

eine euklidische Gradfunktion auf $K[[T]]$. Dazu bemerken wir, dass jedes $f \in K[[T]] \setminus \{0\}$ als $f = T^{\delta(f)} \bar{f}$ geschrieben werden kann, wobei $\delta(\bar{f}) = 0$ und damit $\bar{f} \in K[[T]]^\times$. Also ist jedes f zu $T^{\delta(f)}$ assoziiert. Daraus folgt:

$$g | f \iff T^{\delta(g)} | T^{\delta(f)} \iff \delta(g) \leq \delta(f),$$

so dass die Bedingung für eine euklidische Gradfunktion auf triviale Weise erfüllt ist (man kann immer $q = 0$ oder $r = 0$ nehmen). Also ist $K[[T]]$ ein euklidischer Ring und somit ein Hauptidealring.

8.2.3 Faktorielle Ringe

Faktorielle Ringe sind grob gesagt Ringe, in denen jedes Element ein Produkt von Primelementen ist. In diesem Abschnitt zeigen wir, dass jeder Hauptidealring faktoriell ist. Da $\mathbb{Z}$ ein Hauptidealring ist, erhalten wir insbesondere einen Beweis des *Fundamentalsatzes der Arithmetik* (Satz 1.1.14): Jede natürliche Zahl besitzt eine eindeutige Primfaktorzerlegung.

Definition 8.2.21 (Primfaktorzerlegung, faktorieller Ring)**.**

- Sei R ein kommutativer Ring. Eine *Primfaktorzerlegung* eines Elements $r \in R \setminus \{0\}$ ist eine Darstellung
$$r = up_1 \dots p_n,$$
wobei $u \in R^\times$, $n \in \mathbb{N}$ und $p_1, \dots, p_n \in R$ Primelemente sind.
- Ein *faktorieller Ring* oder *ZPE-Ring* (für „Zerlegung in Primelemente“) ist ein Integritätsring, in dem jedes Nicht-Null-Element eine Primfaktorzerlegung besitzt.

Proposition 8.2.22 (Eindeutigkeit der Primfaktozerlegung)**.** *Sei R ein Integritätsring und sei $r \in R \setminus \{0\}$. Wenn r eine Primfaktorzerlegung besitzt, dann ist sie* im Wesentlichen eindeutig *im folgenden Sinne: Ist*

$$r = up_1 \dots p_n = vq_1 \dots q_m$$

mit Einheiten u, v und Primelementen p_i, q_j, so gibt es eine Bijektion $\sigma\colon \{1, \dots, n\} \to \{1, \dots, m\}$, so dass p_i und $q_{\sigma(i)}$ assoziiert sind für alle $i \in \{1, \dots, n\}$.

Beweis. Wir verwenden Induktion über n. Wenn $n = 0$ ist r eine Einheit. Ein Produkt von Elementen in einem kommutativen Ring kann aber nur eine Einheit sein, wenn alle Faktoren Einheiten sind (denn die Einheiten sind genau die Teiler von 1). Da Primelemente keine Einheiten sind, muss auch $m = 0$ sein. Sei nun $n \geq 1$. Da p_1 prim ist und r teilt, gibt es ein $j \in \{1, \dots, m\}$, so dass $p_1 | q_j$. Da q_j irreduzibel ist (Proposition 8.2.9(i)), gilt $q_j = u_1 p_1$ mit einer Einheit u_1. Es gilt dann

$$up_2 \dots p_n = vu_1 q_1 \dots q_{j-1} q_{j+1} \dots q_m.$$

Nach Induktionsvoraussetzung gibt es eine Bijektion $\tau\colon \{2, \dots, n\} \to \{1, \dots, m\} \setminus \{j\}$, so dass p_i und $q_{\tau(i)}$ assoziiert sind für alle $i \in \{2, \dots, n\}$. Definiert man nun σ durch $\sigma(1) = j$ und $\sigma(i) = \tau(i)$ für $i \in \{2, \dots, n\}$, so hat σ die gewünschte Eigenschaft. □

Lemma 8.2.23. *Sei R ein Hauptidealring und sei $(I_n)_{n \in \mathbb{N}}$ eine aufsteigende Folge von Idealen in R, d.h., es gilt $I_n \subset I_{n+1}$ für alle $n \in \mathbb{N}$. Dann ist die Folge $(I_n)_{n \in \mathbb{N}}$ stationär, d.h., es gibt ein $N \in \mathbb{N}$, so dass für alle $n \geq N$ gilt $I_n = I_N$.*

Beweis. Sei $I = \bigcup_{n \in \mathbb{N}} I_n$. Nach Lemma 6.4.1 ist I wieder ein Untermodul von R, d.h., ein Ideal. Da R ein Hauptidealring ist, gibt es ein $x \in R$ mit $I = (x)$. Nach Definition der Vereinigung gibt es dann ein $N \in \mathbb{N}$ mit $x \in I_N$. Für alle $n \geq N$ gilt nun

$$(x) \subset I_N \subset I_n \subset I = (x),$$

und somit $I_n = I_N$. □

Proposition 8.2.24. *Jeder Hauptidealring ist faktoriell.*

Beweis. Sei R ein Hauptidealring, und sei $X \subset R \setminus \{0\}$ die Menge aller Elemente, die keine Primfaktorzerlegung besitzen.

Behauptung. Für alle $x \in X$ gibt es ein $y \in X$ mit $(x) \subsetneq (y)$.

Diese Behauptung impliziert, dass X leer sein muss: Sonst würde eine nicht-stationäre aufsteigende Folge von Idealen in R existieren, im Widerspruch zum Lemma 8.2.23.

Ein Element $x \in X$ ist keine Einheit und kein Primelement. Nach Proposition 8.2.9(ii) ist x auch nicht irreduzibel. Das heißt, es gibt eine Zerlegung $x = yz$, wobei y und z keine Einheiten sind. Da $x \in X$ muss $y \in X$ oder $z \in X$. Ohne Einschränkung sei $y \in X$. Da R ein Integritätsring ist und z keine Einheit ist, sind x und y *nicht* assoziiert. Das heißt, es gilt $(x) \subsetneqq (y)$, wie behauptet. □

Korollar 8.2.25. *Die folgenden Ringe sind faktoriell: $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$ und $K[T]$ für einen Körper K.*

Beweis. Folgt aus Proposition 8.2.24 und Korollar 8.2.18. □

Bemerkung 8.2.26. Die Umkehrung von Proposition 8.2.24 gilt nicht. Eigentlich ist die Bedingung, ein Hauptidealring zu sein, viel strenger als die, faktoriell zu sein. Zum Beispiel sagt das *Lemma von Gauß*, das in der Vorlesung *Algebra* bewiesen wird, dass der Polynomring $R[T]$ über einem faktoriellen Ring R wieder faktoriell ist. Damit sind $\mathbb{Z}[T]$ und die Polynomringe in ≥ 2 Variablen über einem Körper faktorielle Ringe, die keine Hauptidealringe sind.

Proposition 8.2.9(ii) verallgemeinert sich auf faktorielle Ringe:

Proposition 8.2.27. *Sei R ein faktorieller Ring. Ein Element von R ist genau dann prim, wenn es irreduzibel ist.*

Beweis. Da R ein Integritätsring ist, ist jedes Primelement irreduzibel (Proposition 8.2.9(i)). Sei umgekehrt $x \in R$ ein irreduzibles Element, und sei $x = up_1 \dots p_n$ eine Primfaktorzerlegung von x. Wir zeigen, dass $n = 1$ gilt, so dass $x = up_1$ ein Primelement ist. Da x keine Einheit ist, muss $n \geq 1$ sein. Angenommen wäre $n \geq 2$. Nach Definition der Irreduzibilität muss dann up_1 oder $p_2 \dots p_n$ eine Einheit sein. Insbesondere ist p_1 oder p_2 eine Einheit, im Widerspruch zur Definition eines Primelements. □

Bemerkung 8.2.28. Aus Bemerkung 8.2.11 und Proposition 8.2.27 folgt, dass der Ring $\mathbb{Z}[\sqrt{-5}]$ nicht einmal faktoriell ist.

Definition 8.2.29 (Vielfachheit)**.** Sei R ein faktorieller Ring, sei $r \in R$ und sei $p \in R$ ein Primelement. Die *Vielfachheit* von p in r ist

$$v_p(r) = \sup\{n \in \mathbb{N} \mid p^n \text{ teilt } r\} \in \mathbb{N} \cup \{+\infty\}.$$

Man beachte dabei, dass $v_p(r)$ nur von der Äquivalenzklasse von p modulo Assoziiertheit (d.h., von dem Ideal (p)) abhängt. Zudem gilt $v_p(0) = +\infty$. Ist

$$r = up_1 \dots p_n$$

eine Primfaktorzerlegung von $r \in R \setminus \{0\}$, so folgt aus der Eindeutigkeit der Primfaktorzerlegung, dass

$$v_p(r) = |\{i \in \{1, \dots, n\} \mid p \text{ und } p_i \text{ sind assoziiert}\}|.$$

Insbesondere gibt es nur endlich viele Assoziiertheitsklassen von Primelementen p, so dass $v_p(r) \neq 0$.

Bemerkung 8.2.30. Die Vielfacheit von $a \in K$ in einem Polynom $f \in K[T]$ im Sinne der Definition 6.3.19 ist genau die Vielfachheit des Primelements $T - a$ in f im Sinne der Definition 8.2.29.

Bemerkung 8.2.31. Man kann die Primfaktorzerlegung in einem faktoriellen Ring R eindeutiger machen, indem man ein Repräsentantensystem P der Primelemente von R modulo Assoziiertheit auswählt (Definition 1.4.8). Dann kann jedes $r \in R \setminus \{0\}$ als

$$r = u \cdot \prod_{p \in P} p^{v_p(r)}$$

dargestellt werden, mit einer eindeutigen Einheit $u \in R^\times$. Obwohl P unendlich sein kann, ist das obige Produkt endlich, da es nur endlich viele $p \in P$ mit $v_p(r) \neq 0$ gibt. Zum Beispiel (siehe Beispiel 8.2.5):

(i) Die Menge aller Primzahlen aus $\mathbb{N}$ ist ein Repräsentantensystem der Primelemente von $\mathbb{Z}$ bis auf Assoziiertheit. Jedes $r \in \mathbb{Z} \setminus \{0\}$ hat damit eine eindeutige Darstellung $r = \pm p_1 \dots p_n$, wobei die p_i Primzahlen sind.

(ii) Ist K ein Körper, so ist die Menge aller *monischen* irreduziblen Polynome über K ein Repräsentantensystem der Primelemente von $K[T]$ bis auf Assoziiertheit. Jedes $f \in K[T] \setminus \{0\}$ hat damit eine eindeutige Darstellung $f = up_1 \dots p_n$, wobei $u \in K^\times$ und die p_i monische irreduzible Polynome sind.

Proposition 8.2.32 (Teilbarkeitskriterium)**.** *Sei R ein faktorieller Ring und seien $r, s \in R$. Dann sind die folgenden Aussagen äquivalent:*

(i) *r teilt s.*

(ii) *Für alle Primelemente $p \in R$ gilt $v_p(r) \leq v_p(s)$.*

Beweis. Die Implikation (i) $\Rightarrow$ (ii) folgt unmittelbar aus der Definition der Vielfachheit. Die Implikation (ii) $\Rightarrow$ (i) ist auch klar wegen der Darstellung $r = u \cdot \prod_{p \in P} p^{v_p(r)}$ aus Bemerkung 8.2.31. □

Beispiel 8.2.33.

(i) Folgende Zerlegungen sind Primfaktorzerlegungen in $\mathbb{Z}$:

$$\begin{aligned} 10 &= 2 \cdot 5, \\ 56 &= 2^3 \cdot 7, \\ 57 &= 3 \cdot 19, \\ -36 &= -2^2 \cdot 3^2. \end{aligned}$$

(ii) Sei K ein algebraisch abgeschlossener Körper und sei $f \in K[T] \setminus \{0\}$. Die Primfaktorzerlegung von f lautet

$$f = u(T - a_1)^{v_{a_1}(f)} \dots (T - a_n)^{v_{a_n}(f)},$$

wobei $u \in K^\times$ der Leitkoeffizient von f ist und $a_1, \dots, a_n \in K$ die Nullstellen von f sind.

(iii) Die Zerlegung

$$2T^5 + 10T^3 + 12T = 2T(T^2 + 2)(T^2 + 3)$$

ist eine Primfaktorzerlegung in $\mathbb{R}[T]$.

Bemerkung 8.2.34. Primzahlen spielen eine wichtige Rolle in der Kryptographie und damit in dem Alltag: Die Sicherheit von Bankgeschäften und Online-Authentifizierungsverfahren (wie bei den HTTPS und SSH Protokollen) beruht auf dem RSA-Algorithmus, dessen Wirksamkeit hängt davon ab, dass die Primfaktorzerlegung einer natürlichen Zahl mit großen Primfaktoren selbst mit einem Computer nicht in einer angemessenen Zeit berechenbar ist. (Das Zerlegen in Primfaktoren ist theoretisch einfach mit einem Quantencomputer, so dass der RSA-Algorithmus nicht „quantensicher“ ist, aber ob solche Quantenalgorithmen in der Praxis laufen können, muss noch bestätigt werden.)

8.2.4 Größte gemeinsame Teiler und kleinste gemeinsame Vielfache

Definition 8.2.35 (größter gemeinsamer Teiler, kleinstes gemeinsames Vielfaches)**.** Sei R ein kommutativer Ring und $E \subset R$ eine Teilmenge.

- Ein *größter gemeinsamer Teiler* von E ist ein Element $d \in R$ mit folgender Eigenschaft: d teilt jedes Element von E, und teilt $s \in R$ jedes Element von E, so teilt s auch d.
- Ein *kleinstes gemeinsames Vielfaches* von E ist ein Element $m \in R$ mit folgender Eigenschaft: m ist durch jedes Element von E teilbar, und ist $s \in R$ durch jedes Element von E teilbar, so ist s auch durch m teilbar.

Bemerkung 8.2.36. Nach der Bemerkung 8.2.2 können wir diese Definitionen auch folgendermaße formulieren. Ein größter gemeinsamer Teiler von E ist ein Element $d \in R$, so dass (d) ein *kleinstes* Element der partiell geordneten Menge

$$(\{I \subset R \mid I \text{ Hauptideal und } E \subset I\}, \subset)$$

ist. Auf ähnliche Weise ist ein kleinstes gemeinsames Vielfaches von E ein Element $m \in R$, so dass (m) ein *größtes* Element der partiell geordneten Menge

$$(\{I \subset R \mid I \text{ Hauptideal und } I \subset \textstyle\bigcap_{e\in E}(e)\}, \subset)$$

ist.

Proposition 8.2.37. *Sei R ein kommutativer Ring und $E \subset R$ eine Teilmenge.*

(i) *Sind d und d' größte gemeinsame Teiler von E, so sind d und d' assoziiert.*

(ii) *Sind m und m' kleinste gemeinsame Vielfache von E, so sind m und m' assoziiert.*

(iii) *Ist das Ideal (E) ein Hauptideal, so ist jedes erzeugende Element von (E) ein größter gemeinsamer Teiler von E.*

(iv) *Ist das Ideal $\bigcap_{e\in E}(e)$ ein Hauptideal, so ist jedes erzeugende Element von $\bigcap_{e\in E}(e)$ ein kleinstes gemeinsames Vielfaches von E.*

Beweis. Aussagen (i) und (ii) folgen unmittelbar aus den Definitionen: Zwei größte gemeinsame Teiler bzw. kleinste gemeinsame Vielfache sind durcheinander teilbar. Aussagen (iii) und (iv) folgen aus Bemerkung 8.2.36. □

Notation 8.2.38. Man schreibt $\mathrm{ggT}(E)$ für einen größten geimeinsamen Teiler von E und $\mathrm{kgV}(E)$ für ein kleinstes gemeinsames Vielfaches von E. Nach Proposition 8.2.37(i,ii) sind diese Elemente eindeutig bis auf Assoziiertheit, wenn sie existieren.

Bemerkung 8.2.39 (ggT und kgV in Hauptidealringen)**.** Sei R ein Hauptidealring und $E \subset R$ eine Teilmenge. Nach Proposition 8.2.37(iii,iv) existieren $\mathrm{ggT}(E)$ und $\mathrm{kgV}(E)$, und es gilt:

$$(\mathrm{ggT}(E)) = (E) \quad \text{und} \quad (\mathrm{kgV}(E)) = \bigcap_{e\in E}(e).$$

Proposition 8.2.40 (ggT und kgV in faktoriellen Ringen)**.** *Sei R ein faktorieller Ring und $E \subset R$ eine Teilmenge. Dann existieren* $\mathrm{ggT}(E)$ *und* $\mathrm{kgV}(E)$*. Ist $P \subset R$ ein Repräsentantensystem der Primelemente von R bis auf Assoziiertheit, so gilt genauer:*

(i) $\mathrm{ggT}(E) = \prod_{p\in P} p^{\inf\{v_p(e) \mid e\in E\}}$.

(ii) $\mathrm{kgV}(E) = \prod_{p\in P} p^{\sup\{v_p(e) \mid e\in E\}}$.

Dabei gelten die folgenden Konventionen: $\inf(\varnothing) = +\infty$, $\sup(\varnothing) = -\infty$, $p^{+\infty} = 0$, $p^{-\infty} = 1$, *und ein Produkt mit unendlich vielen Faktoren $\neq 1$ ist null.*

Beweis. Zunächst beobachten wir, dass die Konventionen in Grenzfällen das Gewünschte leisten: In einem beliebigen kommutativen Ring gilt $\mathrm{ggT}(\varnothing) = 0$, $\mathrm{kgV}(\varnothing) = 1$, $\mathrm{ggT}(\{0\}) = 0$ und $\mathrm{kgV}(E) = 0$ wenn $0 \in E$. Wenn $E \neq \varnothing$ und $0 \notin E$, kann noch das Produkt in (ii) eine unendliche Potenz enthalten (wenn es ein $p \in P$ mit $\{v_p(e) \mid e \in E\}$ unbeschränkt gibt) oder unendlich viele Faktoren $\neq 1$ haben (wenn es unendlich viele $p \in P$ mit $\{v_p(e) \mid e \in E\} \neq \{0\}$ gibt). In diesen Fällen gilt auch $\mathrm{kgV}(E) = 0$, denn kein $r \in R \setminus \{0\}$ kann durch alle Elemente von E teilbar sein nach dem Teilbarkeitskriterium 8.2.32. In allen anderen Fällen haben wir endliche Produkte von endlichen Potenzen, und die Aussage folgt unmittelbar aus dem Teilbarkeitskriterium 8.2.32. □

Beispiel 8.2.41. Im Ring $\mathbb{Z}$ gilt $56 = 2^3 \cdot 7$ und $60 = 2^2 \cdot 3 \cdot 5$. Damit ist $\mathrm{ggT}(56, 60) = 2^2 = 4$ und $\mathrm{kgV}(56, 60) = 2^3 \cdot 3 \cdot 5 \cdot 7 = 840$.

Algorithmus 8.2.42 (der euklidische Algorithmus)**.** Sei R ein euklidischer Ring und sei $(a, b) \in R^2$. Gesucht ist $\mathrm{ggT}(a, b)$. Angenommen wird ein Algorithmus für die Division mit Rest bezüglich einer euklidischen Gradfunktion $\delta\colon R \setminus \{0\} \to \mathbb{N}$ gegeben (solche Algorithmen sind bekannt für $R = \mathbb{Z}$ oder $R = K[T]$, siehe Beispiel 6.3.13).

- Ist $b = 0$, so gilt $\mathrm{ggT}(a, b) = a$.
- Ist $b \neq 0$, so bestimmen wir $q, r \in R$ mit

  $$a = qb + r \quad \text{und} \quad (r = 0 \text{ oder } \delta(r) < \delta(b)).$$

  Dann gilt $\mathrm{ggT}(a, b) = \mathrm{ggT}(b, r)$, denn ein Element teilt a und b genau dann, wenn es b und r teilt.

Man wendet dann rekursiv den Algorithmus mit dem Paar $(b, r) \in R^2$ an. Nach höchstens $\delta(b) + 1$ Schritten wird das zweite Element zu Null und somit terminiert der Algorithmus. In jedem Schritt dürfen wir auch a und b durch assoziierte Elemente ersetzen.

Beispiel 8.2.43. Wir berechnen $\mathrm{ggT}(42, 30)$ in $\mathbb{Z}$ mit dem euklidischen Algorithmus:

- $(42, 30)$: Division mit Rest liefert $42 = 1 \cdot 30 + 12$.
- $(30, 12)$: Division mit Rest liefert $30 = 2 \cdot 12 + 6$.
- $(12, 6)$: Division mit Rest liefert $12 = 2 \cdot 6 + 0$.
- $(6, 0)$: Der Algorithmus terminiert und es gilt $\mathrm{ggT}(42, 30) = 6$.

Beispiel 8.2.44. Sei K ein Körper. Wir berechnen $\mathrm{ggT}(T^6 - 1, T^3 - 3T^2 + 3T - 2)$ in $K[T]$ mit dem euklidischen Algorithmus:

- $(T^6 - 1, T^3 - 3T^2 + 3T - 2)$: Division mit Rest liefert

  $$T^6 - 1 = (T^3 + 3T^2 + 6T + 11)(T^3 - 3T^2 + 3T - 2) + 21T^2 - 21T + 21.$$

- $(T^3 - 3T^2 + 3T - 2, 21T^2 - 21T + 21)$: Hier müssen wir zwei Fälle unterscheiden. Falls $\mathrm{char}(K) \in \{3, 7\}$ terminiert bereits der Algorithmus und es gilt

  $$\mathrm{ggT}(T^6 - 1, T^3 - 3T^2 + 3T - 2) = T^3 - 3T^2 + 3T - 2.$$

  Sonst können wir das zweite Polynom durch die Einheit 21 teilen, und Division mit Rest liefert:

  $$T^3 - 3T^2 + 3T - 2 = (T - 2)(T^2 - T + 1) + 0.$$

- $(T^2 - T + 1, 0)$: Falls $\mathrm{char}(K) \notin \{3, 7\}$ terminiert hier der Algorithmus und es gilt

  $$\mathrm{ggT}(T^6 - 1, T^3 - 3T^2 + 3T - 2) = T^2 - T + 1.$$

Bemerkung 8.2.45 (Lemma von Bézout)**.** Sei R ein Hauptidealring und seien $a, b \in R$. Nach Bemerkung 8.2.39 gilt dann $(a, b) = (\mathrm{ggT}(a, b))$, und insbesondere gibt es Elemente $u, v \in R$, so dass $ua + vb = \mathrm{ggT}(a, b)$. Diese Aussage für $R = \mathbb{Z}$ wird manchmal als *Lemma von Bézout* bezeichnet. Im Fall einem euklidischen Ring R können wir solche Elemente u und v mit dem euklidischen Algorithmus bestimmen, durch Rückwärtssubstitution ab dem vorletzten Schritt. Zur Veranschaulichung behandeln wir das Beispiel 8.2.43:

$$6 = 1 \cdot 30 - 2 \cdot 12 = 1 \cdot 30 - 2 \cdot (42 - 1 \cdot 30) = 3 \cdot 30 - 2 \cdot 42.$$

Definition 8.2.46 (teilerfremd, komaximal)**.** Sei R ein kommutativer Ring.

- Zwei Elemente $x, y \in R$ heißen *teilerfremd*, wenn $\mathrm{ggT}(x, y) = 1$.
- Zwei Ideale $I, J \subset R$ heißen *komaximal*, wenn $I + J = R$.

Bemerkung 8.2.47. Sind die Hauptideale (x) und (y) komaximal, so sind x und y teilerfremd nach Proposition 8.2.37(iii). In einem Hauptidealring R gilt auch die Umkehrung, denn $(x) + (y) = (x, y) = (\mathrm{ggT}(x, y))$ (Bemerkung 8.2.39). Im Allgemeinen gilt die Umkehrung aber nicht, selbst wenn R faktoriell ist: Zum Beispiel sind 2 und T in $\mathbb{Z}[T]$ teilerfremd, aber (2) und (T) sind nicht komaximal, da $(2, T) \neq \mathbb{Z}[T]$.

Bemerkung 8.2.48. Da $R = (1)$ sind zwei Ideale I und J genau dann komaximal, wenn $1 \in I + J$.

Satz 8.2.49 (chinesischer Restsatz)**.** *Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$ und seien $I_1, \ldots, I_n \subset R$ Ideale in R, die* paarweise komaximal *sind. Dann ist die R-lineare Abbildung*

$$\varphi \colon R \to \prod_{i=1}^{n} R/I_i, \quad \varphi(x) = (x + I_1, \ldots x + I_n),$$

surjektiv mit $\ker \varphi = \bigcap_{i=1}^{n} I_i$*. Insbesondere gibt es einen induzierten Isomorphismus*

$$\bar{\varphi} \colon R/\left(\textstyle\bigcap_{i=1}^{n} I_i\right) \xrightarrow{\sim} \prod_{i=1}^{n} R/I_i.$$

Beweis. Der Kern der Quotientenabbildung $\varphi_i \colon R \to R/I_i$ ist genau das Ideal I_i. Da $\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x))$ gilt

$$\ker \varphi = \{x \in R \mid \varphi(x) = 0\} = \bigcap_{i=1}^{n} \{x \in R \mid \varphi_i(x) = 0\} = \bigcap_{i=1}^{n} I_i.$$

Damit ist $\bar{\varphi}$ injektiv, und es bleibt zu zeigen, dass φ surjektiv ist. Sei $e_i \in \prod_{i=1}^{n} R/I_n$ das n-Tupel $(\delta_{i1} + I_1, \ldots, \delta_{in} + I_n)$, d.h., alle Koordinaten von e_i sind null außer der i-te Koordinate, die gleich $1 + I_i$ ist. Dann ist $\{e_1, \ldots, e_n\}$ ein Erzeugendensystem des R-Moduls $\prod_{i=1}^{n} R/I_n$. Damit genügt es zu zeigen, dass für alle $i \in \{1, \ldots, n\}$ gilt $e_i \in \operatorname{im} \varphi$.

Behauptung. Für alle $i \in \{1, \ldots, n\}$ sind die Ideale I_i und $\bigcap_{j \neq i} I_j$ komaximal.

Zu jedem $j \neq i$ gibt es Elemente $x_j \in I_j$ und $y_j \in I_i$ mit $x_j + y_j = 1$, da I_j und I_i komaximal sind. Dann gilt

$$1 = \prod_{j \neq i} (x_j + y_j) = \prod_{j \neq i} x_j + \sum_{j \neq i} y_j(\ldots) \in \bigcap_{j \neq i} I_j + I_i,$$

wie behauptet.

Nach der Behauptung gibt es zu jedem $i \in \{1, \ldots, n\}$ Elemente $r_i \in I_i$ und $s_i \in \bigcap_{j \neq i} I_j$ mit $r_i + s_i = 1$. Es gilt dann $\varphi_j(s_i) = 0$ falls $j \neq i$, da $s_i \in I_j$, und es gilt $\varphi_i(s_i) = \varphi_i(1 - r_i) = \varphi_i(1) = 1 + I_i$, da $r_i \in I_i$. Also gilt $\varphi(s_i) = e_i$, wie gewünscht. □

Korollar 8.2.50 (chinesischer Restsatz für Hauptidealringe)**.** *Sei R ein Hauptidealring, sei $n \in \mathbb{N}$ und seien $a_1, \dots a_n \in R$ Elemente von R, die* paarweise teilerfremd *sind. Dann ist die R-lineare Abbildung*

$$\varphi\colon R \to \prod_{i=1}^{n} R/(a_i), \quad \varphi(x) = (x + (a_1), \dots x + (a_n)),$$

surjektiv mit $\ker \varphi = (a_1 \cdot \ldots \cdot a_n)$*. Insbesondere gibt es einen induzierten Isomorphismus*

$$\bar{\varphi}\colon R/(a_1 \cdot \ldots \cdot a_n) \xrightarrow{\sim} \prod_{i=1}^{n} R/(a_i).$$

Beweis. Da die Elemente a_i paarweise teilerfremd sind, sind die Ideale (a_i) paarweise komaximal nach Bemerkung 8.2.47. Zudem gilt $\mathrm{kgV}(a_1, \dots, a_n) = a_1 \cdot \ldots \cdot a_n$ nach Proposition 8.2.40(ii). Nach Proposition 8.2.37(iv) gilt dann $\bigcap_{i=1}^{n}(a_i) = (a_1 \cdot \ldots \cdot a_n)$. Mit diesen Vorbemerkungen ist nun das Korollar ein Sonderfall des Satzes 8.2.49. □

Beispiel 8.2.51.

(i) Für Primzahlen $p \neq q$ liefert der chinesische Restsatz einen Isomorphismus von abelschen Gruppen

$$\mathbb{Z}/pq\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}, \quad n + pq\mathbb{Z} \mapsto (n + p\mathbb{Z}, n + q\mathbb{Z}).$$

Im Gegensatz dazu sind die abelschen Gruppen $\mathbb{Z}/p^2\mathbb{Z}$ und $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ *nicht* isomorph: In $\mathbb{Z}/p^2\mathbb{Z}$ gibt es ein Element x mit $p \cdot x \neq 0$ (z.B., $x = 1 + p^2\mathbb{Z}$), aber in $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ gilt $p \cdot x = 0$ für alle Elemente x.

(ii) Es gibt Isomorphismen von abelschen Gruppen

$$\begin{aligned}\mathbb{Z}/30\mathbb{Z} &\cong \mathbb{Z}/5\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}\\ &\cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/10\mathbb{Z}\\ &\cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/15\mathbb{Z}\\ &\cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z}.\end{aligned}$$

(iii) Sei K ein Körper und seien $a, b \in K$ mit $a \neq b$. Dann sind die Polynome $T - a$ und $T - b$ irreduzibel und nicht assoziiert, und somit teilerfremd. Nach dem chinesischen Restsatz gibt es einen Isomorphismus

$$K[T]/I \xrightarrow{\sim} K[T]/(T - a) \times K[T]/(T - b),$$

wobei $I = (T^2 - (a + b)T + ab)$.

Außerdem ist der Einsetzungshomorphismus $\varepsilon_a\colon K[T] \to K$, $p \mapsto p(a)$, surjektiv mit Kern $(T - a)$ (nach Proposition 6.3.15). Nach dem Homomorphiesatz induziert er einen Isomorphismus $\bar{\varepsilon}_a\colon K[T]/(T - a) \xrightarrow{\sim} K$. Damit erhalten wir einen Isomorphismus

$$K[T]/I \xrightarrow{\sim} K \times K, \quad p + I \mapsto (p(a), p(b)).$$

Beispiel 8.2.52 (simultane Kongruenzen)**.** Der Beweis des chinesischen Restsatzes und der euklidische Algorithmus erlauben die explizite Lösung von Kongruenzsystemen mit teilerfremden Moduln in euklidischen Ringen.

Zum Beispiel: Gesucht ist die Menge X aller $n \in \mathbb{Z}$, so dass

$$\begin{aligned} n &\equiv 1 \mod 2,\\ n &\equiv 2 \mod 3,\\ n &\equiv 0 \mod 5.\end{aligned}$$

Man beachte dabei, dass die Zahlen 2, 3 und 5 paarweise teilerfremd sind. Wir betrachten die Abbildung

$$\begin{aligned}\varphi\colon \mathbb{Z} &\to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/5\mathbb{Z},\\ n &\mapsto (n+2\mathbb{Z}, n+3\mathbb{Z}, n+5\mathbb{Z}).\end{aligned}$$

Die gesuchte Lösungsmenge X ist dann das Urbild von $\{([1],[2],[0])\}$ unter φ. Nach Korollar 8.2.50 ist X nicht leer, und zwar der Gestalt

$$X = n_0 + 30\mathbb{Z},$$

wobei n_0 eine beliebige Lösung ist. Um ein solches n_0 zu finden, verfahren wir wie im Beweis von Satz 8.2.49. Wir suchen nämlich $r_i, s_i \in \mathbb{Z}$, $i \in \{1,2,3\}$, mit folgenden Eigenschaften:

$$\begin{aligned} r_1 &\in 2\mathbb{Z}, \quad s_1 \in 15\mathbb{Z}, \quad r_1 + s_1 = 1,\\ r_2 &\in 3\mathbb{Z}, \quad s_2 \in 10\mathbb{Z}, \quad r_2 + s_2 = 1,\\ r_3 &\in 5\mathbb{Z}, \quad s_3 \in 6\mathbb{Z}, \quad r_3 + s_3 = 1,\end{aligned}$$

so dass $\varphi(s_i) = e_i$. Dazu verwenden wir den euklidischen Algorithmus, genauer die Bemerkung 8.2.45: Da $\operatorname{ggT}(2,15) = 1$, können wir $u, v \in \mathbb{Z}$ finden mit $2u + 15v = 1$:

$$15 = 7 \cdot 2 + 1 \quad \Longrightarrow \quad 1 = 2 \cdot (-7) + 15 \cdot 1.$$

Damit können wir $r_1 = -14$ und $s_1 = 15$ auswählen. Auf ähnliche Weise finden wir $r_2 = -9$, $s_2 = 10$, $r_3 = -5$ und $s_3 = 6$. Schließlich erhalten wir

$$([1],[2],[0]) = 1 \cdot e_1 + 2 \cdot e_2 + 0 \cdot e_3 = \varphi(1 \cdot s_1 + 2 \cdot s_2 + 0 \cdot s_3) = \varphi(35),$$

so dass $X = 35 + 30\mathbb{Z} = 5 + 30\mathbb{Z}$.

8.3 Endlich erzeugte Moduln über Hauptidealringen

8.3.1 Präsentationen von Moduln

Nicht jeder Modul M über einem Ring R ist frei, aber er lässt sich als Kokern einer linearen Abbildung zwischen freien Moduln darstellen. Man wählt nämlich eine erzeugende Familie $F = (x_i)_{i\in I}$ in M aus, um eine surjektive lineare Abbildung

$$\varphi_F\colon R^{(I)} \twoheadrightarrow M$$

zu erhalten. Der Kern $\ker\varphi_F$ heißt den *Relationenmodul* von F: Er besteht genau aus den Familien $(r_i)_{i\in I} \in R^{(I)}$, so dass die entsprechende Linearkombination der Familie F null ist. Wählt man nun eine Familie $G = (y_j)_{j\in J}$ in $R^{(I)}$ aus, die $\ker\varphi_F$ erzeugt, so erhalten wir eine lineare Abbildung

$$\varphi_G\colon R^{(J)} \to R^{(I)}$$

mit $\operatorname{im}\varphi_G = \ker\varphi_F$. Nach dem Homomorphiesatz induziert dann φ_F einen Isomorphismus

$$\bar{\varphi}_F\colon \operatorname{coker}\varphi_G = R^{(I)}/\operatorname{im}\varphi_G \xrightarrow{\sim} M.$$

Eine solche Darstellung von M heißt *Präsentation*:

Definition 8.3.1 (Präsentation eines Moduls, endlich präsentierbar)**.** Sei R ein Ring und M ein R-Modul.

- Eine *Präsentation* von M besteht aus zwei Mengen I, J und zwei lineare Abbildungen

  $$R^{(J)} \xrightarrow{g} R^{(I)} \xrightarrow{f} M,$$

  so dass f surjektiv ist und $\operatorname{im} g = \ker f$.

- M heißt *endlich präsentierbar*, wenn es eine Präsentation besitzt, in der die Mengen I und J endlich sind.

Bemerkung 8.3.2. Nach Definition ist ein R-Modul genau dann endlich erzeugt, wenn er eine Präsentation wie oben besitzt, in der die Menge I endlich ist. Insbesondere sind endlich präsentierbare Moduln endlich erzeugt. Im Allgemeinen gilt die Umkehrung aber nicht.

Beispiel 8.3.3.

(i) Die Folgen

$$\begin{aligned} &\mathbb{Z} \to \mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z} && &&\mathbb{Z}^2 \to \mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z} \\ &x \mapsto 2x && &&x \mapsto 4x_1 + 6x_2 \end{aligned}$$

sind Präsentationen des $\mathbb{Z}$-Moduls $\mathbb{Z}/2\mathbb{Z}$.

(ii) Seien $n_1, \ldots, n_k \in \mathbb{Z}$. Dann ist die Folge

$$\mathbb{Z}^k \xrightarrow{L_A} \mathbb{Z}^k \twoheadrightarrow \mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_k\mathbb{Z} \quad \text{mit} \quad A = \operatorname{diag}(n_1, \ldots, n_k) \in M_k(\mathbb{Z})$$

eine Präsentation des $\mathbb{Z}$-Moduls $\mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_k\mathbb{Z}$.

(iii) Die folgende Folge ist eine Präsentation des $\mathbb{Z}$-Moduls $\mathbb{Z}$:

$$\mathbb{Z} \xrightarrow{L_A} \mathbb{Z}^2 \xrightarrow{L_B} \mathbb{Z} \quad \text{mit} \quad A = \begin{pmatrix} -3 \\ 2 \end{pmatrix} \text{ und } B = \begin{pmatrix} 2 & 3 \end{pmatrix}.$$

Die folgende Definition spielt eine wichtige Rolle in der Kommutativen Algebra und in der Algebraischen Geometrie:

Definition 8.3.4 (noetherscher Ring)**.** Ein Ring R heißt *noethersch*, wenn jedes Ideal $I \subset R$ ein endlich erzeugter R-Modul ist.

Bemerkung 8.3.5.

(i) Jeder Hauptidealring ist nach Definition noethersch.

(ii) Ist $f\colon R \twoheadrightarrow S$ ein surjektiver Ringhomomorphismus und ist R noethersch, so ist auch S noethersch. Denn jedes Ideal $I \subset S$ ist das Bild des Ideals $f^{-1}(I) \subset R$ unter f, und ist E ein endliches Erzeugendensystem von $f^{-1}(I)$, so ist $f(E)$ ein endliches Erzeugendensystem von I.

(iii) Der *Hilbertsche Basissatz* sagt, dass der Polynomring $R[T]$ über einem noetherschen Ring R wieder noethersch ist. Beispielsweise ist der Polynomring in n Variablen über einem Hauptidealring noethersch.

Proposition 8.3.6. *Sei R ein noetherscher Ring und M ein endlich erzeugter R-Modul. Dann ist jeder Untermodul $N \subset M$ endlich erzeugt.*

Beweis. Sei $\{x_1, \ldots, x_n\}$ ein Erzeugendensystem von M. Wir beweisen die Aussage durch Induktion über n. Falls $n = 0$ ist $M = \{0\}$ und damit auch $N = \{0\}$. Falls $n \geq 1$ betrachten wir $M' = \operatorname{Span}_R\{x_1, \ldots, x_{n-1}\} \subset M$ und die Quotientenabbildung $\pi\colon M \to M/M'$:

$$\begin{array}{ccccc} M' & \hookrightarrow & M & \xrightarrow{\pi} & M/M' \\ \uparrow & & \uparrow & & \uparrow \\ N \cap M' & \hookrightarrow & N & \xrightarrow{\pi|_N} & \pi(N). \end{array}$$

Der Modul M/M' wird von $x_n + M'$ erzeugt, und es gibt insbesondere eine surjektive R-lineare Abbildung $f\colon R \to M/M'$. Jeder Untermodul U von M/M' ist das Bild des Ideals $f^{-1}(U)$ unter f. Da R noethersch ist, ist $f^{-1}(U)$ und damit auch U endlich erzeugt. Insbesondere ist $\pi(N)$ endlich erzeugt, d.h., es gibt eine endliche Teilmenge $E \subset N$ mit $\pi(N) = \mathrm{Span}_R(\pi(E))$. Auf der anderen Seite ist $N \cap M'$ endlich erzeugt nach Induktionsvoraussetzung, d.h., es gibt eine endliche Teilmenge $F \subset N$ mit $N \cap M' = \mathrm{Span}_R(F)$. Dann ist die Menge $E \cup F$ endlich und es gilt $N = \mathrm{Span}_R(E \cup F)$, denn: Sei $x \in N$. Da $\pi(\mathrm{Span}_R(E)) = \mathrm{Span}_R(\pi(E)) = \pi(N)$ gibt es ein $y \in \mathrm{Span}_R(E)$ mit $\pi(x) = \pi(y)$. Dann gilt $x - y \in N \cap \ker\pi = N \cap M' = \mathrm{Span}_R(F)$, und damit $x = y + (x-y) \in \mathrm{Span}_R(E) + \mathrm{Span}_R(F) = \mathrm{Span}_R(E \cup F)$. □

Korollar 8.3.7. *Sei R ein noetherscher Ring. Dann ist jeder endlich erzeugte R-Modul endlich präsentierbar.*

Beweis. Sei M ein endlich erzeugter R-Modul. Es gibt dann ein $m \in \mathbb{N}$ und eine surjektive lineare Abbildung $f\colon R^m \twoheadrightarrow M$. Nach Proposition 8.3.6 ist $\ker f$ wieder endlich erzeugt. Es gibt damit ein $n \in \mathbb{N}$ und eine lineare Abbildung $g\colon R^n \to R^m$ mit $\mathrm{im}\, g = \ker f$. Die Abbildungen f und g bilden eine Präsentation des Moduls M, der somit endlich präsentierbar ist. □

8.3.2 Torsion und Länge

Ein neues Phänomen bei Ringen im Vergleich zu Körpern ist die Existenz von Nullteilern: Ein Produkt $r \cdot s$ kann null sein, selbst wenn r und s nicht null sind. Es gibt aber ein *weiteres* neues Phänomen bei Moduln im Vergleich zu Vektorräumen: Ist M ein R-Modul und ist $x \in M \setminus \{0\}$, so kann $r \cdot x$ null sein, selbst wenn $r \in R$ kein Nullteiler ist. Zum Beispiel gilt $2 \cdot [1] = 0$ im $\mathbb{Z}$-Modul $\mathbb{Z}/2\mathbb{Z}$, und $2 \in \mathbb{Z}$ ist kein Nullteiler. Dieses Phänomen heißt *Torsion*.

Definition 8.3.8 (Torsionselement, Torsionsuntermodul, Torsionsmodul, torsionsfrei)**.** Sei R ein kommutativer Ring und M ein R-Modul.

- Ein Element $x \in M$ heißt *Torsionselement*, wenn $r \cdot x = 0$ für ein Element $r \in R$, das kein Nullteiler ist.
- Die Teilmenge aller Torsionselemente von M heißt der *Torsionsuntermodul* von M und wird mit $T(M)$ bezeichnet.
- M heißt *Torsionsmodul*, wenn $T(M) = M$, und *torsionsfrei*, wenn $T(M) = \{0\}$.

Proposition 8.3.9 (Eigenschaften der Torsion)**.** *Sei R ein kommutativer Ring und M ein R-Modul.*

(i) *$T(M)$ ist ein Untermodul von M.*

(ii) *$T(M)$ ist ein Torsionsmodul.*

(iii) *$M/T(M)$ ist torsionsfrei.*

(iv) *Ist $(M_i)_{i\in I}$ eine Familie von R-Moduln, so gilt*

$$T\left(\bigoplus_{i\in I} M_i\right) = \bigoplus_{i\in I} T(M_i).$$

Beweis. Zu (i). Wir verwenden das Kriterium 3.2.8. Das Nullelement $0 \in M$ ist ein Torsionselement, denn $1 \cdot 0 = 0$ und $1 \in R$ ist kein Nullteiler (siehe Bemerkung 8.1.17(i)). Seien $x, y \in T(M)$ und sei $r \in R$. Es gilt $sx = 0$ und $ty = 0$ für Elemente $s, t \in R$, die keine Nullteiler sind. Dann ist $s(rx) = r(sx) = r0 = 0$, so dass $rx \in T(M)$. Da s und t keine Nullteiler

sind, ist auch das Produkt st kein Nullteiler. Zudem gilt $st(x+y) = t(sx)+s(ty) = t0+s0 = 0$, so dass $x + y \in T(M)$.

Zu (ii). Nach Definition sind alle Elemente von $T(M)$ Torsionselemente.

Zu (iii). Sei $x + T(M)$ ein Torsionselement von $M/T(M)$. Es gibt dann einen Nicht-Nullteiler $r \in R$ mit $r(x + T(M)) = 0$, d.h., $rx \in T(M)$. Es gibt dann wiederum einen Nicht-Nullteiler $s \in R$, so dass $srx = 0$. Da das Produkt sr kein Nullteiler ist, gilt $x \in T(M)$, d.h., $x + T(M) = 0$.

Zu (iv). Sei $x = (x_i)_{i\in I} \in \bigoplus_{i\in I} M_i$. Zu zeigen ist, dass x genau dann ein Torsionselement ist, wenn alle x_i Torsionselemente sind. Ist x ein Torsionselement, so gibt es einen Nicht-Nullteiler $r \in R$ mit $rx_i = 0$ für alle $i \in I$. Insbesondere sind alle Elemente x_i Torsionselemente. Seien umgekehrt alle x_i Torsionselemente. Nach Definition der direkten Summe gibt es eine endliche Teilmenge $J \subset I$, so dass $x_i = 0$ für alle $i \in I \setminus J$. Fur jedes $j \in J$ wählt man einen Nicht-Nullteiler $r_j \in R$ aus, so dass $r_j x_j = 0$. Dann ist $r = \prod_{j\in J} r_j$ kein Nullteiler, und es gilt $rx = 0$. □

Beispiel 8.3.10. Sei R ein kommutativer Ring.

(i) Jeder freie R-Modul M ist torsionsfrei. Ohne Einschränkung ist $M = R^{(I)}$. Ist $r \in R$ kein Nullteiler und ist $r \cdot (x_i)_{i\in I} = 0$ in $R^{(I)}$, so muss jedes $x_i \in R$ null sein (nach Definition von Nullteiler). Das gleiche Argument zeigt, dass der R-Modul R^I torsionsfrei ist (aber nicht unbedingt frei, siehe Beispiel 8.1.33(v)).

(ii) Jeder Untermodul eines torsionsfreien R-Moduls ist wieder torsionsfrei. Insbesondere ist jedes Ideal $I \subset R$ ein torsionsfreier R-Modul.

(iii) Sei $I \subset R$ ein Ideal, das einen Nicht-Nullteiler enthält. Dann ist der zyklische R-Modul R/I ein Torsionsmodul. Denn für alle $r \in I$ und $x \in R$ gilt $r(x + I) = 0$.

(iv) Ist K ein Körper und ist $R \subset K$ ein Unterring, so ist K ein torsionsfreier R-Modul. Zum Beispiel ist $\mathbb{Q}$ ein torsionsfreier $\mathbb{Z}$-Modul, der aber kein freier $\mathbb{Z}$-Modul ist (siehe Beispiel 8.1.33(iv)).

Notation 8.3.11. Sei R ein kommutativer Ring, M ein R-Modul und $r \in R$. Dann ist

$$T_r(M) := \{x \in M \mid \text{es gibt ein } n \in \mathbb{N} \text{ mit } r^n x = 0\}$$

ein Untermodul von M. Man schreibt auch $M[r^\infty]$ oder $M(r)$ für diesen Untermodul und bezeichnet ihn als den *r-Torsionsuntermodul* von M. Falls r kein Nullteiler ist, gilt nach Definition $T_r(M) \subset T(M)$.

Notation 8.3.12. Sei R ein kommutativer Ring, M ein R-Modul und $r \in R$. Dann ist

$$rM := \{rx \mid x \in M\}$$

ein Untermodul von M.

In folgender Aussage wird die universelle Eigenschaft der direkten Summe verwendet, siehe Proposition 6.1.3(ii).

Proposition 8.3.13 (Zerlegung der Torsion über Hauptidealringen)**.** *Sei R ein Hauptidealring, P ein Repräsentantensystem der Primelemente von R bis auf Assoziiertheit und M ein R-Modul. Die Inklusionsabbildungen $T_p(M) \hookrightarrow T(M)$ induzieren einen Isomorphismus*

$$\bigoplus_{p\in P} T_p(M) \xrightarrow{\sim} T(M).$$

Ist M endlich erzeugt, so gibt es nur endlich viele $p \in P$ mit $T_p(M) \neq \{0\}$, und es gibt dann ein $n \in \mathbb{N}$ mit $p^n T_p(M) = \{0\}$.

Beweis. Sei f die gegebene Abbildung. Wir beweisen, dass f surjektiv und injektiv ist.

Zur Surjektivität. Sei $x \in T(M)$, so dass $rx = 0$ mit einem $r \in R \setminus \{0\}$. Man schreibt $r = up_1^{e_1} \dots p_n^{e_n}$ mit $u \in R^\times$ und paarweise verschiedenen Elementen $p_1, \dots, p_n \in P$. Wir zeigen durch Induktion über n, dass x im Bild von f liegt. Falls $n = 0$ gilt sogar $x = 0$. Sei also $n \geq 1$, und seien $y = p_1^{e_1} x$ und $z = p_2^{e_2} \dots p_n^{e_n} x$. Da $p_2^{e_2} \dots p_n^{e_n} y = 0$ gilt $y \in \operatorname{im} f$ nach der Induktionsvoraussetzung. Da $p_1^{e_1} z = 0$ gilt $z \in T_{p_1}(M)$ nach Definition. Da $p_1^{e_1}$ und $p_2^{e_2} \dots p_n^{e_n}$ teilerfremd sind und R ein Hauptidealring ist, gibt es $u, v \in R$ mit $up_1^{e_1} + vp_2^{e_2} \dots p_n^{e_n} = 1$. Dann gilt $x = uy + vz \in \operatorname{im} f$. Also ist f surjektiv

Zur Injektivität. Sei $Q \subset P$ eine endliche Teilmenge und zu jedem $p \in Q$ sei $x_p \in T_p(M)$, so dass $\sum_{p \in Q} x_p = 0$. Zu zeigen ist, dass alle x_p null sind. Zu jedem $p \in Q$ gibt es ein $n_p \in \mathbb{N}$ mit $p^{n_p} x_p = 0$. Sei $q \in Q$ beliebig und sei $r = \prod_{p \in Q \setminus \{q\}} p^{n_p}$. Aus der Gleichheit $x_q = -\sum_{p \in Q \setminus \{q\}} x_p$ folgt, dass $rx_q = 0$. Da r und q^{n_q} teilerfremd sind und R ein Hauptidealring ist, gibt es $u, v \in R$ mit $ur + vq^{n_q} = 1$. Dann gilt $x_q = urx_q + vq^{n_q} x_q = 0$, wie gewünscht.

Sei nun M endlich erzeugt. Nach Proposition 8.3.6 sind $T(M)$ und $T_p(M)$ auch endlich erzeugt. Eine direkte Summe von unendlich vielen nicht-trivialen R-Moduln ist aber nicht endlich erzeugt, denn jedes Element und somit jede endliche Teilmenge der Summe muss in einer endlichen Teilsumme enthalten sein. Damit gibt es nur endlich viele p mit $T_p(M) \neq \{0\}$. Sei $\{x_1, \dots, x_r\}$ ein Erzeugendensystem von $T_p(M)$. Zu jedem i gibt es ein $n_i \in \mathbb{N}$ mit $p^{n_i} x_i = 0$. Ist $n \geq \max\{n_1, \dots, n_r\}$, so gilt $p^n x_i = 0$ für alle $i \in \{1, \dots, r\}$, und damit $p^n x = 0$ für alle $x \in \operatorname{Span}_R\{x_1, \dots, x_r\} = T_p(M)$. □

Beispiel 8.3.14. Sei $R = \mathbb{Z}$ und $M = \mathbb{Z}/30\mathbb{Z}$. Dann gilt $T(M) = M$,

$$T_2(M) = 15\mathbb{Z}/30\mathbb{Z}, \quad T_3(M) = 10\mathbb{Z}/30\mathbb{Z}, \quad T_5(M) = 6\mathbb{Z}/30\mathbb{Z}$$

und $T_p(M) = \{0\}$ für alle anderen Primzahlen p. Damit erhalten wir die Zerlegung

$$\mathbb{Z}/30\mathbb{Z} = 15\mathbb{Z}/30\mathbb{Z} \oplus 10\mathbb{Z}/30\mathbb{Z} \oplus 6\mathbb{Z}/30\mathbb{Z}.$$

Nach dem Homomorphiesatz gibt es zudem Isomorphismen $\mathbb{Z}/2\mathbb{Z} \xrightarrow{\sim} 15\mathbb{Z}/30\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z} \xrightarrow{\sim} 10\mathbb{Z}/30\mathbb{Z}$ und $\mathbb{Z}/5\mathbb{Z} \xrightarrow{\sim} 6\mathbb{Z}/30\mathbb{Z}$. Der Isomorphismus aus Proposition 8.3.13 ist dann die Umkehrabbildung zum Isomorphismus aus dem chinesischen Restsatz (Korollar 8.2.50).

Der Begriff der Dimension existiert bei beliebigen Ringen nicht. Es gibt aber mehrere teilweise Erweiterungen dieses Begriffs, die in verschiedenen Situationen nützlich sind. Zum Beispiel kann man den *Rang* eines Moduls über einem Integritätsring R definieren, so dass der Rang von R^n gleich n ist (das werden wir später im Fall eines Hauptidealringes besprechen). Die *Länge* eines Moduls ist eine andere Verallgemeinerung der Dimension auf beliebige Ringe:

Definition 8.3.15 (Länge eines Moduls)**.** Sei R ein Ring und M ein R-Modul. Die *Länge* von M ist

$$\ell_R(M) = \sup \left\{ n \in \mathbb{N} \;\middle|\; \begin{array}{c} \text{es gibt eine Kette von Untermoduln} \\ \{0\} = M_0 \subsetneqq M_1 \subsetneqq \dots \subsetneqq M_n = M \end{array} \right\} \in \mathbb{N} \cup \{\infty\}.$$

Beispiel 8.3.16.

(i) Nach Definition gilt $\ell_R(M) = 0$ genau dann, wenn $M = \{0\}$.

(ii) Sei K ein Körper und V ein K-Vektorraum. Aus Proposition 3.3.35 folgt leicht, dass $\ell_K(V) = \dim_K(V)$ gilt.

(iii) Es gilt $\ell_{\mathbb{Z}}(\mathbb{Z}) = \infty$, denn für jedes $n \in \mathbb{N}$ gibt es zum Beispiel die Kette

$$\{0\} \subsetneqq 2^{n-1}\mathbb{Z} \subsetneqq \dots \subsetneqq 2\mathbb{Z} \subsetneqq \mathbb{Z}.$$

Ist allgemeiner R ein Integritätsring, der kein Körper ist, und ist $n \in \mathbb{N} \setminus \{0\}$, so gilt $\ell_R(R^n) = \infty$.

Proposition 8.3.17 (Eigenschaften der Länge)**.** *Sei R ein Ring.*

(i) *Sind M und N isomorphe R-Moduln, so gilt $\ell_R(M) = \ell_R(N)$.*

(ii) *Sei M ein R-Modul und $M' \subset M$ ein Untermodul. Dann gilt*

$$\ell_R(M) = \ell_R(M') + \ell_R(M/M').$$

(iii) *Seien M und N zwei R-Moduln. Dann gilt*

$$\ell_R(M \oplus N) = \ell_R(M) + \ell_R(N).$$

(iv) *Ist R ein Hauptidealring und ist $r \in R \setminus \{0\}$, so gilt*

$$\ell_R(R/(r)) = \textit{Anzahl der Primfaktoren von } r.$$

Beweis. Zu (i). Sei $f\colon M \xrightarrow{\sim} N$ ein Isomorphismus. Das Bild bzw. das Urbild unter f einer Kette von Untermoduln ist eine Kette derselben Länge, so dass $\ell_R(M) = \ell_R(N)$.

Zu (ii). Sei $\pi\colon M \to M/M'$ die Quotientenabbildung. Seien $\{0\} = M'_0 \subsetneqq \cdots \subsetneqq M'_m = M'$ und $\{0\} = N_0 \subsetneqq \cdots \subsetneqq N_n = M/M'$ Ketten von Untermoduln. Die Kette

$$\{0\} = M'_0 \subsetneqq \cdots \subsetneqq M'_m = M' = \pi^{-1}(N_0) \subsetneqq \cdots \subsetneqq \pi^{-1}(N_n) = M$$

zeigt, dass $\ell_R(M) \geq m + n$ und damit $\ell_R(M) \geq \ell_R(M') + \ell_R(M/M')$.

Sei umgekehrt $\{0\} = M_0 \subsetneqq \cdots \subsetneqq M_n = M$ eine Kette von Untermoduln von M. Sie induziert zwei Ketten

$$M_0 \cap M' \subset \cdots \subset M_n \cap M' \quad \text{und} \quad \pi(M_0) \subset \cdots \subset \pi(M_n),$$

in denen die Inklusionen können Gleichheiten sein. Gilt aber $\pi(M_{i-1}) = \pi(M_i)$, so muss $M_{i-1} \cap M' \neq M_i \cap M'$ gelten, denn: Sei $x \in M_i \setminus M_{i-1}$. Aus $\pi(M_{i-1}) = \pi(M_i)$ folgt, dass ein $y \in M_{i-1}$ mit $x + M' = y + M'$ existiert. Dann liegt $x - y$ in $M_i \cap M'$, aber nicht in $M_{i-1} \cap M'$, da $x = (x - y) + y \notin M_{i-1}$. Dies zeigt, dass die gesamte Länge der zwei Ketten gleich n ist, und damit dass $\ell_R(M) \leq \ell_R(M') + \ell_R(M/M')$.

Zu (iii). Es gibt Isomorphismen $M \cong M \oplus \{0\} \subset M \oplus N$ und $N \cong (M \oplus N)/(M \oplus \{0\})$. Die Aussage folgt damit aus (i) und (ii).

Zu (iv). Sei $r = up_1^{e_1} \dots p_n^{e_n}$ eine Primfaktorzerlegung von r, wobei $u \in R^\times$ und $p_1, \dots, p_n$ paarweise nicht-assoziierte Primelemente sind. Nach dem chinesischen Restsatz für Hauptidealringe (Korollar 8.2.50) gibt es einen Isomorphismus

$$R/(r) \cong R/(p_1^{e_1}) \oplus \cdots \oplus R/(p_n^{e_n}).$$

Wegen (iii) können wir damit annehmen, dass $r = p^e$ mit einem Primelement $p \in R$ und einem $e \in \mathbb{N}$. Die Kette

$$\{0\} = (p^e)/(p^e) \subsetneqq (p^{e-1})/(p^e) \subsetneqq \cdots \subsetneqq (p^0)/(p^e) = R/(p^e)$$

zeigt, dass $\ell_R(R/(p^e)) \geq e$. Das Urbild in R einer beliebigen Kette von Untermoduln von $R/(p^e)$ ist eine Kette von Idealen in R, die (p^e) enthalten. Ist I ein beliebiges Ideal in R mit $(p^e) \subset I$, so ist $I = (a)$ mit $a | p^e$. Nach der Eindeutigkeit der Primfaktorzerlegung ist a zu p^d assoziiert mit einem $d \in \{0, \dots, e\}$. Insbesondere gibt es höchstens $e + 1$ solche Ideale, was $\ell_R(R/(p^e)) \leq e$ zeigt. □

Wir besprechen nun ein paar Anwendungen der Länge auf Hauptidealringe.

Proposition 8.3.18 (Rang freier Moduln über Hauptidealringen)**.** *Sei R ein Hauptidealring und seien $n, m \in \mathbb{N}$. Sind die R-Moduln R^n und R^m isomorph, so gilt $n = m$.*

Beweis. Falls R ein Körper ist, ist die Aussage schon bekannt (Satz 3.3.27). Sonst gibt es nach der Primfaktorzerlegung mindestens ein Primelement $p \in R$. Sei $f\colon R^n \xrightarrow{\sim} R^m$ ein Isomorphismus. Dann schränkt sich f zu einem Isomorphismus zwischen Untermoduln $p(R^n) \xrightarrow{\sim} p(R^m)$, und induziert damit einen Isomorphismus zwischen Quotientenmoduln $R^n/p(R^n) \xrightarrow{\sim} R^m/p(R^m)$. Insbesondere gilt $\ell_R(R^n/p(R^n)) = \ell_R(R^m/p(R^m))$ nach Proposition 8.3.17(i). Mit dem Homomorphiesatz 4.1.35 erhalten wir zudem einen Isomorphismus

$$\begin{aligned} R^n/p(R^n) &\xrightarrow{\sim} (R/(p))^n, \\ (r_1, \dots, r_n) + p(R^n) &\mapsto (r_1 + (p), \dots, r_n + (p)) \end{aligned}$$

(die komponentenweise Quotientabbildung $R^n \to (R/(p))^n$ ist surjektiv mit Kern $p(R^n)$). Aus Proposition 8.3.17(i,iii,iv) folgt nun

$$\ell_R(R^n/p(R^n)) = \ell_R((R/(p))^n) = n \cdot \ell_R(R/(p)) = n \cdot 1 = n.$$

Das Gleiche gilt aber mit m anstelle von n, so dass $n = m$. □

Bemerkung 8.3.19. Proposition 8.3.18 gilt eigentlich für einen beliebigen *kommutativen* Ring $R \neq \{0\}$. Das werden wir später mithilfe der äußeren Potenz beweisen (Korollar 10.2.11).

Die nächste Proposition ist eine Verallgemeinerung der Proposition 8.3.18 (man gewinnt sie zurück, wenn alle Ideale null sind).

Proposition 8.3.20 (Eindeutigkeit der Elementarteiler)**.** *Sei R ein Hauptidealring. Seien $n, m \in \mathbb{N}$ und seien*

$$(\alpha_1) \subset (\alpha_2) \subset \cdots \subset (\alpha_n) \subsetneq R \quad \textit{und} \quad (\beta_1) \subset (\beta_2) \subset \cdots \subset (\beta_m) \subsetneq R$$

Ketten von Idealen in R. Gibt es einen Isomorphismus von R-Moduln

$$\bigoplus_{i=1}^{n} R/(\alpha_i) \cong \bigoplus_{i=1}^{m} R/(\beta_i),$$

so gilt $n = m$ und $(\alpha_i) = (\beta_i)$ für alle $i \in \{1, \dots, n\}$.

Beweis. Seien $M = \bigoplus_{i=1}^{n} R/(\alpha_i)$ und $N = \bigoplus_{i=1}^{m} R/(\beta_i)$. Ist $\alpha \neq 0$, so ist $R/(\alpha)$ ein Torsionsmodul nach Beispiel 8.3.10(iii). Ist aber $\alpha = 0$, so ist $R/(\alpha) \cong R$ torsionsfrei nach Beispiel 8.3.10(i). Ist r die Anzahl der α_i, die gleich Null sind, so gilt nach Proposition 8.3.9(iv)

$$T(M) = \bigoplus_{i=1}^{r} \{0\} \oplus \bigoplus_{i=r+1}^{n} R/(\alpha_i), \quad \text{und damit} \quad M/T(M) \cong R^r.$$

Ein Isomorphismus $M \xrightarrow{\sim} N$ induziert Isomorphismen $T(M) \xrightarrow{\sim} T(N)$ und daher $M/T(M) \xrightarrow{\sim} N/T(N)$. Nach Proposition 8.3.18 muss deswegen r auch die Anzahl der β_i sein, die gleich Null sind. Wir können damit annehmen, dass α_1 und β_1 nicht null sind.

Wir zeigen zunächst, dass für alle $k \leq \min\{m, n\}$ gilt $(\alpha_k) = (\beta_k)$. Durch vollständige Induktion über k können wir annehmen, dass für alle $i < k$ gilt bereits $(\alpha_i) = (\beta_i)$. Ein Isomorphismus $M \xrightarrow{\sim} N$ induziert einen Isomorphismus $\alpha_k M \xrightarrow{\sim} \alpha_k N$. Für $i \in \{k, \dots, n\}$ gilt $\alpha_k \in (\alpha_i)$ und damit $\alpha_k(R/(\alpha_i)) = \{0\}$. Nach Induktionsvoraussetzung gilt außerdem $R/(\alpha_i) = R/(\beta_i)$ für alle $i \in \{1, \dots, k-1\}$. Damit erhalten wir Isomorphismen

$$\bigoplus_{i=1}^{k-1} \alpha_k(R/(\beta_i)) \cong \alpha_k M \cong \alpha_k N = \bigoplus_{i=1}^{m} \alpha_k(R/(\beta_i)).$$

Nach Proposition 8.3.17(iv) ist die Länge jedes Moduls $R/(\beta_i)$ endlich, und nach Proposition 8.3.17(ii) ist die Länge jedes Untermoduls $\alpha_k(R/(\beta_i))$ auch endlich. Berechnet man die Länge von beiden Seiten mit Proposition 8.3.17(iii), so erhalten wir

$$\ell_R\left(\bigoplus_{i=k}^{m} \alpha_k(R/(\beta_i))\right) = 0.$$

Insbesondere ist $\alpha_k(R/(\beta_k)) = \{0\}$, d.h., $(\alpha_k) \subset (\beta_k)$. Vertauschen wir die Rollen der α_i und der β_i, so erhalten wir ebenso $(\beta_k) \subset (\alpha_k)$, und somit $(\alpha_k) = (\beta_k)$, wie gewünscht.

Sei jetzt ohne Einschränkung $m \geq n$. Es gilt also $(\alpha_i) = (\beta_i)$ für alle $i \in \{1, \ldots, n\}$. Nach Vergleich der Längen von M und N mit Proposition 8.3.17(i,iii,iv), erhalten wir

$$\ell_R\left(\bigoplus_{i=n+1}^{m} R/(\beta_i)\right) = 0,$$

und damit $\bigoplus_{i=n+1}^{m} R/(\beta_i) = \{0\}$. Da $(\beta_i) \neq R$ ist, ist das nur möglich, wenn $n = m$. □

8.3.3 Der Elementarteilersatz

Nach Korollar 8.3.7 kann jeder endlich erzeugte Modul M über einem Hauptidealring R als Kokern einer R-linearen Abbildung

$$f\colon R^n \to R^m$$

dargestellt werden. Eine solche lineare Abbildung (und somit der Modul M bis auf Isomorphie) ist eindeutig durch eine Matrix $A \in M_{m\times n}(R)$ bestimmt, was ganz konkret ist. Das Klassifikationsproblem für endlich erzeugte R-Moduln lässt sich damit in ein Klassifikationsproblem für Matrizen über R übersetzen, das durch den folgenden Satz gelöst wird.

Satz 8.3.21 (Smith-Normalform über Hauptidealringen)**.** *Sei R ein Hauptidealring, seien M, N freie R-Moduln vom Rang $m, n \in \mathbb{N}$ und sei $f\colon N \to M$ eine R-lineare Abbildung. Dann existieren Basen B von N und C von M, so dass*

$$[f]_C^B = \begin{pmatrix} D & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix}, \quad D = \operatorname{diag}(d_1, \ldots, d_r), \quad d_1|d_2|\cdots|d_r \neq 0.$$

Außerdem sind r und das r-Tupel von Idealen $((d_1), \ldots, (d_r))$ eindeutig durch f bestimmt.

Eine $m \times n$-Matrix über R wie im Satz heißt in *Smith-Normalform*. Die Zahl $r \in \mathbb{N}$ heißt der *Rang* der linearen Abbildung f und die (bis auf Assoziiertheit bestimmten) Elemente $d_1, \ldots, d_r$ heißen die *Elementarteiler* von f. Ist R ein Körper, so haben wir diesen Satz bereits bewiesen (Satz 4.2.46). Dort wurde aber der Basisergänzungssatz verwendet, der über Hauptidealringen nicht gilt (z.B., die Familie (2) im $\mathbb{Z}$-Modul $\mathbb{Z}$ ist linear unabhängig, aber kann nicht zu einer Basis ergänzt werden).

Für den Beweis brauchen wir noch eine Vorbemerkung zu Elementarmatrizen. Über einem beliebigen Ring R können wir die $n\times n$-Elementarmatrizen V_{ij}, $M_i(\lambda)$ mit $\lambda \in R^\times$ und $A_{ij}(\alpha)$ mit $\alpha \in R$ genau wie in Definition 5.2.4 definieren. Diese Matrizen sind invertierbar, mit Inversen V_{ij}, $M_i(\lambda^{-1})$ und $A_{ij}(-\alpha)$. Multiplikation von links bzw. von rechts mit einer Elementarmatrix entspricht einer elementaren Zeilenumformung bzw. Spaltenumformung. Ein wichtiger Unterschied zu Körpern ist folgender: Im Allgemeinen kann man nicht eine beliebige Matrix über R durch elementare Zeilenumformungen auf Zeilenstufenform bringen; hierzu war die Existenz von multiplikativen Inversen in $R \setminus \{0\}$ notwendig.

Beweis von Satz 8.3.21. Zur Eindeutigkeit. Sei $A = [f]_C^B$. Die Basen B und C liefern ein kommutatives Quadrat

$$\begin{array}{ccc} R^n & \xrightarrow{L_A} & R^m \\ {\scriptstyle\varphi_B}\downarrow\wr & & \wr\downarrow{\scriptstyle\varphi_C} \\ N & \xrightarrow{f} & M, \end{array}$$

und daher einen Isomorphismus zwischen den Kokernen $R^m/\operatorname{im} L_A \xrightarrow{\sim} M/\operatorname{im} f$. Wegen der Gestalt der Matrix A gilt dann

$$M/\operatorname{im} f \cong R^{m-r} \oplus \bigoplus_{i=1}^{r} R/(d_i).$$

Sei k die Anzahl der d_i, die Einheiten sind (d.h., mit $(d_i) = R$). Nach Proposition 8.3.18 sind n und m eindeutig durch N und M bestimmt. Nach Proposition 8.3.20 sind $m-r$ und die Kette von Idealen $(d_n) \subset \cdots \subset (d_{k+1})$ eindeutig durch f bestimmt. Insbesondere sind r und k eindeutig durch f bestimmt, und damit alle Ideale (d_i).

Zur Existenz. Ohne Beschränkung der Allgemeinheit sei $M = R^m$ und $N = R^n$. Dann ist $f = L_A$ mit einer $m \times n$-Matrix $A = (a_{ij})_{i,j}$ über R, und die Existenz der Basen B und C ist äquivalent zur folgenden Aussage: Es gibt invertierbare Matrizen $S \in \mathrm{GL}_m(R)$ und $T \in \mathrm{GL}_n(R)$, so dass die Matrix $S \cdot A \cdot T$ in Smith-Normalform ist; man nimmt dann für B die Spalten von T und für C die Spalten von S^{-1}. Wir beweisen diese Aussage durch Induktion über m. Falls $A = 0$ (insbesondere falls $m = 0$), ist die Aussage trivial. Sonst können wir Zeilen bzw. Spalten vertauschen, so dass $a_{11} \neq 0$.

Wir werden die folgende Abbildung verwenden, die auf einem beliebigen faktoriellen Ring R wohldefiniert ist:

$$\delta\colon R \setminus \{0\} \to \mathbb{N}, \quad \delta(r) = \text{Anzahl der Primfaktoren von } r,$$

wobei die Primfaktoren mit Vielfachheit gezählt werden. Man beachte dabei, dass im Allgemeinen δ keine euklidische Gradfunktion ist. Falls R ein euklidischer Ring ist, können wir aber im Weiteren eine euklidische Gradfunktion anstelle von diesem δ verwenden. Wir verfahren jetzt durch vollständige Induktion über $\delta(a_{11})$, und wir betrachten zwei Fälle.

1. Fall: Es gibt einen Koeffizient a_{i1} oder a_{1j}, der nicht durch a_{11} teilbar ist. Durch Transponieren und Vertauschen von Zeilen können wir annehmen, dass a_{21} nicht durch a_{11} teilbar ist. Sei $d = \mathrm{ggT}(a_{11}, a_{21})$ und seien $r, s \in R$ die Elemente, so dass $rd = a_{11}$ und $sd = a_{21}$. Es gilt dann $\mathrm{ggT}(r, s) = 1$, und damit gibt es $u, v \in R$ mit $ur + vs = 1$. Die Matrix

$$U = \begin{pmatrix} u & v \\ -s & r \end{pmatrix}$$

ist invertierbar, da $\det(U) = 1$, und es gilt:

$$U \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = \begin{pmatrix} d \\ * \end{pmatrix}.$$

Nach der Wahl von d gilt außerdem $\delta(d) < \delta(a_{11})$. Multipliziert man A von links mit der invertierbaren $m \times m$-Matrix

$$\begin{pmatrix} U & 0 \\ 0 & I_{m-2} \end{pmatrix},$$

so reduziert man den δ-Wert des ersten Koeffizienten. Nach Induktionsvoraussetzung können wir dann die Matrix weiter auf Smith-Normalform bringen.

2. Fall: Alle Koeffizienten a_{i1} und a_{1j} sind durch a_{11} teilbar. Durch elementare Zeilen- und Spaltenumformungen können wir dann annehmen, dass die Matrix A die folgende Gestalt hat:

$$A = \begin{pmatrix} a_{11} & 0 \\ 0 & A' \end{pmatrix}.$$

Nach der Induktionsvoraussetzung bzgl. m können wir A' auf Smith-Normalform bringen. Gilt nun $a_{11} | a_{22}$ in der neuen Matrix, so ist die ganze Matrix in Smith-Normalform. Sonst addieren wir die zweite Spalte zur ersten Spalte und wenden wir den ersten Fall an. □

Bemerkung 8.3.22 (Smith-Normalform über euklidischen Ringen)**.** Wenn R ein euklidischer Ring ist, können wir im obigen Beweis δ durch eine euklidische Gradfunktion ersetzen. Im ersten Fall haben wir dann die folgende Vereinfachung: Es gibt $q, r \in R$ mit $a_{21} = qa_{11}+r$ und $\delta(r) < \delta(a_{11})$. Anstatt mit U zu multiplizieren, können wir mit der Matrix $V_{12}A_{21}(-q)$ multiplizieren, um a_{11} durch r zu ersetzen. Mit dieser Veränderung läuft der ganze Beweis mit nur elementaren Zeilen- und Spaltenumformungen, d.h., die invertierbare Matrizen S und T sind Produkte von Elementarmatrizen. Wenden wir jetzt den Satz auf eine invertierbare $n \times n$-Matrix A an, so schließen wir, dass A ein Produkt von Elementarmatrizen ist. Anders gesagt gilt das Korollar 5.2.11 über euklidischen Ringen. Man kann aber zeigen, dass es über beliebigen Hauptidealringen *nicht* gilt.

Beispiel 8.3.23. Der Beweis von Satz 8.3.21 ist völlig konstruktiv (sofern ein Algorithmus zur Bestimmung des ggT gegeben ist) und liefert einen Algorithmus, um eine Matrix auf Smith-Normalform zu bringen. Wir geben ein paar Beispiele dazu (siehe auch Beispiele 8.3.29 und 9.2.14):

(i) Sei

$$A = \begin{pmatrix} 6 & 4 \\ 7 & -3 \end{pmatrix} \in M_2(\mathbb{Z}).$$

Um A auf Smith-Normalform zu bringen, verwenden wir die euklidische Gradfunktion $\delta\colon \mathbb{Z} \setminus \{0\} \to \mathbb{N}$, $\delta(x) = |x|$, wie in der Bemerkung 8.3.22. Der erste Koeffizient a_{11} von A ist bereits nicht null. Da das Verfahren durch vollständige Induktion über $\delta(a_{11})$ läuft, wollen wir aber einen Koeffizient mit minimalem δ-Wert auswählen, nämlich -3. Man vertauscht also die zwei Zeilen und die zwei Spalten von A (d.h., man multipliziert A von beiden Seiten mit V_{12}):

$$\begin{pmatrix} 6 & 4 \\ 7 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} -3 & 7 \\ 4 & 6 \end{pmatrix}.$$

Nun ist 4 nicht durch -3 teilbar. Wir sind also im ersten Fall: Division mit Rest liefert $4 = (-1) \cdot (-3) + 1$. Man multipliziert von links mit $V_{12}A_{21}(1)$:

$$\begin{pmatrix} -3 & 7 \\ 4 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 13 \\ -3 & 7 \end{pmatrix}.$$

Jetzt ist a_{11} eine Einheit und wir sind im zweiten Fall. Man multipliziert von links mit $A_{21}(3)$ und von rechts mit $A_{12}(-13)$:

$$\begin{pmatrix} 1 & 13 \\ -3 & 7 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 13 \\ 0 & 46 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 46 \end{pmatrix}.$$

Die letzte Matrix ist nun in Smith-Normalform.

(ii) Sei K ein Körper und sei

$$A = \begin{pmatrix} T+1 & T^2 \\ T^2-1 & T \end{pmatrix} \in M_2(K[T]).$$

Um A auf Smith-Normalform zu bringen, verwenden wir die euklidische Gradfunktion $\deg\colon K[T] \setminus \{0\} \to \mathbb{N}$ wie in der Bemerkung 8.3.22. Der Grad des ersten Koeffizienten $a_{11} = T + 1$ ist bereits minimal. Es gilt hier $T + 1 | T^2 - 1$, aber $T + 1$ teilt nicht T^2. Divison mit Rest liefert $T^2 = (T - 1)(T + 1) + 1$. Also multiplizieren wir von rechts mit $A_{12}(1 - T)V_{12}$, um den Grad von a_{11} zu reduzieren:

$$\begin{pmatrix} T+1 & T^2 \\ T^2-1 & T \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & T+1 \\ -T^3+T^2+2T-1 & T^2-1 \end{pmatrix}.$$

Wir sind jetzt im zweiten Fall, und man multipliziert von links mit $A_{21}(T^3-T^2-2T+1)$ und von rechts mit $A_{12}(-T-1)$:

$$\begin{pmatrix} 1 & T+1 \\ -T^3+T^2+2T-1 & T^2-1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & T+1 \\ 0 & T^4-2T^2-T \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & T^4-2T^2-T \end{pmatrix}.$$

Die letzte Matrix ist nun in Smith-Normalform.

Beispiel 8.3.24 (lineare Gleichungen über Hauptidealringen)**.** Über einem Hauptidealring R können wir den Satz 8.3.21 anstelle von dem Gaußschen Eliminationsverfahren verwenden, um lineare Gleichungssysteme über R zu lösen. Gesucht seien zum Beispiel alle $x \in \mathbb{Z}^2$, so dass

$$\begin{aligned} 6x_1 + 4x_2 &= b_1, \\ 7x_1 - 3x_2 &= b_2, \end{aligned}$$

d.h., so dass $Ax = b$, wobei A die Matrix aus Beispiel 8.3.23(i) ist. Nach diesem Beispiel gilt

$$SAT = \begin{pmatrix} 1 & 0 \\ 0 & 46 \end{pmatrix} =: D,$$

wobei

$$S = A_{21}(3)V_{12}A_{21}(1)V_{12} = \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} \quad \text{und} \quad T = V_{12}A_{12}(-13) = \begin{pmatrix} 0 & 1 \\ 1 & -13 \end{pmatrix}.$$

Damit gilt:

$$\begin{aligned} \mathcal{L}(A,b) &= \{x \in \mathbb{Z}^2 \mid Ax = b\} \\ &= \{x \in \mathbb{Z}^2 \mid DT^{-1}x = Sb\} \\ &= \{Ty \in \mathbb{Z}^2 \mid Dy = Sb\} \\ &= T \cdot \{y \in \mathbb{Z}^2 \mid y_1 = b_1 + b_2 \text{ und } 46y_2 = 3b_1 + 4b_2\}. \end{aligned}$$

Falls $3b_1 + 4b_2$ nicht durch 46 teilbar ist, gibt es dann keine Lösungen. Sonst gibt es genau eine Lösung. Sind zum Beispiel $b_1 = 14$ und $b_2 = 1$, so erhalten wir $y_1 = 15$ und $y_2 = 1$, und die einzige Lösung ist

$$x = Ty = \begin{pmatrix} 0 & 1 \\ 1 & -13 \end{pmatrix} \begin{pmatrix} 15 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Satz 8.3.25 (Elementarteilersatz)**.** *Sei R ein Hauptidealring und F ein freier R-Modul vom Rang $n \in \mathbb{N}$. Zu jedem Untermodul $M \subset F$ gibt es eine Basis $(x_1, \ldots, x_n)$ von F, ein $r \in \{0, \ldots, n\}$ und Elemente $d_1, \ldots, d_r \in R \setminus \{0\}$ mit folgenden Eigenschaften:*

(i) *Die Familie $(d_1x_1, \ldots, d_rx_r)$ ist eine Basis von M.*

(ii) *Es gilt $d_1 | d_2 | \cdots | d_r$.*

Außerdem sind r und das r-Tupel von Idealen $((d_1), \ldots, (d_r))$ eindeutig durch M bestimmt.

Die (bis auf Assoziiertheit bestimmten) Elemente $d_1, \ldots, d_r$ heißen die *Elementarteiler* des Untermoduls $M \subset F$.

Beweis. Zur Eindeutigkeit. Wie im Beweis der Eindeutigkeitsaussage vom Satz 8.3.21 gibt es einen Isomorphismus

$$F/M \cong R^{n-r} \oplus \bigoplus_{i=1}^{r} R/(d_i).$$

Die Eindeutigkeitsaussage folgt nun aus Proposition 8.3.20.

Zur Existenz. Ohne Einschränkung sei $F = R^n$. Nach Proposition 8.3.6 ist M endlich erzeugt. Es gibt also ein $m \in \mathbb{N}$ und eine R-lineare Abbildung

$$f\colon R^m \to R^n \quad \text{mit} \quad \operatorname{im} f = M.$$

Nach Satz 8.3.21 gibt es Basen $C = (y_1, \dots, y_m)$ von R^m und $B = (x_1, \dots, x_n)$ von R^n, so dass

$$[f]_B^C = \begin{pmatrix} D & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{pmatrix}, \quad D = \operatorname{diag}(d_1, \dots, d_r), \quad d_1 | d_2 | \cdots | d_r \neq 0.$$

Es bleibt zu zeigen, dass die Familie $(d_1 x_1, \dots, d_r x_r)$ eine Basis von M ist. Nach Definition der Darstellungsmatrix gilt $f(y_i) = d_i x_i$ für alle $i \leq r$ und $f(y_i) = 0$ für $i \geq r$. Da $\operatorname{im} f = M$ ist die Familie erzeugend. Sie ist auch linear unabhängig, da R ein Integritätsring ist und B linear unabhängig ist: Ist $\sum_{i=1}^r r_i d_i x_i = 0$ mit $r_i \in R$, so folgt $r_i d_i = 0$ und damit $r_i = 0$ für alle i. □

Korollar 8.3.26. *Sei R ein Hauptidealring und F ein freier R-Modul vom Rang $n \in \mathbb{N}$. Dann ist jeder Untermodul von F frei vom Rang $\leq n$.*

Beweis. Dies folgt unmittelbar aus Satz 8.3.25. □

Bemerkung 8.3.27. Das Korollar 8.3.26 gilt nicht für allgemeinere Ringe. Zum Beispiel ist $(2, T)$ ein Untermodul des $\mathbb{Z}[T]$-Moduls $\mathbb{Z}[T]$, der nicht frei ist.

Bemerkung 8.3.28. Der Zusammenhang zwischen den Elementarteilern aus Satz 8.3.21 und den aus Satz 8.3.25 ist folgender. Ist $f\colon N \to M$ eine R-lineare Abbildung zwischen freien R-Moduln wie im Satz 8.3.21, dann sind die Elementarteiler von f genau die Elementarteiler des Untermoduls $\operatorname{im} f \subset M$. Seien umgekehrt F ein freier R-Modul und $M \subset F$ ein Untermodul wie im Satz 8.3.25. Dann sind die Elementarteiler von $M \subset F$ genau die der Inklusionsabbildung $M \hookrightarrow F$.

Beispiel 8.3.29. Sei $M = 2\mathbb{Z} \oplus 3\mathbb{Z} \subset \mathbb{Z}^2$. Wir suchen eine Basis (x_1, x_2) von $\mathbb{Z}^2$ wie im Elementarteilersatz. Der Untermodul M ist das Bild von $L_A\colon \mathbb{Z}^2 \to \mathbb{Z}^2$, wobei $A = \operatorname{diag}(2, 3)$. Als erster Schritt müssen wir A auf Smith-Normalform bringen. Da $2|0$ sind wir im zweiten Fall, aber 3 ist nicht durch 2 teilbar. Deswegen addieren wir die zweite Spalte zur ersten Spalte:

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \xrightarrow{\cdot A_{21}(1)} \begin{pmatrix} 2 & 0 \\ 3 & 3 \end{pmatrix},$$

und wir sind jetzt im ersten Fall. Division mit Rest liefert $3 = 1 \cdot 2 + 1$. Man multipliziert von links mit $V_{12}A_{21}(-1)$, um $|a_{11}|$ zu reduzieren:

$$\begin{pmatrix} 2 & 0 \\ 3 & 3 \end{pmatrix} \xrightarrow{V_{12}A_{21}(-1)\cdot} \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix}.$$

Da 1 eine Einheit ist, sind wir wieder im zweiten Fall:

$$\begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} \xrightarrow{A_{21}(-2)\cdot} \begin{pmatrix} 1 & 3 \\ 0 & -6 \end{pmatrix} \xrightarrow{\cdot A_{12}(-3)M_2(-1)} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

Es gilt also $SAT = \operatorname{diag}(1, 6)$ mit $S = A_{21}(-2)V_{12}A_{21}(-1)$. Die Basis (x_1, x_2) im Beweis von Satz 8.3.25 besteht aus den Spalten von S^{-1}:

$$S^{-1} = A_{21}(1)V_{12}A_{21}(2) = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} \quad \Longrightarrow \quad x_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Damit ist (x_1, x_2) eine Basis von $\mathbb{Z}^2$, so dass $(x_1, 6x_2)$ eine Basis von M ist.

8.3.4 Strukturs\"atze

Satz 8.3.30 (Hauptsatz über endlich erzeugte Moduln über Hauptidealringen)**.** *Sei R ein Hauptidealring und sei M ein endlich erzeugter R-Modul.*

(i) (Erster Struktursatz: Elementarteiler) *Es gibt $n, r \in \mathbb{N}$ und Elemente $d_1, \ldots, d_r \in R \setminus (\{0\} \cup R^\times)$ mit $d_1 | d_2 | \cdots | d_r$, so dass*

$$M \cong R^n \oplus \bigoplus_{i=1}^{r} R/(d_i).$$

Außerdem sind n, r und das r-Tupel von Idealen $((d_1), \ldots, (d_r))$ eindeutig durch M bestimmt.

(ii) (Zweiter Struktursatz: Primärteiler) *Es gibt $n, s \in \mathbb{N}$, Primelemente $p_1, \ldots, p_s \in R$ und natürliche Zahlen $e_1, \ldots, e_s \in \mathbb{N} \setminus \{0\}$, so dass*

$$M \cong R^n \oplus \bigoplus_{i=1}^{s} R/(p_i^{e_i}).$$

Außerdem sind n, s und das s-Tupel $(((p_1), e_1) \ldots, ((p_s), e_s))$ bis auf die Reihenfolge eindeutig durch M bestimmt.

Beweis. Zu (i). Nach Korollar 8.3.7 besitzt M eine endliche Präsentation

$$R^m \xrightarrow{g} R^n \xrightarrow{f} M,$$

so dass $M \cong R^n / \operatorname{im} g$. Die Existenzaussage folgt nun aus dem Elementarteilersatz 8.3.25, und die Eindeutigkeitsaussage wurde bereits in Proposition 8.3.20 bewiesen.

Zu (ii). Man erhält eine solche Zerlegung von M aus der Elementarteilerzerlegung in (i), indem man jedes $R/(d_i)$ mit dem chinesischen Restsatz weiter zerlegt. Zur Eindeutigkeit bemerken wir, dass für ein Primelement $p \in R$ gilt

$$T_p(M) \cong T_p\left(\bigoplus_{i=1}^{s} R/(p_i^{e_i})\right) = \bigoplus_{i \in I(p)} R/(p_i^{e_i}),$$

wobei $I(p) \subset \{1, \ldots, s\}$ die Teilmenge aller Indizes i ist, so dass p_i und p assoziiert sind. Denn die Inklusion von rechts nach links ist klar, und da sie für jedes p gilt folgt die umgekehrte Inklusion aus Proposition 8.3.13. Es bleibt dann zu zeigen: Für einen R-Modul M mit

$$M \cong \bigoplus_{i=1}^{k} R/(p^{e_i}) \quad \text{und} \quad 1 \leq e_1 \leq \cdots \leq e_k$$

sind k und das k-Tupel $(e_1, \ldots, e_k)$ eindeutig durch M bestimmt. Dies folgt aber aus der Proposition 8.3.20, denn es gilt $(p^{e_k}) \subset \cdots \subset (p^{e_1})$. □

Korollar 8.3.31. *Sei R ein Hauptidealring. Ist M ein endlich erzeugter R-Modul, so ist der Quotientenmodul $M/T(M)$ frei. Insbesondere sind alle endlich erzeugten torsionsfreien R-Moduln frei.*

Beweis. Dies folgt unmittelbar aus Satz 8.3.30. □

Definition 8.3.32 (Rang, Elementarteiler, Primärteiler)**.** Sei R ein Hauptidealring und sei M ein endlich erzeugter R-Modul.

- Die natürliche Zahl n wie im Satz 8.3.30(i,ii) heißt der *Rang* von M.

- Die Elemente $d_1, \ldots, d_r$ wie im Satz 8.3.30(i) heißen die *Elementarteiler* von M.
- Die Primpotenzen $p_1^{e_1}, \ldots, p_s^{e_s}$ wie im Satz 8.3.30(ii) heißen die *Primärteiler* von M.

Bemerkung 8.3.33. Manchmal werden stattdessen die Primpotenzen $p_i^{e_i}$ als Elementarteiler von M bezeichnet, und die Elemente d_i als die *invarianten Faktoren* von M.

Bemerkung 8.3.34. Sei R ein Hauptidealring. Der Begriff der Elementarteiler aus Definition 8.3.32 unterscheidet sich von den aus Abschnitt 8.3.3, indem die Elementarteiler eines R-Moduls keine Einheiten sein dürfen. Der genaue Zusammenhang ist folgender:

(i) Sei F ein freier R-Modul vom Rang $n \in \mathbb{N}$. Die Elementarteiler eines Untermoduls $M \subset F$ wie im Satz 8.3.25, die keine Einheiten sind, sind genau die Elementarteiler des Quotienten F/M.

(ii) Sei $R^m \xrightarrow{g} R^n \xrightarrow{f} M$ eine Präsentation eines R-Moduls M. Die Elementarteiler von g wie im Satz 8.3.21, die keine Einheiten sind, sind genau die Elementarteiler von M.

Bemerkung 8.3.35. Sei R ein Hauptidealring und M ein endlich erzeugter R-Modul.

(i) Der Rang von M ist gleich dem Rang von $M/T(M)$. Die Elementarteiler und Primärteiler von M sind genau die von $T(M)$.

(ii) Die Elementarteiler und Primärteiler von M bestimmen sich gegenseitig. Die Primärteiler sind nämlich die Primpotenzen, die in den Primfaktorzerlegungen der Elementarteiler vorkommen. Seien umgekehrt $p_1, \ldots, p_k$ die paarweise nicht-assoziierten Primelemente, deren Potenzen in den Primärteilern vorkommen, und seien $e(i,0) \geq e(i,1) \geq \cdots$ die Exponenten dieser Potenzen von p_i in absteigender Reihenfolge und mit Nullen fortgesetzt. Dann ist der Elementarteiler d_{r-j} gleich dem Produkt $p_1^{e(1,j)} \ldots p_k^{e(k,j)}$.

Beispiel 8.3.36. Seien $2, 2, 2^2, 2^3, 3, 3^2, 3^2, 5, 5, 7$ die Primärteiler einer abelschen Gruppe A mit 5.443.200 Elementen. Man ordnet die Potenzen der einzelnen Primzahlen in absteigender Reihenfolge:

2	2^3	2^2	2	2
3	3^2	3^2	3	
5	5	5		
7	7			

Die Elementarteiler von A sind dann die Produkte der Spalten: 2|6|180|2.520.

Beispiel 8.3.37.

(i) Seien $p_1, \ldots, p_n \in \mathbb{N}$ paarweise verschiedene Primzahlen. Bis auf Isomorphie gibt es dann genau eine abelsche Gruppe mit $p_1 \ldots p_n$ Elementen, nämlich

$$\mathbb{Z}/p_1 \ldots p_n\mathbb{Z} \cong \mathbb{Z}/p_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p_n\mathbb{Z}.$$

Zum Beispiel ist jede abelsche Gruppe mit 30 Elementen zu $\mathbb{Z}/30\mathbb{Z}$ isomorph.

(ii) Sei $p \in \mathbb{N}$ eine Primzahl. Es gibt dann bis auf Isomorphie genau zwei abelsche Gruppen mit p^2 Elementen, nämlich

$$\mathbb{Z}/p^2\mathbb{Z} \quad \text{und} \quad \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z},$$

genau drei abelsche Gruppen mit p^3 Elementen, nämlich

$$\mathbb{Z}/p^3\mathbb{Z}, \quad \mathbb{Z}/p^2\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \quad \text{und} \quad \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z},$$

und genau fünf abelsche Gruppen mit p^4 Elementen, nämlich

$$\mathbb{Z}/p^4\mathbb{Z}, \quad \mathbb{Z}/p^3\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}, \quad \mathbb{Z}/p^2\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z},$$
$$\mathbb{Z}/p^2\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \quad \text{und} \quad \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p\mathbb{Z}.$$

(iii) Wir bestimmen alle abelschen Gruppen mit 24 Elementen bis auf Isomorphie:

Elementarteiler	Primärteiler	abelsche Gruppe
24	$2^3, 3$	$\mathbb{Z}/24\mathbb{Z}$
2|12	$2, 2^2, 3$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/12\mathbb{Z}$
2|2|6	$2, 2, 2, 3$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$

(iv) Wir bestimmen alle abelschen Gruppen mit 72 Elementen bis auf Isomorphie:

Elementarteiler	Primärteiler	abelsche Gruppe
72	$2^3, 3^2$	$\mathbb{Z}/72\mathbb{Z}$
2|36	$2, 2^2, 3^2$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/36\mathbb{Z}$
2|2|18	$2, 2, 2, 3^2$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/18\mathbb{Z}$
3|24	$2^3, 3, 3$	$\mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/24\mathbb{Z}$
6|12	$2, 2^2, 3, 3$	$\mathbb{Z}/6\mathbb{Z} \oplus \mathbb{Z}/12\mathbb{Z}$
2|6|6	$2, 2, 2, 3, 3$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$

Kapitel 9

Normalformen linearer Endomorphismen

Das Ziel dieses Kapitels ist es, Endomorphismen von endlich-dimensionalen Vektorräumen zu klassifizieren. Zur Erinnerung definiert jeder Endomorphismus f eines K-Vektorraums V einen Modul $V[f]$ über dem Polynomring $K[T]$, so dass für alle $v \in V$ gilt $T \cdot v = f(v)$ (Proposition 8.1.37). Wenn V endlich-dimensional ist, ist $V[f]$ zudem endlich erzeugt. Da $K[T]$ ein Hauptidealring ist, können wir den $K[T]$-Modul $V[f]$ durch den Struktursatz für endlich erzeugte Moduln über Hauptidealringe (Satz 8.3.30) analysieren. Falls K algebraisch abgeschlossen ist, führt diese Analyse zum Begriff der *Jordanschen Normalform* eines Endomorphismus bzw. einer quadratischen Matrix. Die Jordansche Normalform über $\mathbb{C}$ hat viele Anwendungen, zum Beispiel zur Lösung von allgemeinen linearen Differentialgleichungssystemen.

9.1 Das Minimalpolynom

Definition 9.1.1 (algebraisch, transzendent)**.** Sei K ein Körper und A eine K-Algebra. Eine Element $a \in A$ heißt *algebraisch* über K, wenn ein Polynom $p \in K[T] \setminus \{0\}$ existiert mit $p(a) = 0$. Sonst heißt a *transzendent* über K.

Man kann auch die Definition 9.1.1 mithilfe des Einsetzungshomomorphismus

$$\varepsilon_a \colon K[T] \to A, \quad \varepsilon_a(p) = p(a),$$

formulieren: Das Element a ist genau dann transzendent, wenn ε_a injektiv ist. Nach dem Homomorphiesatz gibt es einen induzierten K-linearen Isomorphismus

$$\bar{\varepsilon}_a \colon K[T]/\ker \varepsilon_a \xrightarrow{\sim} \operatorname{im} \varepsilon_a.$$

Proposition 9.1.2 (Charakterisierung der Algebraizität)**.** *Sei K ein Körper, A eine K-Algebra und $a \in A$. Dann sind die folgenden Aussagen äquivalent:*

(i) *a ist algebraisch über K.*

(ii) $\ker \varepsilon_a \neq \{0\}$.

(iii) $\dim_K \ker \varepsilon_a = \infty$.

(iv) $\dim_K \operatorname{im} \varepsilon_a < \infty$.

Beweis. Die Äquivalenz zwischen (i) und (ii) gilt nach Definition, und die Implikation (iii) $\Rightarrow$ (ii) ist klar.

Zu (iv) $\Rightarrow$ (iii). Sei $\operatorname{im}\varepsilon_a$ endlich-dimensional. Da $\dim_K K[T] = \infty$, folgt aus der Dimensionsformel für die lineare Abbildung ε_a, dass $\dim_K \ker\varepsilon_a = \infty$.

Zu (ii) $\Rightarrow$ (iv). Sei $p \in \ker\varepsilon_a \setminus \{0\}$ und sei $d = \deg p$. Wir behaupten, dass die Komposition

$$K[T]_{<d} \hookrightarrow K[T] \xrightarrow{\varepsilon_a} \operatorname{im}\varepsilon_a$$

surjektiv ist. Dies impliziert (iv), denn $\dim_K(K[T]_{<d}) = d < \infty$. Sei $f \in K[T]$ beliebig. Division mit Rest liefert Polynome $q, r \in K[T]$ mit $f = qp + r$ und $\deg(r) < d$. Damit ist $\varepsilon_a(f) = \varepsilon_a(r)$, so dass $\varepsilon_a(f)$ im Bild der obigen Komposition liegt. □

Korollar 9.1.3. *Sei K ein Körper und A eine K-Algebra, so dass $\dim_K A < \infty$. Dann sind alle Elemente von A algebraisch über K.*

Beweis. Dies folgt aus Proposition 9.1.2 (iv) $\Rightarrow$ (i). □

Beispiel 9.1.4.

(i) In der K-Algebra $K[T]$ sind alle Polynome p vom Grad ≥ 1 transzendent uber K. Denn der Einsetzungshomomorphismus $\varepsilon_p\colon K[T] \to K[T]$ bildet ein Polynom von Grad d auf ein Polynom vom Grad $d \cdot \deg(p)$ ab.

(ii) Da $\dim_K M_n(K) = n^2 < \infty$ sind alle quadratischen Matrizen über K algebraisch über K.

(iii) Da $\dim_{\mathbb{R}} \mathbb{C} = 2 < \infty$ ist jede komplexe Zahl algebraisch über $\mathbb{R}$.

(iv) Die komplexe Zahl $i \in \mathbb{C}$ ist algebraisch über $\mathbb{Q}$, denn $i^2 + 1 = 0$, d.h., $T^2 + 1 \in \ker\varepsilon_i$.

(v) Die reelle Zahl $a = \sqrt{2} + \sqrt{3}$ ist algebraisch über $\mathbb{Q}$. Denn

$$a^2 = 5 + 2\sqrt{6} \quad\Longrightarrow\quad (a^2 - 5)^2 = 24 \quad\Longrightarrow\quad a^4 - 10a^2 + 1 = 0.$$

(vi) Die Elemente $\alpha, \beta \in \mathbb{F}_4$ sind algebraisch über $\mathbb{F}_2$, da $\dim_{\mathbb{F}_2} \mathbb{F}_4 = 2 < \infty$. Expliziter gilt $\alpha^3 - 1 = \beta^3 - 1 = 0$ (siehe Bemerkung 2.4.11).

Bemerkung 9.1.5. Komplexe Zahlen, die algebraisch über $\mathbb{Q}$ sind, heißen *algebraische Zahlen*. In der Vorlesung *Algebra* wird gezeigt, dass algebraische Zahlen ein Teilkörper $\bar{\mathbb{Q}} \subset \mathbb{C}$ bilden, der algebraisch abgeschlossen ist. Elemente von $\mathbb{C} \setminus \bar{\mathbb{Q}}$ heißen *transzendente Zahlen*. Es folgt aus Satz 1.3.36, dass $\bar{\mathbb{Q}}$ abzählbar ist, da jedes Polynom über $\mathbb{Q}$ nur endlich viele Koeffizienten aus $\mathbb{Q}$ enthält und nur endlich viele Nullstellen in $\mathbb{C}$ hat. Da $\mathbb{C}$ selbst überabzählbar ist, ist auch die Menge $\mathbb{C} \setminus \bar{\mathbb{Q}}$ überabzählbar. In diesem Sinne sind die meisten komplexen Zahlen transzendent. Es ist jedoch nicht einfach zu entscheiden, ob eine gegebene Zahl transzendent ist oder nicht. Die Zahlen π und e sind bekanntlich transzendent, aber es ist zum Beispiel nicht bekannt, ob $\pi + e$ transzendent ist.

Zur Erinnerung ist $K[T]$ ein Hauptidealring, in dem jedes Nicht-Null-Element zu genau einem monischen Polynom assoziiert ist, d.h., es gibt eine Bijektion

$$\begin{aligned}\{0\} \cup \{\text{monische Polynome in } K[T]\} &\xrightarrow{\sim} \{\text{Ideale in } K[T]\},\\ p &\mapsto (p).\end{aligned}$$

Definition 9.1.6 (Minimalpolynom)**.** Sei K ein Körper, A eine K-Algebra und $a \in A$. Sei $\varepsilon_a\colon K[T] \to A$ der Einsetzungshomomorphismus mit $\varepsilon_a(T) = a$. Das *Minimalpolynom* von a ist das eindeutige Polynom $m_a \in K[T]$, das entweder null oder monisch ist, so dass

$$\ker\varepsilon_a = (m_a).$$

Bemerkung 9.1.7. Sei A eine K-Algebra. Nach Definition ist ein Element $a \in A$ genau dann transzendent über K, wenn $m_a = 0$. Es gilt $m_a = 1$ genau dann, wenn $A = \{0\}$ (siehe Bemerkung 8.1.22). Wenn $A \neq \{0\}$ und $a \in A$ algebraisch ist, ist also m_a ein Polynom vom Grad ≥ 1. Die Elemente a mit $\deg(m_a) = 1$ sind genau die Elemente $\lambda \cdot 1$ mit $\lambda \in K$, für die gilt $m_{\lambda \cdot 1} = T - \lambda$.

Beispiel 9.1.8.

(i) In der $\mathbb{Q}$-Algebra $\mathbb{C}$ gilt $m_i = T^2 + 1$ und $m_{\sqrt{2}} = T^2 - 2$.

(ii) In der $\mathbb{R}$-Algebra $\mathbb{C}$ gilt $m_i = T^2 + 1$ und $m_{\sqrt{2}} = T - \sqrt{2}$.

(iii) In der $\mathbb{F}_2$-Algebra $\mathbb{F}_4$ gilt $m_\alpha = m_\beta = T^2 + T + 1$.

Proposition 9.1.9. *Sei K ein Körper, sei $f\colon A \to B$ ein K-Algebrenhomomorphismus und sei $a \in A$.*

(i) *Es gilt $m_{f(a)} | m_a$.*

(ii) *Ist f injektiv, so gilt $m_{f(a)} = m_a$.*

Beweis. Nach der universellen Eigenschaft des Polynomringes $K[T]$ (Proposition 8.1.45) gilt $\varepsilon_{f(a)} = f \circ \varepsilon_a$. Daraus folgt $\ker \varepsilon_a \subset \ker \varepsilon_{f(a)}$, und die Gleichheit gilt, wenn f injektiv ist. Dies zeigt beide Aussagen. □

9.1.1 Das Minimalpolynom eines Endomorphismus

Wir betrachten jetzt das Minimalpolynom für Elemente der K-Algebra $\mathrm{End}_K(V)$, d.h., für Endomorphismen eines K-Vektorraums V. Jeder Endomorphismus f hat ein Minimalpolynom $m_f \in K[T]$, so dass

$$(m_f) = \ker \varepsilon_f = \{p \in K[T] \mid p(f) = 0\}.$$

Falls V endlich-dimensional ist, ist auch die K-Algebra $\mathrm{End}_K(V)$ endlich-dimensional mit

$$\dim_K \mathrm{End}_K(V) = \dim_K(V)^2.$$

Nach Korollar 9.1.3 ist dann jeder Endomorphismus $f \in \mathrm{End}_K(V)$ algebraisch über K, so dass $m_f \neq 0$. Es gilt sogar $\deg(m_f) \geq 1$, wenn $\dim_K V \in \mathbb{N} \setminus \{0\}$.

Lemma 9.1.10. *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus. Für alle Basen B von V gilt $m_f = m_{[f]_B^B}$.*

Beweis. Dies folgt aus Proposition 9.1.9(ii), denn die bijektive Abbildung

$$\mathrm{End}_K(V) \xrightarrow{\sim} M_n(K), \quad g \mapsto [g]_B^B,$$

ist ein K-Algebrenhomomorphismus, der f auf $[f]_B^B$ abbildet. □

Über einem beliebigen kommutativen Ring R kann man das charakteristische Polynom einer Matrix $A \in M_n(R)$ genau wie über einem Körper definieren:

$$\chi_A = \det(T \cdot I_n - A) \in R[T].$$

Satz 9.1.11. *Sei R ein kommutativer Ring, $n \in \mathbb{N}$ und $A \in M_n(R)$.*

(i) (Satz von Cayley-Hamilton) *Es gilt $\chi_A(A) = 0$.*

(ii) *Ist $m \in R[T]$ ein Polynom mit $m(A) = 0$, so ist m^n durch χ_A teilbar.*

Beweis. Die erste Aussage haben wir schon über einem Körper bewiesen (Satz 6.3.39), und der gegebene Beweis bleibt gültig über einem beliebigen kommutativen Ring. Zur zweiten Aussage sei $m = \sum_{i=0}^{r} a_i T^i$. Wir betrachten die Diagonalmatrix $D = \operatorname{diag}(m, \dots, m) \in M_n(K[T])$, deren Determinante gleich m^n ist. Es gilt

$$D = m(T \cdot I_n) = m(T \cdot I_n) - m(A) = \sum_{i=0}^{r} a_i (T^i \cdot I_n - A^i) = \sum_{i=1}^{r} a_i (T^i \cdot I_n - A^i).$$

Für $i \geq 1$ sei $B_i = \sum_{k=1}^{i} A^{i-k} T^{k-1}$ und sei $B = \sum_{i=1}^{r} a_i B_i$. Eine direkte Berechnung zeigt

$$T^i \cdot I_n - A^i = (T \cdot I_n - A) B_i \quad \text{und somit} \quad D = (T \cdot I_n - A) B.$$

Mit der Multiplikativität der Determinante erhalten wir

$$m^n = \det(D) = \det(T \cdot I_n - A) \det(B) = \chi_A \det(B).$$

Insbesondere ist m^n durch χ_A teilbar, wie gewünscht. □

Korollar 9.1.12 (Minimalpolynom vs. charakteristisches Polynom)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}_K(V)$ ein Endomorphismus.*

(i) *Es gilt $m_f | \chi_f$.*

(ii) *m_f und χ_f haben dieselben Primfaktoren: Für jedes irreduzible Polynom $q \in K[T]$ gilt*

$$q | m_f \iff q | \chi_f.$$

Beweis. Ist B eine Basis von V und ist $A = [f]_B^B$, so gilt $\chi_f = \chi_A$ nach Definition und $m_f = m_A$ nach Lemma 9.1.10. Beide Aussagen folgen nun aus Satz 9.1.11. □

Korollar 9.1.13 (Nullstellen des Minimalpolynoms)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}_K(V)$ ein Endomorphismus. Dann sind die Nullstellen von $m_f \in K[T]$ genau die Eigenwerte von f.*

Beweis. Für jedes $\lambda \in K$ haben wir die folgende Kette von Äquivalenzen:

$$\begin{aligned}
\lambda \text{ ist ein Eigenwert von } f &\iff \lambda \text{ ist eine Nullstelle von } \chi_f && \text{(Proposition 6.3.32(iv))} \\
&\iff \chi_f \text{ ist durch } T - \lambda \text{ teilbar} && \text{(Proposition 6.3.15)} \\
&\iff m_f \text{ ist durch } T - \lambda \text{ teilbar} && \text{(Korollar 9.1.12(ii))} \\
&\iff \lambda \text{ ist eine Nullstelle von } m_f. && \text{(Proposition 6.3.15)}
\end{aligned}$$

□

Rezept 9.1.14 (Bestimmung des Minimalpolynoms)**.** Sei $n \in \mathbb{N}$ und sei $A \in M_n(K)$. Es gibt zwei verschiedene Rezepte, um das Minimalpolynom m_A zu bestimmen:

(i) Falls die Primfaktorzerlegung von χ_A bekannt ist, kann man das Korollar 9.1.12 anwenden. Das Minimalpolynom ist nämlich der monische Teiler p kleinsten Grades von χ_A mit $p(A) = 0$. Man weiß zudem, dass jeder Primfaktor von χ_A mindestens einmal in m_A auftreten muss.

(ii) Im Allgemeinen kann man mit dem Rezept 5.2.22 das kleinste $n \in \mathbb{N}$ bestimmen, so dass die Familie $(A^0, A^1, \dots, A^n)$ linear abhängig ist (dabei identifiziert man $M_n(K)$ mit $K^{n \cdot n}$). Das Rezept liefert dann eine Linearkombination $\sum_{i=0}^{n} a_i A^i = 0$ mit $a_n \neq 0$, und es gilt $m_A = a_n^{-1} \sum_{i=0}^{n} a_i T^i$. Denn nach Konstruktion gilt $p(A) \neq 0$ für alle Polynome $p \neq 0$ vom Grad $< n$.

Beispiel 9.1.15.

(i) Ist $n \geq 1$ und ist $\lambda \in K$, so ist das Minimalpolynom von $\lambda I_n \in M_n(K)$ gleich $T - \lambda$. Dies folgt bereits aus Bemerkung 9.1.7. Alternativ dazu ist $T - \lambda$ der monische Teiler kleinsten Grades von $\chi_{\lambda I_n} = (T - \lambda)^n$, der im Kern von $\varepsilon_{\lambda I_n}$ liegt.

(ii) Sei

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in M_2(K).$$

Es gilt $\chi_A = (T-1)^2$. Es gibt also nur zwei Möglichkeiten für m_A: $T-1$ oder $(T-1)^2$. Da $A - I_2 \neq 0$, ist $m_A \neq T-1$, und damit gilt $m_A = (T-1)^2$.

(iii) Sei

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ -1 & -1 & 0 \end{pmatrix} \in M_3(\mathbb{F}_3).$$

Das charakteristische Polynom von A ist $\chi_A = T^3 + T^2 - T + 1$. Es ist irreduzibel, da es keine Nullstellen hat: $\chi_A(0) = 1$ und $\chi_A(1) = \chi_A(-1) = -1$. Damit ist $m_A = \chi_A$.

(iv) Sei

$$A = \begin{pmatrix} 1 & 4 & -2 \\ 0 & 3 & -1 \\ 0 & 2 & 0 \end{pmatrix} \in M_3(\mathbb{R}).$$

Es gilt $\chi_A = (T-1)^2(T-2)$, und damit gibt es zwei Möglichkeiten für das Minimalpolynom: $(T-1)(T-2)$ oder $(T-1)^2(T-2)$. Man berechnet $(A - I_3)(A - 2I_3) = 0$, so dass $m_A = (T-1)(T-2)$.

Proposition 9.1.16 (Minimalpolynom und Trigonalisierbarkeit/Diagonalisierbarkeit)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$.*

(i) *f ist genau dann trigonalisierbar, wenn m_f in seine Linearfaktoren zerfällt.*

(ii) *f ist genau dann diagonalisierbar, wenn m_f in seine Linearfaktoren zerfällt und die Vielfachheit jeder seiner Nullstellen gleich 1 ist.*

Beweis. Zu (i). Nach Satz 6.4.14 ist f genau dann trigonalisierbar, wenn χ_f in seine Linearfaktoren zerfällt. Nach Korollar 9.1.12(ii) ist dies genau dann der Fall, wenn m_f in seine Linearfaktoren zerfällt.

Zu (ii). Für eine Diagonalmatrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ gilt $m_D = \prod_{\lambda \in \{\lambda_1, \ldots, \lambda_n\}} (T - \lambda)$. Ist f diagonalisierbar, so gibt es eine Basis B von V, so dass $[f]_B^B$ eine Diagonalmatrix ist. Nach Lemma 9.1.10 hat dann m_f die gewünschte Form. Sei umgekehrt $m_f = \prod_{i=1}^{d} (T - \lambda_i)$ mit paarweise verschiedenen Skalaren $\lambda_1, \ldots, \lambda_d$. Nach (i) ist f trigonalisierbar. Um zu schließen, dass f diagonalisierbar ist, genügt es zu zeigen, dass jeder Hauptvektor von f ein Eigenvektor ist. Sei also $v \in V$ ein Hauptvektor zu λ_j von f. Nach Lemma 6.4.5 ist v kein Hauptvektor zu λ_i für $i \neq j$, und insbesondere gilt $f(v) - \lambda_i v \neq 0$. Aus

$$0 = m_f(f)(v) = \prod_{i=1}^{d} (f(v) - \lambda_i v)$$

folgt nun $f(v) - \lambda_j v = 0$, d.h., v ist ein Eigenvektor zu λ_j. □

Sei $K \subset L$ eine Körpererweiterung und sei $A \in M_n(K)$. Dann können wir A als Matrix über L betrachten. Das charakeristische Polynom χ_A ist aber unabhängig von dem Grundkörper, nach der Leibniz-Formel für die Determinante $\det(T \cdot I_n - A)$. Die analoge Aussage gilt auch für das Minimalpolynom m_A, aber sie ist nicht offensichtlich, denn das Minimalpolynom von A über K kann neue Teiler über L bekommen:

Proposition 9.1.17 (Minimalpolynom unter Körpererweiterungen)**.** *Sei $K \subset L$ eine Körpererweiterung, sei $n \in \mathbb{N}$ und seien $\zeta\colon K[T] \hookrightarrow L[T]$ und $\xi\colon M_n(K) \hookrightarrow M_n(L)$ die Inklusionsabbildungen. Für alle $A \in M_n(K)$ gilt dann $\zeta(m_A) = m_{\xi(A)}$ in $L[T]$.*

Beweis. Nach Definition des Minimalpolynoms genügt es zu zeigen, dass $\ker \varepsilon_{\xi(A)} = (\zeta(m_A))$, und die Inklusion $(\zeta(m_A)) \subset \ker \varepsilon_{\xi(A)}$ ist klar. Da die Inklusionsabbildung $K \hookrightarrow L$ ein Ringhomomorphismus ist, hat L nach Beispiel 8.1.28(iii) eine Struktur von K-Vektorraum. Sei $(\mu_i)_{i\in I}$ eine Basis von L über K, so dass die K-lineare Abbildung

$$K^{(I)} \to L, \quad (\lambda_i)_{i\in I} \mapsto \sum_{i\in I} \lambda_i \mu_i,$$

bijektiv ist. Man kann dann leicht nachrechnen, dass die folgenden zwei K-linearen Abbildungen auch bijektiv sind:

$$\begin{aligned} K[T]^{(I)} &\to L[T], \qquad (p_i)_{i\in I} \mapsto \sum_{i\in I} \mu_i \zeta(p_i), \\ M_n(K)^{(I)} &\to M_n(L), \quad (A_i)_{i\in I} \mapsto \sum_{i\in I} \mu_i \xi(A_i). \end{aligned}$$

Unter diesen Isomorphismen kann der Einsetzungshomomorphismus $\varepsilon_{\xi(A)}$ mit der Abbildung

$$\begin{aligned} \varepsilon_A^{(I)} \colon K[T]^{(I)} &\to M_n(K)^{(I)}, \\ (p_i)_{i\in I} &\mapsto (\varepsilon_A(p_i))_{i\in I}, \end{aligned}$$

identifiziert werden, d.h., das folgende Quadrat ist kommutativ:

$$\begin{array}{ccc} K[T]^{(I)} & \xrightarrow{\varepsilon_A^{(I)}} & M_n(K)^{(I)} \\ \wr\downarrow & & \downarrow\wr \\ L[T] & \xrightarrow{\varepsilon_{\xi(A)}} & M_n(L). \end{array}$$

Deswegen gibt es einen induzierten Isomorphismus

$$\ker(\varepsilon_A^{(I)}) \xrightarrow{\sim} \ker \varepsilon_{\xi(A)}, \quad (p_i)_{i\in I} \mapsto \sum_{i\in I} \mu_i \zeta(p_i).$$

Aber der Kern von $\varepsilon_A^{(I)}$ ist $(\ker \varepsilon_A)^{(I)} = (m_A)^{(I)} \subset K[T]^{(I)}$. Damit erhalten wir

$$\ker \varepsilon_{\xi(A)} = \left\{ \sum_{i\in I} \mu_i \zeta(p_i) \;\middle|\; \text{für alle } i \in I \text{ gilt } p_i \in (m_A) \right\}.$$

Daraus folgt unmittelbar, dass $\ker \varepsilon_{\xi(A)} \subset (\zeta(m_A))$, wie gewünscht. □

9.2 Strukturs̈atze für Endomorphismen

In diesem Abschnitt übersetzen wir den Struktursatz 8.3.30 für den Hauptidealring $K[T]$ in eine konkretere Aussage über Darstellungsmatrizen von Endomorphismen.

9.2.1 Begleitmatrizen und Jordanblöcke

Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus. Nach dem Struktursatz für endlich erzeugte Moduln über $K[T]$ ist das Paar (V, f) isomorph zu einer direkten Summe von Paaren der Gestalt $(K[T]/(p^e), x \mapsto T \cdot x)$. Wir untersuchen jetzt besondere Darstellungsmatrizen der Multiplikation mit T auf $K[T]/(p^e)$, die später als Bausteine der Jordanschen Normalform von f verwendet werden. Alle Resultate gelten tatsächlich über einem beliebigen kommutativen Ring.

Definition 9.2.1 (Begleitmatrix)**.** Sei R ein kommutativer Ring und sei

$$p = T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0$$

ein monisches Polynom vom Grad n über R. Die *Begleitmatrix* zu p ist die $n \times n$-Matrix

$$A(p) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix} \in M_n(R).$$

Definition 9.2.2 (verallgemeinerter Jordanblock, Jordanblock)**.** Sei R ein kommutativer Ring, sei $p \in R[T]$ ein monisches Polynom vom Grad n und sei $e \in \mathbb{N}$. Der *verallgemeinerte Jordanblock* (oder die *Hyperbegleitmatrix*) zu p der Länge e ist die Matrix

$$A_e(p) = \begin{pmatrix} A(p) & & & 0 \\ N & A(p) & & \\ & \ddots & \ddots & \\ 0 & & N & A(p) \end{pmatrix} \in M_{ne}(R), \quad \text{wobei} \quad N = \begin{pmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 \end{pmatrix} \in M_n(R).$$

Der *Jordanblock* der Größe e zu einem $\lambda \in R$ ist die Matrix

$$J_e(\lambda) := A_e(T - \lambda) = \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix} \in M_e(R).$$

Bemerkung 9.2.3. Nach Definition gilt $A(p) = A_1(p)$.

Beispiel 9.2.4.

(i) Die Begleitmatrix zum Polynom $T - a$ ist die 1×1-Matrix (a).

(ii) Es gilt

$$A(T^3 + 1) = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

(iii) Es gilt

$$A_2(T^2 - 2T - 3) = \begin{pmatrix} 0 & 3 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

Lemma 9.2.5. *Sei R ein kommutativer Ring und sei $p \in R[T]$ monisch vom Grad n. Dann ist $R[T]/(p)$ ein freier R-Modul vom Rang n mit Basis $B = (1+(p), T+(p), \dots, T^{n-1}+(p))$.*

Beweis. Da p monisch ist, ist jedes Nicht-Null-Vielfache von p vom Grad $\geq n$. Deswegen kann eine Linearkombination $\sum_{i=0}^{n-1} r_iT^i$ mit $r_i \in R$ nur in (p) liegen, wenn alle r_i gleich Null sind, so dass B linear unabhängig ist. Sei $k \geq n$. Da $T^{k-n}p \in (p)$ ist $T^k + (p)$ eine Linearkombination der Elemente $T^{k-i}+(p)$ mit $i \in \{1, \dots, n\}$. Durch vollständige Induktion über $k \geq n$ schließen wir, dass B auch erzeugend ist. □

Bemerkung 9.2.6. Die Begleitmatrix $A(p)$ ist genau die Darstellungsmatrix der Multiplikation mit T auf $R[T]/(p)$ bezüglich der Basis B aus Lemma 9.2.5, denn

$$T \cdot (T^{i-1} + (p)) = \begin{cases} T^i + (p), & \text{falls } i \in \{1, \dots, n-1\}, \\ -a_0 - a_1T - \cdots - a_{n-1}T^{n-1} + (p), & \text{falls } i = n. \end{cases}$$

Proposition 9.2.7 (Eigenschaften der verallgemeinerten Jordanblöcke)**.** *Sei R ein kommutativer Ring, sei $p = T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0$ ein monisches Polynom über R und sei $e \in \mathbb{N}$.*

(i) *Es gilt $\chi_{A_e(p)} = p^e$.*

(ii) *Der Kern des Einsetzungshomomorphismus $\varepsilon_{A_e(p)} \colon R[T] \to M_{ne}(R)$ ist das Ideal (p^e).*

(iii) *Es gibt einen Isomorphismus von $R[T]$-Moduln*

$$\begin{aligned} \varphi \colon R[T]/(p^e) &\xrightarrow{\sim} R^{ne}[L_{A_e(p)}], \\ q + (p^e) &\mapsto q \cdot e_1. \end{aligned}$$

Bemerkung 9.2.8. Falls R ein Körper ist, sagt die zweite Aussage der Proposition 9.2.7, dass das Minimalpolynom von $A_e(p)$ gleich p^e ist.

Beweis. Zu (i). Nach Korollar 5.3.36 genügt es den Fall $e = 1$ zu behandeln. Wir verfahren durch Induktion über $\deg(p) \in \mathbb{N}$. Falls $\deg(p) = 0$ ist $p = 1$, und die Determinante einer 0×0-Matrix ist auch gleich 1. Sei also $\deg(p) \geq 1$. Das Polynom $\chi_{A(p)}$ ist die Determinante der Matrix

$$T \cdot I_n - A(p) = \begin{pmatrix} T & 0 & \dots & 0 & a_0 \\ -1 & T & \dots & 0 & a_1 \\ 0 & -1 & \dots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & T + a_{n-1} \end{pmatrix} \in M_n(R[T]).$$

Wir wenden den Laplaceschen Entwicklungssatz (Satz 5.3.31) mit der ersten Zeile an:

$$\chi_{A(p)} = T \cdot \chi_{A(q)} + (-1)^{n+1} a_0 \cdot \det \begin{pmatrix} -1 & T & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{pmatrix},$$

wobei $q = T^{n-1} + a_{n-1}T^{n-2} + \cdots + a_1$. Die letzte Matrix ist eine obere Dreiecksmatrix, und nach Korollar 5.3.34 ist ihre Determinante gleich $(-1)^{n-1}$. Nach Induktionsvorausetzung ist $\chi_{A(q)} = q$. Damit gilt

$$\chi_{A(p)} = T \cdot q + (-1)^{n+1}(-1)^{n-1} a_0 = T \cdot q + a_0 = p.$$

Zu (iii). Wir zeigen zunächst, dass $p^e \cdot e_1 = 0$, so dass die Abbildung φ wohldefiniert ist. Dies folgt aus (i) und dem Satz von Cayley-Hamilton: $p^e \cdot e_1 = \chi_{A_e(p)} \cdot e_1 = \chi_{A_e(p)}(A_e(p))e_1 = 0$. Man kann aber auch einen direkten Beweis geben: Im $R[T]$-Modul $R^{ne}[L_{A_e(p)}]$ gilt

$$T \cdot e_i = A_e(p)_{*i} = \begin{cases} e_{i+1}, & \text{falls } i \notin n\mathbb{Z}, \\ e_{nd+1} - \sum_{k=0}^{n-1} a_k e_{n(d-1)+k+1}, & \text{falls } i = nd \text{ mit } d < e, \\ -\sum_{k=0}^{n-1} a_k e_{n(e-1)+k+1}, & \text{falls } i = ne. \end{cases}$$

Daraus folgt $p \cdot e_{n(d-1)+1} = e_{nd+1}$ für alle $d \in \{1, \dots, e-1\}$ und $p \cdot e_{n(e-1)+1} = 0$, so dass $p^e \cdot e_1 = 0$. Sei nun B die Basis von $R[T]/(p^e)$ aus Lemma 9.2.5 und sei E die Standardbasis von R^{ne}. Die obige Berechnung zeigt, dass für alle i gilt $T^{i-1} \cdot e_1 \in e_i + \mathrm{Span}_R\{e_1, \dots, e_{i-1}\}$. Die Darstellungsmatrix $[\varphi]_E^B$ ist daher eine obere Dreiecksmatrix mit 1 auf der Hauptdiagonale. Insbesondere ist ihre Determinante gleich 1, so dass φ ein Isomorphismus ist.

Zu (ii). Nach (i) und dem Satz von Cayley-Hamilton (Satz 9.1.11(i)) gilt $(p^e) \subset \ker \varepsilon_{A_e(p)}$, so dass eine induzierte Abbildung $\bar{\varepsilon}_{A_e(p)} \colon R[T]/(p^e) \to M_{ne}(R)$ existiert. Die Komposition

$$\begin{aligned} R[T]/(p^e) \xrightarrow{\bar{\varepsilon}_{A_e(p)}} M_{ne}(R) &\to R^{ne}, \\ A &\mapsto Ae_1, \end{aligned}$$

bildet $q + (p^e)$ auf $q(A_e(p))e_1$ ab, und sie ist bijektiv nach (iii). Insbesondere ist $\bar{\varepsilon}_{A_e(p)}$ injektiv, so dass $\ker \varepsilon_{A_e(p)} \subset (p^e)$. □

9.2.2 Die Frobenius- und Jordansche Normalformen

Lemma 9.2.9. *Sei K ein Körper und V ein K-Vektorraum mit einem Endomorphismus $f \in \mathrm{End}_K(V)$. Der $K[T]$-Modul $V[f]$ ist genau dann ein endlich erzeugter Torsionsmodul, wenn $\dim_K V < \infty$.*

Beweis. Eine Basis von V ist insbesondere ein Erzeugendensystem von $V[f]$. Falls $\dim_K V < \infty$ ist deswegen $V[f]$ endlich erzeugt. Sei nun $V[f]$ endlich erzeugt. Nach Satz 8.3.30 gibt es einen Isomorphismus

$$V[f] \cong K[T]^n \oplus \bigoplus_{i=1}^{r} K[T]/(d_i)$$

mit Polynomen $d_i \in K[T] \setminus \{0\}$, und $V[f]$ ist genau dann ein Torsionsmodul, wenn $n = 0$ gilt. Falls $n \geq 1$ ist $\dim_K V = \infty$, da $\dim_K K[T] = \infty$. Falls $n = 0$ ist aber $\dim_K V < \infty$, denn $\dim_K K[T]/(d_i) = \deg(d_i) < \infty$ nach Lemma 9.2.5. □

Satz 9.2.10 (Normalformen von Endomorphismen)**.** *Sei K ein Körper, V ein endlichdimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus.*

(i) (Frobenius-Normalform) *Es gibt eine Basis B von V, so dass*

$$[f]_B^B = \begin{pmatrix} A(d_1) & & 0 \\ & \ddots & \\ 0 & & A(d_r) \end{pmatrix}$$

mit einem $r \in \mathbb{N}$ und monischen Polynomen $d_1 | d_2 | \cdots | d_r$. Außerdem sind r und das r-Tupel $(d_1, \ldots, d_r)$ eindeutig durch (V, f) bestimmt. Zudem gilt

$$\chi_f = \prod_{i=1}^{r} d_r \quad \textit{und} \quad m_f = d_r.$$

(ii) (verallgemeinerte Jordansche Normalform) *Es gibt eine Basis B von V, so dass*

$$[f]_B^B = \begin{pmatrix} A_{e_1}(p_1) & & 0 \\ & \ddots & \\ 0 & & A_{e_s}(p_s) \end{pmatrix}$$

mit einem $s \in \mathbb{N}$, monischen irreduziblen Polynomen $p_1, \ldots, p_s \in K[T]$ und natürlichen Zahlen $e_1, \ldots, e_s \in \mathbb{N} \setminus \{0\}$. Außerdem sind s und das s-Tupel $((p_1, e_1), \ldots, (p_s, e_s))$ bis auf die Reihenfolge eindeutig durch (V, f) bestimmt. Zudem gilt

$$\chi_f = \prod_{i=1}^{s} p_i^{e_i} \quad \textit{und} \quad m_f = \prod_{p \in \{p_1, \ldots, p_s\}} p^{\max\{e_i \mid p_i = p\}}.$$

Beweis. Nach Lemma 9.2.9 ist $V[f]$ ein endlich erzeugter Torsionsmodul über $K[T]$. Die Existenz- und Eindeutigkeitsaussagen in (i) und (ii) folgen nun aus dem Struktursatz 8.3.30 und der Proposition 9.2.7(iii). Wir erklären stellvertretend die verallgemeinerte Jordansche Normalform. Nach den angeführten Sätzen gibt es einen Isomorphismus von $K[T]$-Moduln

$$V[f] \cong \bigoplus_{i=1}^{s} K^{n_i e_i}[L_{A_{e_i}(p_i)}],$$

wobei $n_i = \deg(p_i)$. Für eine Matrix $A \in M_n(K)$ ist die Darstellungsmatrix der Multiplikation mit T auf $K^n[L_A]$ bezüglich der Standarbasis von K^n genau die Matrix A. Setzt man die Standardbasen der Vektorräume $K^{n_i e_i}$ zusammen, so erhält man eine Basis B von V, so dass die Matrix $[f]_B^B$ die gewünschte Form hat.

Die Berechnung von $\chi_f = \chi_{[f]_B^B}$ folgt aus Korollar 5.3.36 und Proposition 9.2.7(i). Für ein Polynom $p \in K[T]$ und eine Blockdiagonalmatrix

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_s \end{pmatrix} \quad \text{gilt} \quad p(A) = \begin{pmatrix} p(A_1) & & 0 \\ & \ddots & \\ 0 & & p(A_s) \end{pmatrix}.$$

Damit ist das Minimalpolynom von A das kleinste gemeinsame Vielfache der Minimalpolynome der Blöcke A_i. Die Berechnung von $m_f = m_{[f]_B^B}$ folgt nun aus Proposition 8.2.40(ii) und Proposition 9.2.7(ii). □

Bemerkung 9.2.11. Die Polynome d_i im Satz 9.2.10(i) sind genau die Elementarteiler des $K[T]$-Moduls $V[f]$, und die Polynome $p_i^{e_i}$ im Satz 9.2.10(ii) sind die Primärteiler von $V[f]$.

Bemerkung 9.2.12. Für eine endliche abelsche Gruppe A ist das Produkt der Elementarteiler genau die Mächtigkeit von A, und der größte Elementarteiler ist die kleinste natürliche Zahl $n \geq 1$ mit $nA = \{0\}$, die auch als *Exponent* von A bezeichnet wird. Es gibt also die folgenden Parallelen zwischen $\mathbb{Z}$ und $K[T]$:

Hauptidealring R	$\mathbb{Z}$	$K[T]$
Modul über R	abelsche Gruppe A	Endomorphismus f von V
endlich erzeugter Torsionsmodul	$\lvert A\rvert < \infty$	$\dim_K V < \infty$
Produkt der Elementarteiler	Mächtigkeit von A	charakteristisches Polynom χ_f
größter Elementarteiler	Exponent von A	Minimalpolynom m_f

Das Analogon des Satzes von Cayley-Hamilton für endliche abelsche Gruppen ist die Aussage, dass $|A|A = \{0\}$.

Rezept 9.2.13 (Bestimmung der Frobenius- und verallgemeinerten Jordan-Normalformen)**.** Sei $A \in M_n(K)$. Der $K[T]$-Modul $K^n[L_A]$ hat die Präsentation

$$\begin{aligned} K[T]^n \xrightarrow{L_{TI_n - A}} K[T]^n &\twoheadrightarrow K^n[L_A], \\ e_i &\mapsto e_i. \end{aligned}$$

Denn aus $T^{k+1}e_i - AT^k e_i \in \operatorname{im} L_{TI_n - A}$ folgt induktiv, dass $\{e_1, \ldots, e_n\}$ ein Erzeugendensystem des K-Vektorraums $K[T]^n / \operatorname{im} L_{TI_n - A}$ ist, so dass die induzierte surjektive Abbildung

$$K[T]^n / \operatorname{im} L_{TI_n - A} \twoheadrightarrow K^n[L_A]$$

ein Isomorphismus sein muss. Nach Bemerkung 8.3.34(ii) findet man die Elementarteiler von $K^n[L_A]$ und damit die Frobenius-Normalform von A, indem man die *Hilfsmatrix* $TI_n - A \in$

$M_n(K[T])$ auf Smith-Normalform bringt. Dazu kann man die Bemerkung 8.3.22 mit der euklidischen Gradfunktion $\deg\colon K[T]\setminus\{0\}\to\mathbb{N}$ anwenden.

Um die Primärteiler von $K^n[L_A]$ (und damit die verallgemeinerte Jordansche Normalform von A) zu bestimmen, braucht man noch die Elementarteiler in ihre Primfaktoren zu zerlegen (siehe Bemerkung 8.3.35(iii)). Dazu gibt es aber keinen Algorithmus.

Beispiel 9.2.14.

(i) Sei

$$A=\begin{pmatrix}3&0&1\\-2&2&-2\\-1&0&1\end{pmatrix}\in M_3(\mathbb{Q}).$$

Wir berechnen die Elementarteiler der Hilfsmatrix TI_3-A, indem wir sie auf Smith-Normalform bringen:

$$\begin{pmatrix}T-3&0&-1\\2&T-2&2\\1&0&T-1\end{pmatrix}\xrightarrow{V_{13}\cdot}\begin{pmatrix}1&0&T-1\\2&T-2&2\\T-3&0&-1\end{pmatrix}$$
$$\xrightarrow[A_{31}(3-T)\cdot]{A_{21}(-2)\cdot}\begin{pmatrix}1&0&T-1\\0&T-2&-2(T-2)\\0&0&-(T-2)^2\end{pmatrix}\xrightarrow{\cdot A_{13}(1-T)}\begin{pmatrix}1&0&0\\0&T-2&-2(T-2)\\0&0&-(T-2)^2\end{pmatrix}$$
$$\xrightarrow{\cdot A_{23}(2)}\begin{pmatrix}1&0&0\\0&T-2&0\\0&0&-(T-2)^2\end{pmatrix}.$$

Die Elementarteiler des $\mathbb{Q}[T]$-Moduls $\mathbb{Q}^3[L_A]$ sind damit $T-2|(T-2)^2$, und seine Primärteiler sind $T-2,(T-2)^2$. Die Frobenius-Normalform F und die verallgemeinerte Jordansche Normalform J von A sind dann

$$F=\begin{pmatrix}2&0&0\\0&0&-4\\0&1&4\end{pmatrix}\quad\text{und}\quad J=\begin{pmatrix}2&0&0\\0&2&0\\0&1&2\end{pmatrix}.$$

(ii) Sei

$$A=\begin{pmatrix}1&-1&1\\1&1&0\\0&0&2\end{pmatrix}\in M_3(\mathbb{R}).$$

Wir bringen die Hilfsmatrix TI_3-A auf Smith-Normalform:

$$\begin{pmatrix}T-1&1&-1\\-1&T-1&0\\0&0&T-2\end{pmatrix}\xrightarrow{\cdot V_{12}}\begin{pmatrix}1&T-1&-1\\T-1&-1&0\\0&0&T-2\end{pmatrix}$$
$$\xrightarrow[\cdot A_{13}(1)]{\cdot A_{12}(1-T)}\begin{pmatrix}1&0&0\\T-1&-T^2+2T-2&T-1\\0&0&T-2\end{pmatrix}\xrightarrow[\cdot V_{23}]{A_{21}(1-T)\cdot}\begin{pmatrix}1&0&0\\0&T-1&-T^2+2T-2\\0&T-2&0\end{pmatrix}$$
$$\xrightarrow{\cdot A_{23}(T-1)V_{23}}\begin{pmatrix}1&0&0\\0&-1&T-1\\0&(T-2)(T-1)&T-2\end{pmatrix}\xrightarrow[A_{32}((T-2)(T-1))\cdot]{\cdot A_{23}(T-1)}\begin{pmatrix}1&0&0\\0&-1&0\\0&0&(T-2)(T^2-2T+2)\end{pmatrix}.$$

Im vorletzten Schritt haben wir die Division von $-T^2+2T-2$ mit Rest durch $T-1$ berechnet, um den Grad an der Stelle $(2,2)$ zu reduzieren. Der Modul $\mathbb{R}^3[L_A]$ hat damit den einzigen Elementarteiler $(T-2)(T^2-2T+2)=T^3-4T^2+6T-4$. Die Begleitmatrix davon ist dann die Frobenius-Normalform F von A, und die verallgemeinerte

Jordansche Normalform J von A besteht aus den Begleitmatrizen der Primfaktoren $T-2$ und T^2-2T+2:

$$F = \begin{pmatrix} 0 & 0 & 4 \\ 1 & 0 & -6 \\ 0 & 1 & 4 \end{pmatrix} \quad \text{und} \quad J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 1 & 2 \end{pmatrix}.$$

Betrachtet man A als Matrix über $\mathbb{C}$, so ist die Frobenius-Normalform unverändert (die Elementarteiler einer Matrix von Polynomen sind unabhängig von dem Grundkörper, wegen ihrer Eindeutigkeit), aber die verallgemeinerte Jordansche Normalform ist nun die Diagonalmatrix $\operatorname{diag}(2, 1+i, 1-i)$.

Die verallgemeinerte Jordansche Normalform vereinfacht sich, wenn der Endomorphismus f trigonalisierbar ist (siehe Definition 6.4.8 und außerdem Satz 6.4.14 für verschiedene Charakterisierungen der Trigonalisierbarkeit). Man erinnert daran, dass f automatisch trigonalisierbar ist, wenn K algebraisch abgeschlossen ist (z.B., wenn $K = \mathbb{C}$).

Korollar 9.2.15 (Jordansche Normalform)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}_K(V)$ ein* trigonalisierbarer *Endomorphismus. Dann gibt es eine Basis B von V, so dass*

$$[f]_B^B = \begin{pmatrix} J_{e_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{e_s}(\lambda_s) \end{pmatrix}$$

mit einem $s \in \mathbb{N}$, Skalaren $\lambda_1, \ldots, \lambda_s \in K$ und natürlichen Zahlen $e_1, \ldots, e_s \in \mathbb{N} \setminus \{0\}$. Außerdem sind s und das s-Tupel $((\lambda_1, e_1), \ldots, (\lambda_s, e_s))$ bis auf die Reihenfolge eindeutig durch (V, f) bestimmt. Zudem gilt

$$\chi_f = \prod_{i=1}^{s} (T-\lambda_i)^{e_i} \quad \textit{und} \quad m_f = \prod_{\lambda \in \sigma(f)} (T-\lambda)^{\max\{e_i \mid \lambda_i = \lambda\}}.$$

Eine Basis B vie im Korollar heißt eine *Jordan-Basis* für f, und die (bis auf die Reihenfolge der Jordanblöcke eindeutig bestimmte) Matrix $[f]_B^B$ heißt die *Jordansche Normalform* von f.

Beweis. Die Trigonalisierbarkeit von f bedeutet, dass χ_f in seine Linearfaktoren zerfällt (Satz 6.4.14). Damit ist jedes irreduzible Polynom p_i im Satz 9.2.10(ii) gleich $T - \lambda_i$ mit einem $\lambda_i \in K$, was das Korollar liefert. □

Bemerkung 9.2.16 (Vielfachheiten und Jordanblöcke)**.** Mit Worten sagt die letzte Aussage im Korollar 9.2.15 folgendes: Die Vielfachheit von λ in χ_f (d.h., die algebraische Vielfachheit von λ bzgl. f) ist die Summe der Größen der Jordanblöcke zu λ in der Jordanschen Normalform von f, und die Vielfachheit von λ in m_f ist die Größe des *größten* Jordanblocks zu λ.

Auf der anderen Seite ist die geometrische Vielfachheit von λ bzgl. f genau die Anzahl der Jordanblöcke zu λ, denn bezüglich jedes Blocks hat λ die geometrische Vielfachheit 1: Ist $n \geq 1$, so gilt

$$\operatorname{Eig}_\lambda(J_n(\lambda)) = \operatorname{Span}_K\{e_n\}.$$

Nach Korollar 6.3.36(ii) ist der Endomorphismus f genau dann diagonalisierbar, wenn alle Blöcke die Größe 1 haben, d.h., wenn seine Jordansche Normalform eine Diagonalmatrix ist.

Beispiel 9.2.17. Sei

$$A = \begin{pmatrix} 3 & 0 & 1 \\ -2 & 2 & -2 \\ -1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Es gilt (Entwicklung nach der dritten Zeile):

$$\begin{aligned}\chi_A &= \det\begin{pmatrix} 0 & -1 \\ T-2 & 2 \end{pmatrix} + (T-1)\begin{pmatrix} T-3 & 0 \\ 2 & T-2 \end{pmatrix} \\ &= T-2+(T-1)(T-3)(T-2) \\ &= (T-2)(T^2-4T+4) = (T-2)^3.\end{aligned}$$

Insbesondere ist A trigonalisierbar mit dem einzigen Eigenwert 2. Sei J eine Jordansche Normalform von A. Es gibt dann drei Möglichkeiten für J:

- Drei Jordanblöcke ($\mu_A^{\text{geom}}(2) = 3$) und der größte hat die Größe 1 ($m_A = (T-2)$).
- Zwei Jordanblöcke ($\mu_A^{\text{geom}}(2) = 2$) und der größte hat die Größe 2 ($m_A = (T-2)^2$).
- Ein Jordanblock ($\mu_A^{\text{geom}}(2) = 1$) der Größe 3 ($m_A = (T-2)^3$).

Um J zu bestimmen, kann man in diesem Fall entweder die geometrische Vielfachheit von 2 berechnen (mit dem Gaußschen Eliminationsverfahren) oder das Minimalpolynom bestimmen, was wir jetzt tun. Es gilt $m_A \neq T-2$, denn

$$A - 2I_3 = \begin{pmatrix} 1 & 0 & 1 \\ -2 & 0 & -2 \\ -1 & 0 & -1 \end{pmatrix} \neq 0.$$

Es gilt aber $(A-2I_3)^2 = 0$, so dass $m_A = (T-2)^2$. Damit hat die Matrix J zwei Jordanblöcke:

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Diese Antwort hatten wir schon im Beispiel 9.2.14(i) mit einer anderen Methode erhalten.

Bemerkung 9.2.18. Sei $A \in M_n(K)$ trigonalisierbar. Falls alle Eigenwerte von A die algebraische Vielfachheit ≤ 3 haben, dann ist die Jordansche Normalform von A durch das charakteristische Polynom und entweder das Minimalpolynom oder die geometrischen Vielfachheiten der Eigenwerte bestimmt, wie im Beispiel 9.2.17. Von den 4×4-Matrizen

$$\begin{pmatrix} \lambda & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \quad \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{pmatrix} \quad \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$$

in Jordanscher Normalform mit charakteristischem Polynom $(T-\lambda)^4$ haben die ersten beiden dasselbe Minimalpolynom $(T-\lambda)^2$ aber verschiedene geometrische Vielfachheiten 3 und 2, und die letzten beiden dieselbe geometrische Vielfachheit 2 aber verschiedene Minimalpolynome $(T-\lambda)^2$ und $(T-\lambda)^3$.

Haben alle Eigenwerte von A die algebraische Vielfachheit ≤ 6, so lässt sich die Jordansche Normalform von A allein aus χ_A, m_A und μ_A^{geom} ablesen. Das kleinste Gegenbeispiel dazu ist das Paar von nicht-ähnlichen 7×7-Matrizen

$$\begin{pmatrix} \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \lambda \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{pmatrix},$$

die beide $\chi_A = (T-\lambda)^7$, $m_A = (T-\lambda)^3$ und $\mu_A^{\text{geom}}(\lambda) = 3$ erfüllen.

Bemerkung 9.2.19 (Klassifikation von trigonalisierbaren Endomorphismen bis auf Isomorphie)**.** Sei $\mathfrak{Trig}\mathfrak{End}$ die Menge aller Isomorphieklassen von Paaren (V, f), wobei V ein endlich-dimensionaler K-Vektorraum ist und f ein trigonalisierbarer Endomorphismus von V ist. Sei $\sim$ die folgende Äquivalenzrelation auf $(K\times(\mathbb{N}\setminus\{0\}))^s$: $x \sim y$ genau dann, wenn eine Permutation $\sigma \in S_s$ existiert, so dass $y_i = x_{\sigma(i)}$ für alle $i \in \{1, \dots, s\}$. Nach Korollar 9.2.15 gibt es dann eine bijektive Abbildung

$$\coprod_{s\in\mathbb{N}} (K \times (\mathbb{N} \setminus \{0\}))^s/\sim \xrightarrow{\sim} \mathfrak{Trig}\mathfrak{End},$$
$$[((\lambda_1, e_1), \dots, (\lambda_s, e_s))] \mapsto [(K^{e_1+\dots+e_s}, L_{\mathrm{diag}(J_{e_1}(\lambda_1),\dots,J_{e_s}(\lambda_s))})].$$

Die Teilmenge $\mathfrak{Diag}\mathfrak{End} \subset \mathfrak{Trig}\mathfrak{End}$ der diagonalisierbaren Endomorphismen entspricht der Teilmenge der linken Seite, in der alle e_i gleich 1 sind. Vergleiche dazu die Bemerkung 6.2.35. Diese Klassifikation verallgemeinert sich auf *beliebige* Endomorphismen, indem man K durch die Menge aller monischen irreduziblen Polynome über K ersetzt.

9.2.3 Berechnung einer Jordan-Basis

In diesem Abschnitt erklären wir, wie man eine Jordan-Basis für einen trigonalisierbaren Endomorphismus bestimmen kann. Mit einer ähnlichen Methode (indem man $T - \lambda$ durch ein beliebiges irreduzibles Polynom ersetzt) kann man sogar eine verallgemeinerte Jordan-Basis für einen beliebigen Endomorphismus finden. Der Einfachheit halber behandeln wir aber nur den trigonalisierbaren Fall. Für die meisten Anwendungen ist das eigentlich hinreichend, denn man kann oft den Grundkörper durch einen algebraisch abgeschlossenen Körper ersetzen (z.B., $\mathbb{R}$ durch $\mathbb{C}$).

Bemerkung 9.2.20. Die Beweismethoden in diesem Abschnitt verwenden nur die Resultate aus Kapitel 6 und sind unabhängig von den Strukturätzen aus Kapitel 8. Sie liefern also einen neuen Beweis der Existenz der Jordanschen Normalform.

Sei K ein Körper, V ein K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein Endomorphismus. Für einen Vektor $v \in V$ bezeichnen wir mit $\langle v\rangle_f$ den von v erzeugten zyklischen Untermodul von $V[f]$, d.h.:

$$\langle v\rangle_f = \mathrm{Span}_{K[T]}\{v\} = \mathrm{Span}_K\{v, f(v), f^2(v), \dots\} \subset V.$$

Die folgende Proposition ist der Kernpunkt zur Bestimmung einer Jordan-Basis:

Proposition 9.2.21 (Jordanblöcke aus Hauptvektoren)**.** *Sei V ein K-Vektorraum, $f \in \mathrm{End}_K(V)$ und $\lambda \in K$. Sei $v \in V$ ein Hauptvektor zu λ von f der Stufe $n \in \mathbb{N}\setminus\{0\}$. Dann ist die Familie*

$$B_v = (v, (f - \lambda\,\mathrm{id}_V)(v), \dots, (f - \lambda\,\mathrm{id}_V)^{n-1}(v))$$

eine Basis von $\langle v\rangle_f$, für die gilt

$$[f_{\langle v\rangle_f}]^{B_v}_{B_v} = J_n(\lambda).$$

Die Familie B_v heißt die *Jordankette* zum Hauptvektor v.

Beweis. Zur linearen Unabhängigkeit sei $\sum_{i=1}^n \mu_i(f-\lambda\,\mathrm{id}_V)^{i-1}(v) = 0$. Wir beweisen $\mu_j = 0$ durch vollständige Induktion über j. Es gilt

$$0 = (f - \lambda\,\mathrm{id}_V)^{n-j}\left(\sum_{i=1}^n \mu_i(f - \lambda\,\mathrm{id}_V)^{i-1}(v)\right) = \sum_{i=1}^n \mu_i(f - \lambda\,\mathrm{id}_V)^{n-j+i-1}(v).$$

Aus $(f-\lambda\,\mathrm{id}_V)^n(v) = 0$ folgt, dass alle Summanden mit $i > j$ null sind, und nach Induktionsvoraussetzung sind alle Summanden mit $i < j$ auch null. Es gilt also $\mu_j(f-\lambda\,\mathrm{id}_V)^{n-1}(v) = 0$.

Da v ein Hauptvektor der Stufe n ist, gilt nach Definition $(f-\lambda\,\mathrm{id}_V)^{n-1}(v)\neq 0$, und damit $\mu_j=0$.

Sei nun $U\subset V$ der von B_v erzeugte Untervektorraum. Es ist klar, dass $U\subset\langle v\rangle_f$. Zur umgekehrten Inklusion genügt es zu zeigen, dass U f-invariant ist. Da $(f-\lambda\,\mathrm{id}_V)^n(v)=0$ enthält U den Vektor $(f-\lambda\,\mathrm{id}_V)^i(v)$ für alle $i\in\mathbb{N}$. Es gilt dann

$$f((f-\lambda\,\mathrm{id}_V)^{i-1}(v))=\lambda(f-\lambda\,\mathrm{id}_V)^{i-1}(v)+(f-\lambda\,\mathrm{id}_V)^i(v)\in U,$$

wie gewünscht. Diese Formel zeigt außerdem, dass die i-te Spalte der Darstellungsmatrix $[f_{\langle v\rangle_f}]_{B_v}^{B_v}$ gleich λe_i+e_{i+1} (oder λe_n wenn $i=n$) ist. □

Zur Erinnerung ist $f\in\mathrm{End}_K(V)$ genau dann trigonalisierbar, wenn sich V als direkte Summe der Haupträume von f zerlegt (Proposition 6.4.10):

$$V=\bigoplus_{\lambda\in K}\mathrm{Hau}_\lambda(f).$$

Um eine Jordan-Basis für f zu finden, genügt es jeden Hauptraum $\mathrm{Hau}_\lambda(f)$ als direkte Summe von Untervektorräumen der Gestalt $\langle v\rangle_f$ zu zerlegen: Man erhält dann eine Jordan-Basis, indem man die Jordanketten B_v aus Proposition 9.2.21 zusammensetzt.

Bemerkung 9.2.22. Es gibt ein paar Varianten der Proposition 9.2.21:

- Ersetzt man $f-\lambda\,\mathrm{id}_V$ durch $\lambda\,\mathrm{id}_V-f$ in der Basis B_v, so erhalten wir einen Jordanblock mit -1 anstelle von 1 unter der Diagonale.
- Kehrt man die Reihenfolge der Basis B_v um, so erhalten wir für die Darstellungsmatrix den transponierten Jordanblock $J_n(\lambda)^\intercal$. Viele Quellen verwenden diese Konvention für die Jordanblöcke, so dass die Jordansche Normalform eine obere statt einer unteren Dreiecksmatrix ist (bei der verallgemeinerten Jordanschen Normalform muss man dann auch die Begleitmatrizen transponieren).

Wir beginnen mit einer ausführlicheren Strukturanalyse der Haupträume. Sei V ein K-Vektorraum und $f\in\mathrm{End}_K(V)$ ein Endomorphismus. Für $n\in\mathbb{N}$ definieren wir den Untervektorraum

$$\mathrm{Hau}_\lambda^n(f):=\ker\big((f-\lambda\cdot\mathrm{id}_V)^n\big)\subset V.$$

Nach Proposition 6.1.12(i) ist $\mathrm{Hau}_\lambda^n(f)$ ein f-invarianter Untervektorraum von V. Es gilt $\mathrm{Hau}_\lambda^0(f)=\{0\}$, $\mathrm{Hau}_\lambda^1(f)=\mathrm{Eig}_\lambda(f)$ und

$$\mathrm{Hau}_\lambda^0(f)\subset\mathrm{Hau}_\lambda^1(f)\subset\mathrm{Hau}_\lambda^2(f)\subset\cdots\subset\mathrm{Hau}_\lambda(f)=\bigcup_{n\in\mathbb{N}}\mathrm{Hau}_\lambda^n(f).$$

Die Hauptvektoren zu λ der Stufe n sind genau die Elemente von $\mathrm{Hau}_\lambda^n(f)\setminus\mathrm{Hau}_\lambda^{n-1}(f)$. Auf der anderen Seite gilt nach Definition $(f-\lambda\,\mathrm{id}_V)(\mathrm{Hau}_\lambda^{n+1}(f))\subset\mathrm{Hau}_\lambda^n(f)$.

Man definiert

$$\mu_f^n(\lambda):=\dim_K\mathrm{Hau}_\lambda^n(f).$$

Es gilt also $\mu_f^{\mathrm{geom}}(\lambda)=\mu_f^1(\lambda)$ und $\mu_f^n(\lambda)\leq\mu_f^{n+1}(\lambda)$ für alle $n\in\mathbb{N}$. Die Dimensionen $\mu_f^n(\lambda)$ heißen die *verallgemeinerten geometrischen Vielfachheiten* von λ bzgl. f. Man definiert ferner

$$\gamma_f^n(\lambda):=\dim_K\left(\frac{\mathrm{Hau}_\lambda^n(f)}{\mathrm{Hau}_\lambda^{n-1}(f)}\right)$$

für alle $n\geq 1$. Anders gesagt ist $\gamma_f^n(\lambda)$ die Dimension eines zu $\mathrm{Hau}_\lambda^{n-1}(f)$ komplementären Untervektorraums von $\mathrm{Hau}_\lambda^n(f)$, d.h., die „Anzahl der linear unabhängigen Hauptvektoren zu λ der Stufe n“. Nach der Dimensionsformel für Quotientenvektorräume gilt

$$\mu_f^{n-1}(\lambda)+\gamma_f^n(\lambda)=\mu_f^n(\lambda).$$

Schließlich setzt man

$$\delta_f^n(\lambda) = \dim_K \left(\frac{\operatorname{Hau}_\lambda^n(f)}{\operatorname{Hau}_\lambda^{n-1}(f) + (f - \lambda \operatorname{id}_V)(\operatorname{Hau}_\lambda^{n+1}(f))} \right).$$

Diese Dimension ist die „Anzahl der linear unabhängigen Hauptvektoren zu λ der Stufe n, die nicht von einer höheren Stufe stammen“.

Lemma 9.2.23. *Sei V ein K-Vektorraum, $f \in \operatorname{End}_K(V)$ und $\lambda \in K$. Für alle $n \in \mathbb{N} \setminus \{0\}$ induziert $f - \lambda \operatorname{id}_V \colon \operatorname{Hau}_\lambda^{n+1}(f) \to \operatorname{Hau}_\lambda^n(f)$ eine* injektive *Abbildung*

$$\frac{\operatorname{Hau}_\lambda^{n+1}(f)}{\operatorname{Hau}_\lambda^n(f)} \hookrightarrow \frac{\operatorname{Hau}_\lambda^n(f)}{\operatorname{Hau}_\lambda^{n-1}(f)},$$

und es gilt $\gamma_f^{n+1}(\lambda) + \delta_f^n(\lambda) = \gamma_f^n(\lambda)$. Insbesondere gilt $\gamma_f^{n+1}(\lambda) \leq \gamma_f^n(\lambda)$.

Beweis. Dies folgt aus dem Homomorphiesatz 4.1.35, denn der Kern der Komposition

$$\operatorname{Hau}_\lambda^{n+1}(f) \xrightarrow{f - \lambda \operatorname{id}_V} \operatorname{Hau}_\lambda^n(f) \twoheadrightarrow \frac{\operatorname{Hau}_\lambda^n(f)}{\operatorname{Hau}_\lambda^{n-1}(f)}$$

ist genau

$$(f - \lambda \operatorname{id}_V)^{-1}(\operatorname{Hau}_\lambda^{n-1}(f)) = \operatorname{Hau}_\lambda^n(f)$$

nach Definition von $\operatorname{Hau}_\lambda^n(f)$. Die zweite Aussage folgt nun aus der Dimensionsformel für Quotientenvektorräume, denn nach dem Homomorphiesatz gibt es zu jeder linearen Abbildung $h \colon W \to V$ und jedem Untervektorraum $U \subset V$ einen Isomorphismus

$$V/(U + \operatorname{im} h) \xrightarrow{\sim} (V/U)/\operatorname{im}(q \circ h),$$

wobei $q \colon V \to V/U$ die Quotientenabbildung ist. □

Ist V endlich-dimensional, so ist die aufsteigende Folge $(\mu_f^n(\lambda))_{n \in \mathbb{N}}$ stationär. Nach Proposition 6.4.12 gilt genauer

$$\mu_f^n(\lambda) = \mu_f^{\text{alg}}(\lambda) \quad \text{für alle } n \geq \mu_f^{\text{alg}}(\lambda).$$

Das kleinste $n \in \mathbb{N}$ mit dieser Eigenschaft ist eigentlich die Vielfachheit von λ im Minimalpolynom m_f (dies kann man zum Beispiel aus Satz 9.2.10(ii) schließen). Die absteigende Folge $(\gamma_f^n(\lambda))_{n \in \mathbb{N} \setminus \{0\}}$ ist dann schließlich null, und nach Lemma 9.2.23 gilt

$$\mu_f^{\text{geom}}(\lambda) = \gamma_f^1(\lambda) = \sum_{n=1}^{\infty} \delta_f^n(\lambda).$$

Proposition 9.2.24. *Sei V ein endlich-dimensionaler K-Vektorraum, $f \in \operatorname{End}_K(V)$ und $\lambda \in K$. Dann gibt es eine Teilmenge $Q \subset \operatorname{Hau}_\lambda(f) \setminus \{0\}$ mit $|Q| = \mu_f^{\text{geom}}(\lambda)$, so dass die kanonische Abbildung*

$$\bigoplus_{v \in Q} \langle v \rangle_f \to \operatorname{Hau}_\lambda(f)$$

ein Isomorphismus ist. Die Zusammensetzung der Jordanketten B_v mit $v \in Q$ ist dann eine Jordan-Basis für $f_{\operatorname{Hau}_\lambda(f)}$.

Beweis. Die letzte Aussage folgt aus Proposition 9.2.21. Sei $d \in \mathbb{N}$ die kleinste Zahl, so dass $\operatorname{Hau}_\lambda(f) = \operatorname{Hau}_\lambda^d(f)$ (d.h., die Vielfachheit von λ im Minimalpolynom m_f). Wir bilden die Teilmenge Q als Vereinigung $Q = \bigcup_{s=1}^d Q_s$, wobei Q_s aus Hauptvektoren der Stufe s besteht.

Sei $\Gamma_{d+1} = \{0\} \subset \operatorname{Hau}_\lambda(f)$. Man definiert Untervektorräume $\Gamma_s, \Delta_s \subset \operatorname{Hau}_\lambda^s(f)$ für $s \in \{1, \ldots, d\}$ rekursiv wie folgt: Δ_s ist ein komplementärer Untervektorraum zu $(f - \lambda \operatorname{id}_V)(\Gamma_{s+1}) + \operatorname{Hau}_\lambda^{s-1}(f)$ und $\Gamma_s = \Delta_s + (f - \lambda \operatorname{id}_V)(\Gamma_{s+1})$ (nach Definition von Δ_s ist die Summe direkt). Sei dann $Q_s \subset \Delta_s$ eine Basis und sei $Q = \bigcup_{s=1}^d Q_s$.

Behauptung. Für alle $s \in \{1, \dots, d\}$ gilt:

(i) Γ_s ist komplementär zu $\operatorname{Hau}_\lambda^{s-1}(f)$ in $\operatorname{Hau}_\lambda^s(f)$,

(ii) Die Einschränkung von $f - \lambda \operatorname{id}_V$ auf Γ_{s+1} ist injektiv.

Zu (i). Nach Definition von Γ_s und Δ_s gilt

$$\Gamma_s + \operatorname{Hau}_\lambda^{s-1}(f) = \Delta_s + (f - \lambda \operatorname{id}_V)(\Gamma_{s+1}) + \operatorname{Hau}_\lambda^{s-1}(f) = \operatorname{Hau}_\lambda^s(f).$$

Die Aussage $\Gamma_s \cap \operatorname{Hau}_\lambda^{s-1}(f) = \{0\}$ beweisen wir für alle $s \in \{1, \dots, d+1\}$ durch absteigende Induktion über s. Falls $s = d+1$ ist die Aussage trivial, denn $\Gamma_{d+1} = \{0\}$. Sei $s \leq d$ und sei $v \in \Gamma_s \cap \operatorname{Hau}_\lambda^{s-1}(f)$. Nach Definition von Γ_s kann man schreiben $v = u + (f - \lambda \operatorname{id}_V)(w)$ mit $u \in \Delta_s$ und $w \in \Gamma_{s+1}$. Nach Definition von Δ_s ist dann $u = 0$. Aus $v \in \operatorname{Hau}_\lambda^{s-1}(f)$ folgt dann $w \in \operatorname{Hau}_\lambda^s(f)$. Nach Induktionsvoraussetzung ist nun $w = 0$, so dass $v = 0$.

Zu (ii). Wir betrachten das kommutative Diagramm

$$\begin{array}{ccccc} \Gamma_{s+1} & \hookrightarrow & \operatorname{Hau}_\lambda^{s+1}(f) & \xrightarrow{f-\lambda \operatorname{id}_V} & \operatorname{Hau}_\lambda^s(f) \\ & \searrow^{\sim} & \downarrow & & \downarrow \\ & & \dfrac{\operatorname{Hau}_\lambda^{s+1}(f)}{\operatorname{Hau}_\lambda^s(f)} & \hookrightarrow & \dfrac{\operatorname{Hau}_\lambda^s(f)}{\operatorname{Hau}_\lambda^{s-1}(f)}. \end{array}$$

Nach (i) ist der diagonale Pfeil ein Isomorphismus, und nach Lemma 9.2.23 ist der untere Pfeil injektiv. Deswegen ist auch die Einschränkung von $f - \lambda \operatorname{id}_V$ auf Γ_{s+1} injektiv, wie gewünscht.

Nach (i) und der Definition von Γ_s erhalten wir eine direkte Zerlegung

$$\operatorname{Hau}_\lambda(f) = \bigoplus_{s=1}^{d} \Gamma_s.$$

Aus (ii) folgt, dass $f - \lambda \operatorname{id}_V$ direkte Zerlegungen von Γ_{s+1} erhält. Jedes Γ_s lässt sich also induktiv zerlegen als:

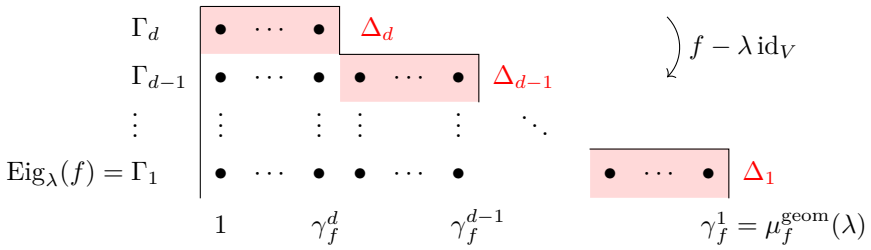
$$\begin{aligned} \Gamma_s &= \Delta_s \oplus (f - \lambda \operatorname{id}_V)(\Gamma_{s+1}) \\ &= \Delta_s \oplus (f - \lambda \operatorname{id}_V)(\Delta_{s+1}) \oplus (f - \lambda \operatorname{id}_V)^2(\Gamma_{s+2}) \\ &= \cdots = \bigoplus_{i=0}^{d-s} (f - \lambda \operatorname{id}_V)^i(\Delta_{s+i}). \end{aligned}$$

Insgesamt erhalten wir die direkte Zerelgung

$$\operatorname{Hau}_\lambda(f) = \bigoplus_{s=1}^{d} \bigoplus_{i=0}^{d-s} (f - \lambda \operatorname{id}_V)^i(\Delta_{s+i}) = \bigoplus_{t=1}^{d} \bigoplus_{i=0}^{t-1} (f - \lambda \operatorname{id}_V)^i(\Delta_t).$$

Aus (ii) folgt außerdem, dass für alle $i \leq t-1$ die Abbildung $(f - \lambda \operatorname{id}_V)^i$ Basen von Δ_t auf Basen von $(f - \lambda \operatorname{id}_V)^i(\Delta_t)$ abbildet. Damit ist die Zusammensetzung der Jordanketten B_v mit $v \in Q_t$ eine Basis von $\bigoplus_{i=0}^{t-1} (f - \lambda \operatorname{id}_V)^i(\Delta_t)$, so dass die Zusammensetzung der Jordanketten B_v für alle $v \in Q$ eine Basis des ganzen Hauptraums $\operatorname{Hau}_\lambda(f)$ ist, wie gewünscht. Zudem bilden die letzten Vektoren der Ketten B_v eine Basis von $\Gamma_1 = \operatorname{Eig}_\lambda(f)$, so dass $|Q| = \dim_K \operatorname{Eig}_\lambda(f) = \mu_f^{\text{geom}}(\lambda)$. □

Die Basis von $\mathrm{Hau}_\lambda(f)$, die im Beweis der Proposition 9.2.24 konstruiert wurde, können wir folgendermaße darstellen:



Die Punkte im roten Bereich bilden die Teilmenge Q, und die Spalten sind die zugehörigen Jordanketten B_v mit $v \in Q$ (die den einzelnen Jordanblöcken in der Darstellungsmatrix entsprechen). Jeder Punkt im weißen Bereich ist also $f-\lambda \mathrm{id}_V$ des darüberliegenden Punktes. Die ersten s Zeilen von unten bilden eine Basis von $\mathrm{Hau}_\lambda^s(\lambda)$. Die s-te Zeile selbst ist eine Basis des Untervektorraums Γ_s, und der rote Teil davon ist genau die Basis Q_s von Δ_s. Nach Konstruktion gilt $\dim_K \Gamma_s = \gamma_f^s(\lambda)$ und $\dim_K \Delta_s = \delta_f^s(\lambda)$, was das folgende Korollar liefert:

Korollar 9.2.25 (Anzahl der Jordanblöcke)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein trigonalisierbarer Endomorphismus. Sei $\lambda \in K$ und $n \in \mathbb{N}\backslash\{0\}$. Die Anzahl der Jordanblöcke zu λ der Größe n in einer Jordanschen Normalform von f ist genau*

$$\delta_f^n(\lambda) = \gamma_f^n(\lambda) - \gamma_f^{n+1}(\lambda) = 2\mu_f^n(\lambda) - \mu_f^{n+1}(\lambda) - \mu_f^{n-1}(\lambda).$$

Eine allgemeine Methode zur Bestimmung der Jordanschen Normalform einer Matrix haben wir bereits gesehen (Rezept 9.2.13). Das Korollar 9.2.25 liefert aber eine verschiedene Methode auf Basis des Gaußschen Eliminationsverfahrens:

Rezept 9.2.26 (Bestimmung der Jordanschen Normalform mit dem Gaußschen Verfahren)**.** Gegeben sei $A \in M_n(K)$ eine trigonalisierbare Matrix. Gesucht ist die Jordansche Normalform von A. Mit dem Gaußschen Eliminationsverfahren (genauer mit dem Rezept 5.2.14) bestimmt man die verallgemeinerten geometrischen Vielfachheiten $\mu_A^n(\lambda)$ jedes Eigenwerts λ von A für alle $n = 1, 2, \ldots$, bis $\mu_A^n(\lambda)$ gleich $\mu_A^{\mathrm{alg}}(\lambda)$ ist. Nach Korollar 9.2.25 ist dann $2\mu_f^n(\lambda) - \mu_f^{n+1}(\lambda) - \mu_f^{n-1}(\lambda)$ die Anzahl der Jordanblöcke zu λ der Größe n, was die Jordansche Normalform eindeutig (bis auf die Reihenfolge der Jordanblöcke) bestimmt.

Rezept 9.2.27 (Bestimmung einer Jordan-Basis)**.** Der Beweis von Proposition 9.2.24 liefert einen Algorithmus, um eine Jordan-Basis für eine trigonalisierbare Matrix $A \in M_n(K)$ zu finden. Man führt nämlich die folgenden Schritte mit jedem Eigenwert λ von A durch:

- Mit dem Gaußschen Eliminationsverfahren bestimmt man eine Basis H_s der Lösungsmenge $\mathrm{Hau}_\lambda^s(A) = \mathcal{L}((\lambda I_n - A)^s, 0)$ für alle $s \in \{1, \ldots, d\}$, wobei d die kleinste Zahl mit $|H_d| = \mu_A^{\mathrm{alg}}(\lambda)$ ist.
- Man bestimmt dann Teilmengen $Q_s \subset \mathrm{Hau}_\lambda(A)$ induktiv über $s \in \{1, \ldots, d\}$, beginnend mit $s = d$, so dass Q_s eine Basis eines komplementären Untervektorraums zu $\mathrm{Span}_K(H_{s-1} \cup \bigcup_{i=1}^{d-s}(A - \lambda I_n)^i(Q_{s+i}))$ in $\mathrm{Hau}_\lambda^s(A)$ ist. Dazu verwendet man das Rezept 5.2.29 mit der Basis H_s von $\mathrm{Hau}_\lambda^s(A)$ anstelle von der Standardbasis von K^n.

Die Jordanketten B_v mit $v \in \bigcup_{s=1}^d Q_s$ bilden dann eine Jordan-Basis von $\mathrm{Hau}_\lambda(A)$. Schließlich setzt man diese Basen für alle Eigenwerte λ zusammen, um eine Jordan-Basis $B = (v_1, \ldots, v_n)$ für A zu erhalten. Ist $S \in \mathrm{GL}_n(K)$ die Basiswechselmatrix $T_E^B = \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}$, so gilt $[L_A]_B^B = S^{-1}AS$ nach der Basiswechselformel, so dass die Matrix $S^{-1}AS$ in Jordanscher Normalform ist.

Beispiel 9.2.28. Sei

$$A = \begin{pmatrix} 3 & 0 & 1 \\ -2 & 2 & -2 \\ -1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Im Beispiel 9.2.17 haben wir bereits $\chi_A = (T-2)^3$ berechnet, so dass A trigonalisierbar ist. Wir suchen jetzt eine Jordan-Basis für A mit Rezept 9.2.27.

- *Basis von* $\mathrm{Hau}_2^1(A) = \mathrm{Eig}_2(A)$.

$$A - 2I_3 = \begin{pmatrix} 1 & 0 & 1 \\ -2 & 0 & -2 \\ -1 & 0 & -1 \end{pmatrix} \xrightarrow[A_{31}(1)]{A_{21}(2)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies \mathrm{Eig}_2(A) = \mathrm{Span}_{\mathbb{Q}}\{e_1 - e_3, e_2\}.$$

- *Basis von* $\mathrm{Hau}_2^2(A)$. Da $(A-2I_3)^2 = 0$ gilt $\mathrm{Hau}_2^2(A) = \mathrm{Hau}_2(A) = \mathbb{Q}^3$.

- *Basis von* $\Delta_2 = \Gamma_2$. Dies soll ein komplementärer Untervektorraum zu $\mathrm{Eig}_2(A)$ in $\mathbb{Q}^3$ sein, z.B. $\Delta_2 = \mathrm{Span}_{\mathbb{Q}}\{e_1\}$. Man setzt $Q_2 = \{e_1\}$.

- *Basis von* Δ_1. Dies soll ein komplementärer Untervektorraum zu $L_{A-2I_3}(\Gamma_2)$ in $\mathrm{Eig}_2(A)$ sein. Es gilt

$$L_{A-2I_3}(\Gamma_2) = \mathrm{Span}_{\mathbb{Q}}\{v\}, \quad \text{wobei} \quad v = (A-2I_3)e_1 = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}.$$

  Zur Bestimmung von Δ_1 wenden wir das Rezept 5.2.29 mit der obigen Basis von $\mathrm{Eig}_2(A)$ an:

$$\begin{pmatrix} v & e_1 - e_3 & e_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \xrightarrow[A_{31}(1)]{A_{21}(2)} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

  Die Matrix ist jetzt in Zeilenstufenform und die Pivotspalten sind die ersten zwei Spalten, so dass $(v, e_1 - e_3)$ eine Basis von $\mathrm{Eig}_2(A)$ ist. Damit hat $\Delta_1 = \mathrm{Span}_{\mathbb{Q}}\{e_1 - e_3\}$ die gewünschte Eigenschaft. Man setzt $Q_1 = \{e_1 - e_3\}$.

Also erhalten wir $Q = Q_2 \cup Q_1 = \{e_1, e_1 - e_3\}$, mit zugehörigen Jordanketten $B_{e_1} = (e_1, v)$ und $B_{e_1-e_3} = (e_1 - e_3)$. Damit ist $B = (e_1, v, e_1 - e_3)$ eine Jordan-Basis für A, und es gilt

$$S^{-1}AS = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \text{wobei} \quad S = \begin{pmatrix} e_1 & v & e_1 - e_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & -1 & -1 \end{pmatrix}.$$

Bemerkung 9.2.29. Durch explizite Beispiele kann man beobachten, dass nicht alle Basen von $\mathrm{Eig}_\lambda(f)$ zu einer Jordan-Basis von $\mathrm{Hau}_\lambda(f)$ ergänzt werden können. Deshalb ist es im Rezept 9.2.27 wirklich notwending, die Teilmengen Q_s von der höchsten zu der niedrigsten Stufe zu bestimmen, was leider recht rechenintensiv sein kann. Zum Beispiel liefert dieses Rezept *zwei* Basen jedes Untervektorraums $\mathrm{Hau}_\lambda^s(A)$, die Basis H_s sowie die entsprechende Teilfamilie der erhaltenen Jordan-Basis.

9.2.4 Die Jordan-Chevalley-Zerlegung

Definition 9.2.30 (nilpotentes Element)**.** Sei R ein Ring. Ein Element $r \in R$ heißt *nilpotent*, wenn ein $n \in \mathbb{N}$ mit $r^n = 0$ existiert.

Beispiel 9.2.31.

(i) In einem Integritätsring ist 0 das einzige nilpotente Element.

(ii) Im Ring $\mathbb{Z}/4\mathbb{Z}$ ist $[2]$ nilpotent, denn $[2]^2 = [4] = 0$.

(iii) Im Matrizenring $M_n(R)$ ist jede Matrix der Gestalt

$$A = \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} \quad \text{bzw.} \quad A = \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ * & & 0 \end{pmatrix}$$

nilpotent: Es gilt $A^n = 0$.

Eine Matrix $J \in M_n(K)$ in Jordanscher Normalform lässt sich als $J = D + N$ zerlegen, wobei

$$D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \quad \text{und} \quad N = \begin{pmatrix} 0 & & & 0 \\ \varepsilon_1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & \varepsilon_{n-1} & 0 \end{pmatrix}$$

mit $\lambda_i \in K$ und $\varepsilon_i \in \{0, 1\}$. Insbesondere ist N nilpotent mit $N^n = 0$ (eigentlich gilt bereits $N^d = 0$, wobei d die Größe des größten Jordanblocks ist). Diese Zerlegung hat viele Anwendungen, weil beide Diagonalmatrizen und nilpotente Matrizen besonders einfach sind. Die Jordan-Chevalley-Zerlegung ist eine kanonische Verallgemeinerung dieser Zerlegung auf beliebige trigonalisierbare Endomorphismen:

Satz 9.2.32 (Jordan-Chevalley-Zerlegung)**.** *Sei K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$ ein trigonalisierbarer Endomorphismus. Dann gibt es eine Zerlegung*

$$f = f_{\mathrm{diag}} + f_{\mathrm{nil}}$$

in $\mathrm{End}_K(V)$, wobei f_{diag} ein diagonalisierbarer und f_{nil} ein nilpotenter Endomorphismus sind, die miteinander kommutieren, d.h., $f_{\mathrm{diag}} \circ f_{\mathrm{nil}} = f_{\mathrm{nil}} \circ f_{\mathrm{diag}}$. Außerdem sind f_{diag} und f_{nil} eindeutig durch f bestimmt.

Beweis. Zur Existenz. Der Jordanblock $J_e(\lambda)$ besitzt eine solche Zerlegung:

$$J_e(\lambda) = \lambda I_e + J_e(0),$$

denn $J_e(0)^e = 0$ und $\lambda I_e \in M_e(K)$ ist zentral. Nach Korollar 9.2.15 gibt es eine Basis B von V, so dass die Matrix $A = [f]_B^B$ in Jordanscher Normalform ist. Wendet man die obige Zerlegung auf jeden Jordanblock von A an, so erhält man eine Zerlegung $A = A_{\mathrm{diag}} + A_{\mathrm{nil}}$ mit den gewünschten Eigenschaften. Durch den Ringisomorphismus $\mathrm{End}_K(V) \xrightarrow{\sim} M_{\dim(V)}(K)$, $g \mapsto [g]_B^B$, erhalten wir schließlich eine entsprechende Zerlegung von f.

Zur Eindeutigkeit. Da der Endomorphismus f_{diag} diagonalisierbar ist, ist er durch seine Eigenräume eindeutig bestimmt (denn $f_{\mathrm{diag}}(v) = \lambda v$ für alle $v \in \mathrm{Eig}_\lambda(f_{\mathrm{diag}})$ und V ist die Summe der Eigenräume von f_{diag}). Wir behaupten, dass für jedes $\lambda \in K$ gilt

$$\mathrm{Eig}_\lambda(f_{\mathrm{diag}}) = \mathrm{Hau}_\lambda(f).$$

Dies impliziert, dass f_{diag} (und somit auch $f_{\mathrm{nil}} = f - f_{\mathrm{diag}}$) eindeutig durch f bestimmt ist. Da f trigonalisierbar und f_{diag} diagonalisierbar ist, gibt es Isomorphismen

$$\bigoplus_{\lambda \in K} \mathrm{Hau}_\lambda(f) \xrightarrow{\sim} V \xleftarrow{\sim} \bigoplus_{\lambda \in K} \mathrm{Eig}_\lambda(f_{\mathrm{diag}}).$$

Deswegen genügt es die Inklusion $\mathrm{Eig}_\lambda(f_{\mathrm{diag}}) \subset \mathrm{Hau}_\lambda(f)$ nachzuprüfen. Ist $v \in \mathrm{Eig}_\lambda(f_{\mathrm{diag}})$, so gilt

$$(f - \lambda \,\mathrm{id}_V)(v) = f(v) - \lambda v = f(v) - f_{\mathrm{diag}}(v) = f_{\mathrm{nil}}(v).$$

Außerdem liegt $f_{\mathrm{nil}}(v)$ wieder in $\mathrm{Eig}_\lambda(f_{\mathrm{diag}})$, denn

$$f_{\mathrm{diag}}(f_{\mathrm{nil}}(v)) = f_{\mathrm{nil}}(f_{\mathrm{diag}}(v)) = f_{\mathrm{nil}}(\lambda v) = \lambda f_{\mathrm{nil}}(v).$$

Aus der Nilpotenz von f_{nil} folgt nun, dass v im Hauptraum $\mathrm{Hau}_\lambda(f)$ liegt, wie behauptet. □

Bemerkung 9.2.33. Die Jordan-Chevalley-Zerlegung gilt auch für Endomorphismen, die nicht trigonalisierbar sind, sofern K ein sogenannter *vollkommener* Körper ist. Dieser Begriff wird in der Vorlesung *Algebra* eingeführt. Körper der Charakteristik 0 sowie endliche Körper sind vollkommen.

Abschließend besprechen wir ein paar konkrete Anwendungen der Jordan-Chevalley-Zerlegung.

Bemerkung 9.2.34 (Potenzen einer trigonalisierbaren Matrix)**.** Sei $A \in M_n(K)$ eine trigonalisierbare Matrix. Mithilfe der Jordanschen Normalform kann man eine einheitliche Formel für alle Potenzen A^k erhalten (siehe Bemerkung 6.2.32 für den diagonaliserbaren Fall). Sei $S \in \mathrm{GL}_n(K)$, so dass $S^{-1}AS = J$ in Jordanscher Normalform ist. Dann gilt

$$A^k = SJ^kS^{-1}.$$

Es genügt also J^k zu berechnen. Dazu schreibt man $J = D + N$ mit D diagonal, $N^n = 0$ und $DN = ND$. Da D und N miteinander kommutieren gilt der binomische Lehrsatz:

$$J^k = (D+N)^k = \sum_{i=0}^{k} \binom{k}{i} D^{k-i}N^i = \sum_{i=0}^{\min\{k,n-1\}} \binom{k}{i} D^{k-i}N^i.$$

Außerdem werden die Potenzen der Diagonalmatrix D komponentenweise berechnet. Man braucht also nur die (endlich vielen) Potenzen von N zu berechnen, um eine Formel für beliebige Potenzen von A zu erhalten.

Bemerkung 9.2.35. Eine reelle Matrix $A \in M_n(\mathbb{R})$ ist nicht unbedingt trigonalisierbar. Man kann trotzdem das Verfahren aus Bemerkung 9.2.34 verwenden, indem man A als komplexe Matrix betrachtet. Die Potenzen A^k sind natürlich wieder reelle Matrizen, aber die Hilfsmatrizen J und S können komplexe Zahlen enthalten. Allgemeiner wird jede Matrix $A \in M_n(K)$ trigonalisierbar, wenn man K durch einen algebraischen Abschluss $\bar{K}$ ersetzt (siehe Bemerkung 6.4.18).

Bemerkung 9.2.36 (Exponentialfunktion für Matrizen)**.** Die Funktion $\exp\colon \mathbb{C} \to \mathbb{C} \setminus \{0\}$ lässt sich auf quadratische Matrizen verallgemeinern. Sie ist tatsächlich der Sonderfall $n = 1$ einer Abbildung

$$\begin{aligned} \exp\colon M_n(\mathbb{C}) &\to \mathrm{GL}_n(\mathbb{C}), \\ A &\mapsto \sum_{k=0}^{\infty} \frac{A^k}{k!}. \end{aligned}$$

Um diese Reihe zu verstehen kann man $M_n(\mathbb{C})$ mit dem Skalarprodukt $\langle A, B\rangle = \mathrm{tr}(A^{\mathsf{H}}B)$ versehen, das $M_n(\mathbb{C})$ zu einem metrischen Raum befördert. Man kann dann zeigen wie im Fall $n = 1$, dass die Reihe für alle A konvergiert. Die üblichen Beweise zeigen zudem:

(i) Es gilt $\exp(0_n) = I_n$.

(ii) Falls $AB = BA$ gilt $\exp(A+B) = \exp(A)\exp(B)$ (dies gilt aber nicht im Allgemeinen).

Diese Aussagen implizieren übrigens, dass $\exp(A)$ invertierbar ist, mit $\exp(A)^{-1} = \exp(-A)$.

Die Jordansche Normalform ist sehr hilfreich bei der Berechnung der Exponentialfunktion von Matrizen. Denn sei $S \in \mathrm{GL}_n(\mathbb{C})$, so dass $S^{-1}AS = J$ in Jordanscher Normalform ist, und sei $J = D + N$ mit D diagonal, $N^n = 0$ und $DN = ND$. Aus den Gleichungen $(SJS^{-1})^k = SJ^kS^{-1}$ für alle $k \in \mathbb{N}$ folgt die Gleichung $\exp(SJS^{-1}) = S\exp(J)S^{-1}$. Damit gilt

$$\exp(A) = \exp(SJS^{-1}) = S\exp(J)S^{-1} = S\exp(D)\exp(N)S^{-1}.$$

Für eine Diagonalmatrix wird die Exponentialfunktion komponentweise berechnet (da dies für die einzelnen Potenzen gilt):

$$\exp(\operatorname{diag}(\lambda_1, \dots, \lambda_n)) = \operatorname{diag}(e^{\lambda_1}, \dots, e^{\lambda_n}).$$

Schließlich ist $\exp(N)$ gleich der *endlichen* Summe $\sum_{i=0}^{n-1} \frac{N^k}{k!}$.

Beispiel 9.2.37.

(i) Sei $A = \begin{pmatrix} \lambda & 0 \\ \alpha & \lambda \end{pmatrix} = \lambda I_2 + N_\alpha$. Dann gilt

$$\exp(\lambda I_2) = \begin{pmatrix} e^\lambda & 0 \\ 0 & e^\lambda \end{pmatrix} \quad \text{und} \quad \exp(N_\alpha) = \sum_{k=0}^{\infty} \frac{N_\alpha^k}{k!} = I_2 + N_\alpha = \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix},$$

und damit

$$\exp(A) = \exp(\lambda I_2) \exp(N_\alpha) = \begin{pmatrix} e^\lambda & 0 \\ e^\lambda \alpha & e^\lambda \end{pmatrix}.$$

(ii) Sei

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix} \in M_2(\mathbb{C}).$$

Es gilt $\chi_A = (T-1)^2$ und $\operatorname{Eig}_1(A) = \operatorname{Span}_{\mathbb{C}}\{e_2 - e_1\}$. Zudem gilt $(A - I_2)e_1 = e_2 - e_1$, so dass e_1 ein Hauptvektor zu 1 der Stufe 2 ist. Damit erhalten wir die Jordan-Basis $(e_1, e_2 - e_1)$ für A, so dass

$$S^{-1}AS = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \text{wobei} \quad S = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad S^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Nach (i) gilt nun

$$\exp(A) = S \exp \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} S^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e & 0 \\ e & e \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -e \\ e & 2e \end{pmatrix}.$$

(iii) Sei

$$A = \begin{pmatrix} 3 & 0 & 1 \\ -2 & 2 & -2 \\ -1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{C}).$$

Nach Beispiel 9.2.28 gilt

$$S^{-1}AS = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 2I_3 + N, \quad \text{wobei} \quad S = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & -1 & -1 \end{pmatrix}.$$

Wir berechnen noch die inverse Matrix S^{-1}:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 0 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[V_{23}]{M_3(-1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & -2 & 0 & 0 & 1 & 0 \end{array}\right)$$

$$\xrightarrow{A_{32}(2)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 2 & 0 & 1 & -2 \end{array}\right) \xrightarrow{M_3(\frac{1}{2})} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & -1 \end{array}\right)$$

$$\xrightarrow[A_{13}(-1)]{A_{23}(-1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & -\frac{1}{2} & 1 \\ 0 & 1 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & -1 \end{array}\right) \xrightarrow{A_{12}(-1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & -1 \end{array}\right).$$

Aus $N^2 = 0$ folgt $\exp(N) = I_3 + N$. Schließlich erhalten wir

$$\begin{aligned}\exp(A) &= S\exp(2I_3)\exp(N)S^{-1}\\ &= \begin{pmatrix}1&1&1\\0&-2&0\\0&-1&-1\end{pmatrix}\begin{pmatrix}e^2&0&0\\e^2&e^2&0\\0&0&e^2\end{pmatrix}\begin{pmatrix}1&0&1\\0&-\frac{1}{2}&0\\0&\frac{1}{2}&-1\end{pmatrix}\\ &= \begin{pmatrix}2e^2&e^2&e^2\\-2e^2&-2e^2&0\\-e^2&-e^2&-e^2\end{pmatrix}\begin{pmatrix}1&0&1\\0&-\frac{1}{2}&0\\0&\frac{1}{2}&-1\end{pmatrix} = \begin{pmatrix}2e^2&0&e^2\\-2e^2&e^2&-2e^2\\-e^2&0&0\end{pmatrix}.\end{aligned}$$

Bemerkung 9.2.38 (lineare Differentialgleichungssysteme)**.** Sei $A \in M_n(\mathbb{C})$ eine Matrix und sei $x_0 \in \mathbb{C}^n$ beliebig. Wir betrachten das homogene lineare Differentialgleichungssystem erster Ordnung $x' = Ax$ unter der Anfangsbedingung $x(0) = x_0$. Das heißt, wir suchen alle differenzierbaren Funktionen $x\colon \mathbb{R} \to \mathbb{C}^n$, so dass

$$x'(t) = A \cdot x(t) \quad \text{für alle } t \in \mathbb{R}, \quad \text{und} \quad x(0) = x_0.$$

Das Matrixexponential definiert eine Funktion $\mathbb{R} \to M_n(\mathbb{C})$, $t \mapsto \exp(At)$, für die gilt

$$\frac{d}{dt}\exp(At) = A \cdot \exp(At).$$

Daraus folgt leicht, dass die Funktion $x(t) = \exp(At)\cdot x_0$ eine Lösung unseres Gleichungssystems ist. Sie ist eigentlich die einzige Lösung unter der gegebenen Anfangsbedingung, denn: Sei $x\colon \mathbb{R} \to \mathbb{C}^n$ eine beliebige Lösung mit $x(0) = x_0$. Dann gilt

$$\begin{aligned}\frac{d}{dt}(\exp(-At)x(t)) &= \exp(-At)x'(t) - A\exp(-At)x(t)\\ &= \exp(-At)Ax(t) - A\exp(-At)x(t) = 0,\end{aligned}$$

da A und $\exp(-At)$ kommutieren (da A mit allen ihrer Potenzen kommutiert). Also ist $t \mapsto \exp(-At)x(t)$ eine konstante Funktion, deren Wert in $t = 0$ gleich $x(0) = x_0$ ist, so dass für alle $t \in \mathbb{R}$ gilt $\exp(-At)x(t) = x_0$, d.h., $x(t) = \exp(At)x_0$.

Sei $\mathcal{L}_A = \{x \in \mathrm{Diff}(\mathbb{R}, \mathbb{C}^n) | x' = Ax\}$ die Lösungsmenge des Differentialgleichungssystems ohne Anfangsbedingung. Dann ist $\mathcal{L}_A$ ein Untervektorraum des $\mathbb{C}$-Vektorraums $\mathrm{Diff}(\mathbb{R}, \mathbb{C}^n)$, und die $\mathbb{C}$-lineare Abbildung

$$\mathcal{L}_A \to \mathbb{C}^n, \quad x \mapsto x(0),$$

ist ein Isomorphismus mit Umkehrabbildung $x_0 \mapsto (t \mapsto \exp(At) \cdot x_0)$. Insbesondere bilden die Spalten von $\exp(At)$ eine Basis von $\mathcal{L}_A$.

Beispiel 9.2.39. Wir lösen das lineare Differentialgleichungssystem

$$\begin{aligned}x_1' &= 3x_1 + x_3,\\ x_2' &= -2x_1 + 2x_2 - 2x_3,\\ x_3' &= -x_1 + x_3.\end{aligned}$$

Die allgemeine Lösung ist $x(t) = \exp(At)x_0$ mit $x_0 \in \mathbb{C}^n$ beliebig, wobei A die Koeffizientenmatrix ist. Die Matrix $\exp(A)$ haben wir schon im Beispiel 9.2.37(iii) berechnet, und die Berechnung von $\exp(At)$ ist ganz ähnlich: Es gilt $\exp(Nt) = I_3 + Nt$ und daher

$$\begin{aligned}\exp(At) &= S\exp(2I_3t)\exp(Nt)S^{-1}\\ &= \begin{pmatrix}1&1&1\\0&-2&0\\0&-1&-1\end{pmatrix}\begin{pmatrix}e^{2t}&0&0\\te^{2t}&e^{2t}&0\\0&0&e^{2t}\end{pmatrix}\begin{pmatrix}1&0&1\\0&-\frac{1}{2}&0\\0&\frac{1}{2}&-1\end{pmatrix}\\ &= \begin{pmatrix}(1+t)e^{2t}&e^{2t}&e^{2t}\\-2te^{2t}&-2e^{2t}&0\\-te^{2t}&-e^{2t}&-e^{2t}\end{pmatrix}\begin{pmatrix}1&0&1\\0&-\frac{1}{2}&0\\0&\frac{1}{2}&-1\end{pmatrix} = \begin{pmatrix}(1+t)e^{2t}&0&te^{2t}\\-2te^{2t}&e^{2t}&-2te^{2t}\\-te^{2t}&0&(1-t)e^{2t}\end{pmatrix}.\end{aligned}$$

Die Spalten von $\exp(At)$ bilden dann eine Basis der Lösungsmenge.

Kapitel 10

Multilineare Algebra

Die multilineare Algebra befasst sich mit *multilinearen* Abbildungen. Multilineare Abbildungen zwischen Vektorräumen haben wir bereits definiert (Definition 5.3.11), und die Definition bleibt sinnvoll für Moduln über einem beliebigen kommutativen Ring R.

Das *Tensorprodukt* von R-Moduln ist eine Konstruktion, die multilineare Abbildungen in *lineare* Abbildungen umwandelt. Sind genauer $M_1, \dots, M_n, N$ Moduln über R, so gibt es eine Bijektion zwischen n-linearen Abbildungen

$$M_1 \times \cdots \times M_n \to N$$

und linearen Abbildungen aus dem Tensorprodukt

$$M_1 \otimes_R \cdots \otimes_R M_n \to N.$$

Elemente von $M_1 \otimes_R \cdots \otimes_R M_n$ heißen *Tensoren*. Die *symmetrische Potenz* $\mathrm{Sym}^n_R(M)$ und die *äußere Potenz* $\Lambda^n_R(M)$ spielen eine ähnliche Rolle wie das Tensorprodukt aber für symmetrische und alternierende n-lineare Abbildungen $M^n \to N$.

Die direkte Summe $\oplus$ und das Tensorprodukt $\otimes_R$ von R-Moduln verhalten sich ganz ähnlich wie die Addition $+$ und die Multiplikation $\cdot$ von Zahlen: Beide $\oplus$ und $\otimes_R$ sind assoziativ und kommutativ (bis auf Isomorphie) und besitzen neutrale Elemente. Außerdem gilt das Distributivgesetz: $M \otimes_R (N \oplus P) \cong (M \otimes_R N) \oplus (M \otimes_R P)$. Für Vektorräume U, V über einem Körper K gilt zudem

$$\begin{aligned} \dim_K(U \oplus V) &= \dim_K(U) + \dim_K(V), \\ \dim_K(U \otimes_K V) &= \dim_K(U) \cdot \dim_K(V). \end{aligned}$$

Historisch gesehen wurden Tensoren als gemeinsame Verallgemeinerung von Skalaren, Vektoren und Matrizen eingeführt. Nämlich, wie wir bereits im Kapitel 7 bemerkt haben, können wir Bilinearformen $R^k \times R^l \to R$ mit $k \times l$-Matrizen über R identifizieren. Für $n \in \mathbb{N}$ beliebig können wir auf ähnliche Weise eine n-lineare Form

$$R^{k_1} \times \cdots \times R^{k_n} \to R$$

mit einer „n-dimensionale Matrix“ der Größe $k_1 \times \cdots \times k_n$ darstellen. Nennt man ein solches Objekt *Tensor* der Stufe n, so gilt:

- Ein Tensor nullter Stufe ist ein Skalar (denn eine 0-lineare Form ist eine beliebige Abbildung $\{*\} \to R$).
- Ein Tensor erster Stufe ist ein Vektor.
- Ein Tensor zweiter Stufe ist eine Matrix.

In vielen Situationen ist aber diese konkrete Definition nicht hinreichend, denn nicht alle R-Moduln sind frei mit einer bevorzugten Basis.

Das Tensorprodukt ist ein grundlegender Begriff der Linearen Algebra, der in allen Bereichen der Mathematik sowie in der Physik häufig benutzt wird. Tensoren wurden tatsächlich durch Einsteins Allgemeine Relativitätstheorie weithin bekannt gemacht. Tensorprodukte sind besonders wichtig in der Differentialgeometrie, die sich mit glatten Mannigfaltigkeiten beschäftigt. Zum Beispiel ist eine Riemannsche oder Lorentzsche Metrik ein Schnitt der zweiten symmetrischen Potenz des Kotangentialbündels einer glatten Mannigfaltigkeit. Äußere Potenzen sind auch wichtig in der Integrationstheorie über Mannigfaltigkeiten. Die übliche Notation „$dx_1 \dots dx_n$“ in einem Integral über eine Teilmenge von $\mathbb{R}^n$ steht eigentlich für eine Differentialform n-ten Grades über $\mathbb{R}^n$, die jedem Punkt von $\mathbb{R}^n$ ein Element der dualen äußeren Potenz $\Lambda^n_{\mathbb{R}}(\mathbb{R}^n)^*$ zuordnet.

10.1 Das Tensorprodukt

Wir erinnern an die Definition von multilinearen Abbildungen (Definition 5.3.11):

Definition 10.1.1 (multilineare Abbildung)**.** Sei R ein kommutativer Ring und seien M_1, $\dots, M_n, N$ Moduln über R. Eine Abbildung

$$f\colon M_1 \times \cdots \times M_n \to N$$

heißt (R-)*multilinear* oder (R-)n*-linear*, wenn sie bezüglich jedes ihrer n Argumente R-linear ist, d.h.: Für jedes $i \in \{1, \dots, n\}$ und jedes

$$(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in M_1 \times \cdots \times M_{i-1} \times M_{i+1} \times \cdots \times M_n$$

ist die folgende Abbildung R-linear:

$$\begin{aligned} M_i &\to N, \\ x &\mapsto f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n). \end{aligned}$$

Notation 10.1.2. Man schreibt

$$\mathrm{Hom}^n_R(M_1, \dots, M_n, N) \subset \mathrm{Abb}(M_1 \times \cdots \times M_n, N)$$

für den Untermodul von n-linearen Abbildungen $M_1 \times \cdots \times M_n \to N$.

Proposition 10.1.3 (Exponentialgesetz für multilineare Abbildungen)**.** *Sei R ein kommutativer Ring, seien $m, n \in \mathbb{N}$ und seien $M_1, \dots, M_m, N_1, \dots, N_n, P$ Moduln über R. Dann gibt es einen Isomorphismus von R-Moduln*

$$\begin{aligned} \mathrm{Hom}^{m+n}_R(M_1, \dots, M_m, N_1, \dots, N_n, P) &\xrightarrow{\sim} \mathrm{Hom}^m_R(M_1, \dots, M_m, \mathrm{Hom}^n_R(N_1, \dots, N_n, P)), \\ f &\mapsto (x \mapsto (y \mapsto f(x, y))). \end{aligned}$$

Beweis. Da f in jedem seiner $m + n$ Argumente linear ist, ist die Abbildung $x \mapsto f(x, -)$ m-linear, und für jedes feste $x \in M_1 \times \cdots \times M_m$ ist die Abbildung $y \mapsto f(x, y)$ n-linear. Das heißt, die gegebene Abbildung ist wohldefiniert. Sie ist auch R-linear, da die Modulstruktur auf beiden Seiten punktweise definiert ist. Die Umkehrabbildung schickt g auf die Abbildung $(x, y) \mapsto g(x)(y)$. □

Bemerkung 10.1.4. Der Sonderfall $m = n = 1$ der Proposition 10.1.3 haben wir schon im Kapitel 7 verwendet: Der Isomorphismus

$$\mathrm{Bil}_K(V, W) = \mathrm{Hom}^2_K(V, W, K) \xrightarrow{\sim} \mathrm{Hom}_K(V, \mathrm{Hom}_K(W, K)) = \mathrm{Hom}_K(V, W^*)$$

schickt eine Bilinearform $b\colon V \times W \to K$ auf die lineare Abbildung $b_l\colon V \to W^*$.

Definition 10.1.5 (Tensorprodukt)**.** Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$ und seien $M_1, \dots, M_n$ Moduln über R. Ein *Tensorprodukt* der Familie $(M_1, \dots, M_n)$ ist ein Paar (T, τ) bestehend aus einem R-Modul T und einer n-linearen Abbildung

$$\tau \colon M_1 \times \cdots \times M_n \to T$$

mit folgender universellen Eigenschaft: Zu jedem R-Modul N und jeder n-lineare Abbildung $f \colon M_1 \times \cdots \times M_n \to N$ gibt es *genau eine* lineare Abbildung $g \colon T \to N$ mit $g \circ \tau = f$:

$$\begin{array}{ccc} M_1 \times \cdots \times M_n & \xrightarrow{f} & N. \\ {\scriptstyle\tau}\big\downarrow & \nearrow_{\exists! g} & \\ T & & \end{array}$$

Die universelle Eigenschaft eines Tensorprodukts (T, τ) kann man auch folgendermaße formulieren: Es gibt eine Bijektion

$$\begin{aligned} \operatorname{Hom}_R(T, M) &\xrightarrow{\sim} \operatorname{Hom}_R^n(M_1, \dots, M_n, N), \\ g &\mapsto g \circ \tau. \end{aligned}$$

(Diese Abbildung ist zudem R-linear und damit ein Isomorphismus von R-Moduln.)

Satz 10.1.6 (Existenz und Eindeutigkeit des Tensorprodukts)**.** *Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$ und seien $M_1, \dots, M_n$ Moduln über R.*

(i) *Ein Tensorprodukt (T, τ) von $(M_1, \dots, M_n)$ existiert.*

(ii) *Seien (T, τ) und (T', τ') zwei Tensorprodukte von $(M_1, \dots, M_n)$. Dann gibt es genau eine lineare Abbildung $\varphi \colon T \to T'$ mit $\varphi \circ \tau = \tau'$. Außerdem ist φ ein Isomorphismus.*

Beweis. Zu (i). Sei $F := R^{(M_1 \times \cdots \times M_n)}$. Für ein n-Tupel $(x_1, \dots, x_n) \in M_1 \times \cdots \times M_n$ bezeichnen wir mit $[x_1, \dots, x_n]$ den entsprechenden Standardeinheitsvektor in F. Nach der universellen Eigenschaft von Basen gibt es zu jedem R-Modul N eine Bijektion

$$\operatorname{Abb}(M_1 \times \cdots \times M_n, N) \xrightarrow{\sim} \operatorname{Hom}_R(F, N), \quad f \mapsto \hat{f},$$

wobei $\hat{f}([x_1, \dots, x_n]) = f(x_1, \dots, x_n)$. Sei $E \subset F$ die Teilmenge bestehend aus allen Elementen der Gestalt

$$\begin{gathered} [x_1, \dots, x_i + x_i', \dots, x_n] - [x_1, \dots, x_i, \dots, x_n] - [x_1, \dots, x_i', \dots, x_n], \\ [x_1, \dots, rx_i, \dots, x_n] - r \cdot [x_1, \dots, x_i, \dots, x_n], \end{gathered}$$

mit $i \in \{1, \dots, n\}$ beliebig. Dann ist eine Abbildung $f \colon M_1 \times \cdots \times M_n \to N$ genau dann n-linear, wenn $E \subset \ker \hat{f}$. Da $\ker \hat{f}$ ein Untermodul von F ist, ist die letzte Bedingung auch äquivalent zu $\operatorname{Span}_R(E) \subset \ker \hat{f}$. Sei nun T der Quotientenmodul $F/\operatorname{Span}_R(E)$, sei $q \colon F \to T$ die Quotientenabbildung und sei

$$\begin{aligned} \tau \colon M_1 \times \cdots \times M_n &\to T, \\ (x_1, \dots, x_n) &\mapsto q([x_1, \dots, x_n]). \end{aligned}$$

Da $E \subset \ker q$ ist τ n-linear. Nach der universellen Eigenschaft des Quotientenmoduls gibt es dann zu jeder n-linearen Abbildung $f \colon M_1 \times \cdots \times M_n \to N$ genau eine R-lineare Abbildung $g \colon T \to N$ mit $g \circ q = \hat{f}$. Damit ist auch g die einzige R-lineare Abbildung mit $g \circ \tau = f$, wie gewünscht. Dieses Argument lässt sich mit folgendem Diagramm zusammenfassen:

$$\begin{array}{ccccc} M_1 \times \cdots \times M_n & \xrightarrow{[-]} & F & \xrightarrow{q} & T. \\ {\scriptstyle f}\big\downarrow & \swarrow_{\exists! \hat{f}} & & \nearrow_{\exists! g} & \\ N & \longleftarrow & & & \end{array}$$

Zu (ii). Der Beweis ist ein Standardargument mit universellen Eigenschaften. Nach der universellen Eigenschaft von (T, τ) und der n-Linearität von τ' gibt es genau eine R-lineare Abbildung $\varphi\colon T \to T'$ mit $\varphi \circ \tau = \tau'$. Nach der universellen Eigenschaft von (T', τ') und der n-Linearität von τ gibt es genau eine R-lineare Abbildung $\varphi'\colon T' \to T$ mit $\varphi' \circ \tau' = \tau$. Die Komposition $\varphi' \circ \varphi\colon T \to T$ erfüllt

$$\varphi' \circ \varphi \circ \tau = \varphi' \circ \tau' = \tau.$$

Die Indentitätsabbildung id_T erfüllt auch $\mathrm{id}_T \circ \tau = \tau$. Aus der universellen Eigenschaft von (T, τ) folgt, dass $\varphi' \circ \varphi = \mathrm{id}_T$. Auf ähnliche Weise gilt $\varphi \circ \varphi' = \mathrm{id}_{T'}$, so dass φ und φ' zueinander invers sind. □

Notation 10.1.7. Wegen der wesentlichen Eindeutigkeit des Tensorprodukts schreiben wir

$$M_1 \otimes_R \cdots \otimes_R M_n$$

für „das“ Tensorprodukt der Familie $(M_1, \ldots, M_n)$. Wir schreiben weiter

$$\begin{aligned} \tau\colon M_1 \times \cdots \times M_n &\to M_1 \otimes_R \cdots \otimes_R M_n, \\ (x_1, \ldots, x_n) &\mapsto x_1 \otimes \cdots \otimes x_n \end{aligned}$$

für die universelle n-lineare Abbildung auf $M_1 \times \cdots \times M_n$.

Elemente von $M_1 \otimes_R \cdots \otimes_R M_n$ heißen *Tensoren*, und die im Bild von τ heißen *reine Tensoren* oder *Elementartensoren*. Der reine Tensor $x_1 \otimes \cdots \otimes x_n$ heißt das *Tensorprodukt* der Elemente $x_1, \ldots, x_n$.

Proposition 10.1.8 (Funktorialität des Tensorprodukts)**.** *Sei R ein kommutativer Ring, $n \in \mathbb{N}$, $M_1, \ldots, M_n$ und $N_1, \ldots, N_n$ Moduln über R und $f_i\colon M_i \to N_i$ R-lineare Abbildungen. Dann gibt es genau eine R-lineare Abbildung*

$$\begin{aligned} f_1 \otimes \cdots \otimes f_n\colon M_1 \otimes_R \cdots \otimes_R M_n &\to N_1 \otimes_R \cdots \otimes_R N_n, \\ x_1 \otimes \cdots \otimes x_n &\mapsto f_1(x_1) \otimes \cdots \otimes f_n(x_n). \end{aligned}$$

Außerdem:

(i) *Es gilt* $\mathrm{id}_{M_1} \otimes \cdots \otimes \mathrm{id}_{M_n} = \mathrm{id}_{M_1 \otimes_R \cdots \otimes_R M_n}$

(ii) *Sind $f_i\colon M_i \to N_i$ und $g_i\colon N_i \to P_i$ R-lineare Abbildungen, so gilt*

$$(g_1 \circ f_1) \otimes \cdots \otimes (g_n \circ f_n) = (g_1 \otimes \cdots \otimes g_n) \circ (f_1 \otimes \cdots \otimes f_n).$$

(iii) *Sind $f_1, \ldots, f_n$ Isomorphismen, so ist $f_1 \otimes \cdots \otimes f_n$ ein Isomorphismus.*

Beweis. Die Existenz und Eindeutigkeit von $f_1 \otimes \cdots \otimes f_n$ folgt aus der universellen Eigenschaft des Tensorprodukts $M_1 \otimes_R \cdots \otimes_R M_n$, da die Abbildung

$$\begin{aligned} M_1 \times \cdots \times M_n &\to N_1 \otimes_R \cdots \otimes_R N_n, \\ (x_1, \ldots, x_n) &\mapsto f_1(x_1) \otimes \cdots \otimes f_n(x_n), \end{aligned}$$

n-linear ist. Aussagen (i) und (ii) folgen unmittelbar aus der Eindeutigkeit der linken Seite. Aussage (iii) folgt dann aus (i) und (ii), die implizieren, dass $f_1^{-1} \otimes \cdots \otimes f_n^{-1}$ eine Umkehrabbildung zu $f_1 \otimes \cdots \otimes f_n$ ist. □

Proposition 10.1.9 (Eigenschaften des Tensorprodukts)**.** *Sei R ein kommutativer Ring.*

(i) (nulläres Tensorprodukt) $(R, 1)$ *ist ein Tensorprodukt der leeren Familie von R-Moduln.*

(ii) (unäres Tensorprodukt) *Für jeden R-Modul M ist (M, id_M) ein Tensorprodukt der einelementigen Familie (M).*

(iii) (Exponentialgesetz für das binäre Tensorprodukt) *Seien M, N, P Moduln über R. Dann gibt es einen Isomorphismus von R-Moduln*

$$\begin{aligned}\mathrm{Hom}_R(M \otimes_R N, P) &\xrightarrow{\sim} \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)),\\ f &\mapsto (x \mapsto (y \mapsto f(x \otimes y))).\end{aligned}$$

Sei nun $n \in \mathbb{N}$ und seien $M_1, \dots, M_n$ Moduln über R.

(iv) (Assoziativität) *Sei $k \in \mathbb{N}$ und seien $0 = n_0 \leq n_1 \leq \cdots \leq n_k = n$ natürliche Zahlen. Zu jedem $j \in \{1, \dots, k\}$ sei (T_j, τ_j) ein Tensorprodukt von $(M_{n_{j-1}+1}, \dots, M_{n_j})$, und sei (T, τ) ein Tensorprodukt von $(T_1, \dots, T_k)$. Dann ist $(T, \tau \circ (\tau_1 \times \cdots \times \tau_k))$ ein Tensorprodukt von $(M_1, \dots, M_n)$. Anders gesagt gibt es einen Isomorphismus*

$$\begin{aligned}M_1 \otimes_R \cdots \otimes_R M_n &\xrightarrow{\sim} (M_1 \otimes_R \cdots \otimes_R M_{n_1}) \otimes_R \cdots \otimes_R (M_{n_{k-1}+1} \otimes_R \cdots \otimes_R M_n),\\ x_1 \otimes \cdots \otimes x_n &\mapsto (x_1 \otimes \cdots \otimes x_{n_1}) \otimes \cdots \otimes (x_{n_{k-1}+1} \otimes \cdots \otimes x_n).\end{aligned}$$

(v) (Kommutativität) *Sei (T, τ) ein Tensorprodukt von $(M_1, \dots, M_n)$ und sei $\sigma \in S_n$ eine Permutation. Sei*

$$\begin{aligned}\sigma^*\colon M_1 \times \cdots \times M_n &\xrightarrow{\sim} M_{\sigma(1)} \times \cdots \times M_{\sigma(n)},\\ (x_1, \dots, x_n) &\mapsto (x_{\sigma(1)}, \dots, x_{\sigma(n)}).\end{aligned}$$

Dann ist $(T, \tau \circ (\sigma^*)^{-1})$ ein Tensorprodukt von $(M_{\sigma(1)}, \dots, M_{\sigma(n)})$. Anders gesagt gibt es einen Isomorphismus

$$\begin{aligned}M_1 \otimes_R \cdots \otimes_R M_n &\xrightarrow{\sim} M_{\sigma(1)} \otimes_R \cdots \otimes_R M_{\sigma(n)},\\ x_1 \otimes \cdots \otimes x_n &\mapsto x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(n)}.\end{aligned}$$

(vi) (Distributivität über direkte Summen) *Sei $i \in \{1, \dots, n\}$ und sei $M_i = \bigoplus_{j \in J} N_j$. Zu jedem $j \in J$ sei (T_j, τ_j) ein Tensorprodukt von $(M_1, \dots, N_j, \dots, M_n)$, wobei N_j an der i-ten Stelle ist. Sei $T = \bigoplus_{j \in J} T_j$ und sei*

$$\begin{aligned}\tau\colon M_1 \times \cdots \times M_i \times \cdots \times M_n &\to T,\\ (x_1, \dots, (y_j)_{j \in J}, \dots, x_n) &\mapsto (\tau_j(x_1, \dots, y_j, \dots, x_n))_{j \in J}.\end{aligned}$$

Dann ist (T, τ) ein Tensorprodukt von $(M_1, \dots, M_n)$. Anders gesagt gibt es einen Isomorphismus

$$\begin{aligned}M_1 \otimes_R \cdots \otimes_R \left(\bigoplus_{j \in J} N_j\right) \otimes_R \cdots \otimes_R M_n &\xrightarrow{\sim} \bigoplus_{j \in J} M_1 \otimes_R \cdots \otimes_R N_j \otimes_R \cdots \otimes_R M_n,\\ x_1 \otimes \cdots \otimes (y_j)_{j \in J} \otimes \cdots \otimes x_n &\mapsto (x_1 \otimes \cdots \otimes y_j \otimes \cdots \otimes x_n)_{j \in J}.\end{aligned}$$

Beweis. Zu (i). Ein Produkt von Mengen mit leerer Indexmenge ist eine einelementige Menge $\{*\}$. Deswegen ist eine 0-lineare Abbildung nach einem R-Modul N eine beliebige Abbildung $\{*\} \to N$, die mit einem Element von N identifiziert werden kann. Nach der universellen Eigenschaft von Basen gibt es zu jedem Element $x \in N$ genau eine lineare Abbildung $g\colon R \to N$ mit $g(1) = x$. Dies ist aber genau die Aussage, dass $(R, 1)$ ein Tensorprodukt der leeren Familie ist.

Zu (ii). Eine 1-lineare Abbildung $f\colon M \to N$ ist einfach eine lineare Abbildung. Es gibt dann genau eine lineare Abbildung $g\colon M \to N$ mit $g \circ \mathrm{id}_M = f$, nämlich $g = f$.

Zu (iii). Dies folgt aus Proposition 10.1.3: Es gibt Isomorphismen

$$\mathrm{Hom}_R(M \otimes_R N, P) \xrightarrow{\sim} \mathrm{Hom}^2_R(M, N, P) \xrightarrow{\sim} \mathrm{Hom}_R(M, \mathrm{Hom}_R(N, P)).$$

Der erste bildet f auf die bilineare Abbildung $(x, y) \mapsto f(x \otimes y)$ ab und der zweite bildet die weiter auf $x \mapsto (y \mapsto f(x \otimes y))$ ab.

Zu (iv). Sei P ein R-Modul. Nach Definition des Tensorprodukts gibt es eine Bijektion

$$\mathrm{Hom}_R(T_1 \otimes_R \cdots \otimes_R T_k, P) \xrightarrow{\sim} \mathrm{Hom}_R^k(T_1, \ldots, T_k, P), \quad g \mapsto g \circ \tau.$$

Es genügt also zu zeigen, dass die Abbildung

$$\mathrm{Hom}_R^k(T_1, \ldots, T_k, P) \to \mathrm{Hom}_R^n(M_1, \ldots, M_n, P), \quad g \mapsto g \circ (\tau_1 \times \cdots \times \tau_k),$$

bijektiv ist. Dies beweisen wir durch Induktion über k. Falls $k = 0$ ist auch $n = 0$ und die Aussage ist trivial. Sei also $k \geq 1$. Es gibt ein kommutatives Diagramm

$$\begin{array}{ccc}
\mathrm{Hom}_R(T_1, \mathrm{Hom}_R^{k-1}(T_2, \ldots, T_k, P)) & \xrightarrow{\sim} & \mathrm{Hom}_R^k(T_1, \ldots, T_k, P) \\
\downarrow & & \\
\mathrm{Hom}_R^{n_1}(M_1, \ldots, M_{n_1}, \mathrm{Hom}_R^{k-1}(T_2, \ldots, T_k, P)) & & \downarrow \\
\downarrow & & \\
\mathrm{Hom}_R^{n_1}(M_1, \ldots, M_{n_1}, \mathrm{Hom}_R^{n-n_1}(M_{n_1+1}, \ldots, M_n, P)) & \xrightarrow{\sim} & \mathrm{Hom}_R^n(M_1, \ldots, M_n, P),
\end{array}$$

wobei die horizontalen Pfeile bijektiv sind nach dem Exponentialgesetz für multilineare Abbildungen (Proposition 10.1.3). Der erste Pfeil auf der linken Seite ist bijektiv nach Definition von T_1, und der zweite ist bijektiv nach der Induktionsvoraussetzung. Damit ist der Pfeil auf der rechten Seite auch bijektiv, wie gewünscht.

Zu (v). Es ist klar, dass die Abbildung $\tau \circ (\sigma^*)^{-1}$ n-linear ist. Sei $f\colon M_{\sigma(1)} \times \cdots \times M_{\sigma(n)} \to N$ eine n-lineare Abbildung. Dann ist $f \circ \sigma^*$ auch n-linear. Damit gibt es genau eine lineare Abbildung $g\colon T \to N$ mit $g \circ \tau = f \circ \sigma^*$, d.h., es gibt genau eine lineare Abbildung $g\colon T \to N$ mit $g \circ (\tau \circ (\sigma^*)^{-1}) = f$, wie gewünscht.

Zu (vi). Dies folgt aus der Kombination der universellen Eigenschaften des Tensorprodukts und der direkten Summe. Genauer sei P ein R-Modul, sei $\iota_j\colon N_j \hookrightarrow M_i$ die kanonische Abbildung und sei

$$\begin{aligned}
\kappa_j\colon M_1 \times \cdots \times N_j \times \cdots \times M_n &\hookrightarrow M_1 \times \cdots \times M_i \times \cdots \times M_n, \\
(x_1, \ldots, y, \ldots, x_n) &\mapsto (x_1, \ldots, \iota_j(y), \ldots, x_n),
\end{aligned}$$

so dass $\tau \circ \kappa_j = \iota_j \circ \tau_j$. Daraus folgt, dass folgendes Quadrat kommutativ ist:

$$\begin{array}{ccc}
g & \longmapsto & (g \circ \iota_j)_j \\
\downarrow & & \\
g \circ \tau & \mathrm{Hom}_R^n(M_1, \ldots, M_i, \ldots, M_n, P) \to \prod_{j \in J} \mathrm{Hom}_R^n(M_1, \ldots, N_j, \ldots, M_n, P). & (g_j \circ \tau_j)_j \\
f & \longmapsto & (f \circ \kappa_j)_j
\end{array}$$

$$\begin{array}{ccc}
g & \longmapsto & (g \circ \iota_j)_j \\
\downarrow & & \\
g \circ \tau & \mathrm{Hom}_R\left(\bigoplus_{j \in J} T_j, P\right) \xrightarrow{\quad} \prod_{j \in J} \mathrm{Hom}_R(T_j, P) & (g_j)_j \\
& \downarrow \qquad\qquad \downarrow & \downarrow \\
& \mathrm{Hom}_R^n(M_1, \ldots, M_i, \ldots, M_n, P) \to \prod_{j \in J} \mathrm{Hom}_R^n(M_1, \ldots, N_j, \ldots, M_n, P). & (g_j \circ \tau_j)_j \\
& f \longmapsto (f \circ \kappa_j)_j &
\end{array}$$

Der rechte vertikale Pfeil ist bijektiv nach der universellen Eigenschaft von (T_j, τ_j), und der obere horizontale Pfeil ist bijektiv nach der universellen Eigenschaft der direkten Summe. Um zu schließen, dass der linke vertikale Pfeil bijektiv ist, bleibt es zu zeigen, dass der untere horizontale Pfeil injektiv ist (seine Surjektivität folgt bereits aus der Kommutativität des Quadrates). Jedes f ist aber durch seine Einschränkungen $f \circ \kappa_j$ bestimmt, denn jedes Element von M_i ist eine Summe von Elementen in den Bildern der Abbildungen ι_j, und f ist linear in seinem i-ten Argument. □

Bemerkung 10.1.10 (Sonderfälle der Assoziativität)**.** Seien M, N, P Moduln über R. Nach Proposition 10.1.9 (ii) und (iv) gibt es Isomorphismen

$$\begin{array}{ccccc} (M \otimes_R N) \otimes_R P & \xleftarrow{\sim} & M \otimes_R N \otimes_R P & \xrightarrow{\sim} & M \otimes_R (N \otimes_R P), \\ (x \otimes y) \otimes z & \longmapsto & x \otimes y \otimes z & \longmapsto & x \otimes (y \otimes z). \end{array}$$

die die Assoziativität „im üblichen Sinne“ des binären Tensorprodukts nachweisen. Nach Proposition 10.1.9 (i) und (iv) (angewendet mit $n = 1$ und $k = 2$) gibt es Isomorphismen

$$\begin{array}{ccccc} M \otimes_R R & \xleftarrow{\sim} & M & \xrightarrow{\sim} & R \otimes_R M, \\ x \otimes 1 & \longmapsto & x & \longmapsto & 1 \otimes x. \end{array}$$

Das heißt, der R-Modul R ist ein „neutrales Element“ bezüglich des binären Tensorprodukts.

Bemerkung 10.1.11 (Sonderfall der Kommutativität)**.** Seien M, N Moduln über R. Nach Proposition 10.1.9(v), angewendet mit $n = 2$ und $\sigma = (1\ 2)$, gibt es einen Isomorphismus

$$\begin{aligned} M \otimes_R N &\xrightarrow{\sim} N \otimes_R M, \\ x \otimes y &\mapsto y \otimes x. \end{aligned}$$

Bemerkung 10.1.12 (Sonderfälle der Distributivität)**.** Seien M, N, P Moduln über R. Nach Proposition 10.1.9(vi), angewendet mit $n = 2$ und $|J| = 2$, gibt es Isomorphismen

$$\begin{array}{ll} (M \oplus N) \otimes_R P \xrightarrow{\sim} (M \otimes_R P) \oplus (N \otimes_R P), & M \otimes_R (N \oplus P) \xrightarrow{\sim} (M \otimes_R N) \oplus (M \otimes_R P), \\ (x, y) \otimes z \mapsto (x \otimes z, y \otimes z), & x \otimes (y, z) \mapsto (x \otimes y, x \otimes z). \end{array}$$

Nach Proposition 10.1.9(vi), angewendet mit $n = 2$ und $|J| = 0$, gibt es Isomorphismen

$$\{0\} \otimes_R M \xrightarrow{\sim} \{0\} \xleftarrow{\sim} M \otimes_R \{0\}.$$

Bemerkung 10.1.13. Proposition 10.1.9(iii) ist ein lineares Analogon des Exponentialgesetzes für Mengen X, Y, Z:

$$\mathrm{Abb}(X \times Y, Z) \xrightarrow{\sim} \mathrm{Abb}(X, \mathrm{Abb}(Y, Z)), \quad f \mapsto (x \mapsto (y \mapsto f(x, y))).$$

Sind X, Y, Z endliche Mengen der Mächtigkeit $a, b, c \in \mathbb{N}$, so ist diese Bijektion eine Verfeinerung der Gleichheit $c^{a \cdot b} = (c^b)^a$.

Korollar 10.1.14 (Dimension des Tensorprodukts)**.** *Sei R ein kommutativer Ring und $n \in \mathbb{N}$. Zu jedem $i \in \{1, \dots, n\}$ sei M_i ein freier R-Modul mit Basis $(b_{ij})_{j \in I_i}$. Dann ist das Tensorprodukt $M_1 \otimes_R \cdots \otimes_R M_n$ ein freier R-Modul mit Basis*

$$(b_{1j_1} \otimes \cdots \otimes b_{nj_n})_{(j_1, \dots, j_n) \in I_1 \times \cdots \times I_n}.$$

Insbesondere: Ist jedes M_i frei vom Rang $d_i \in \mathbb{N}$, so ist $M_1 \otimes_R \cdots \otimes_R M_n$ frei vom Rang $\prod_{i=1}^n d_i$.

Beweis. Nach Proposition 10.1.8(iii) können wir annehmen, dass $M_i = R^{(I_i)}$ und dass die gegebene Basis die Standardbasis $(e_j)_{j \in I_i}$ ist. Nach der Distributivität des Tensorprodukts über direkte Summen (Proposition 10.1.9(vi)) gibt es einen Isomorphismus

$$\begin{aligned} R^{(I_1)} \otimes_R \cdots \otimes_R R^{(I_n)} &\xrightarrow{\sim} (R \otimes_R \cdots \otimes_R R)^{(I_1 \times \cdots \times I_n)}, \\ x_1 \otimes \cdots \otimes x_n &\mapsto (x_{1k_1} \otimes \cdots \otimes x_{nk_n})_{(k_1, \dots, k_n) \in I_1 \times \cdots \times I_n}. \end{aligned}$$

Ferner gibt es einen Isomorphismus

$$\begin{aligned} R \otimes_R \cdots \otimes_R R &\xrightarrow{\sim} R, \\ r_1 \otimes \cdots \otimes r_n &\mapsto r_1 \dots r_n, \end{aligned}$$

nach Proposition 10.1.9 (i) und (iv) (angewendet mit $n = 0$ und $k = n$). Durch diese Isomorphismen werden die Tensoren $e_{j_1} \otimes \cdots \otimes e_{j_k}$ auf die Elemente

$$(\delta_{j_1 k_1} \dots \delta_{j_n k_n})_{(k_1,\dots,k_n) \in I_1 \times \cdots \times I_n} = e_{(j_1,\cdots,j_n)}$$

abgebildet, die die Standardbasis von $R^{(I_1 \times \cdots \times I_n)}$ bilden. □

Proposition 10.1.15 (Erzeugung durch reine Tensoren)**.** *Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$ und seien $M_1, \dots, M_n$ Moduln über R. Sei (T, τ) ein Tensorprodukt von $(M_1, \dots, M_n)$. Dann ist das Bild von τ ein Erzeugendensystem von T. Wenn $n \geq 1$ ist sogar jedes Element von T eine Summe von reinen Tensoren.*

Beweis. Sei $T' = \mathrm{Span}_R(\tau(M_1 \times \cdots \times M_n)) \subset T$, sei $i\colon T' \hookrightarrow T$ die Inklusionsabbildung und sei $\tau'\colon M_1 \times \cdots \times M_n \to T'$ die Abbildung mit $i \circ \tau' = \tau$. Da τ' n-linear ist, gibt es eine lineare Abbildung $f\colon T \to T'$ mit $f \circ \tau = \tau'$. Für die Komposition $i \circ f$ gilt

$$i \circ f \circ \tau = i \circ \tau' = \tau.$$

Nach der universellen Eigenschaft von (T, τ) ist dann $i \circ f = \mathrm{id}_T$. Insbesondere ist i surjektiv, d.h., es gilt $T' = T$, wie gewünscht. Zur letzten Aussage muss man noch bemerken: Wenn $n \geq 1$ sind Skalarvielfache von reinen Tensoren wieder reine Tensoren, nach der n-Linearität von τ: $\lambda(x_1 \otimes \cdots \otimes x_n) = (\lambda x_1) \otimes \cdots \otimes x_n$. □

Beispiel 10.1.16 (unreine Tensoren)**.** Nicht alle Tensoren sind reine Tensoren. Sei zum Beispiel $R \neq \{0\}$ und sei

$$t = e_1 \otimes e_2 + e_2 \otimes e_1 \in R^2 \otimes_R R^2.$$

Wir behaupten, dass t kein reiner Tensor ist, d.h.: Für alle $x, y \in R^2$ gilt $x \otimes y \neq t$. Denn es gilt

$$\begin{aligned} x \otimes y &= (x_1 e_1 + x_2 e_2) \otimes y \\ &= x_1(e_1 \otimes y) + x_2(e_2 \otimes y) \\ &= x_1(e_1 \otimes (y_1 e_1 + y_2 e_2)) + x_2(e_2 \otimes (y_1 e_1 + y_2 e_2)) \\ &= x_1 y_1 (e_1 \otimes e_1) + x_1 y_2 (e_1 \otimes e_2) + x_2 y_1 (e_2 \otimes e_2) + x_2 y_2 (e_2 \otimes e_2). \end{aligned}$$

Da die Tensoren $e_i \otimes e_j$ eine Basis von $R^2 \otimes_R R^2$ bilden (Korollar 10.1.14), ist die Gleichung $x \otimes y = t$ äquivalent zu den vier Gleichungen $x_1 y_2 = x_2 y_1 = 1$ und $x_1 y_1 = x_2 y_2 = 0$. Die ersten beiden implizieren, dass x_1 und y_1 Einheiten sind, so dass $0 = x_1 y_1$ auch eine Einheit ist. Das ist aber nur möglich, wenn $R = \{0\}$.

Beispiel 10.1.17 (Tensorprodukte von abelschen Gruppen)**.**

(i) Seien $n, m \in \mathbb{N}$. Es gilt dann

$$\mathbb{Z}/n\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/(n, m) = \mathbb{Z}/d\mathbb{Z},$$

wobei $d = \mathrm{ggT}(n, m)$. Denn sei

$$\tau\colon \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/(n, m), \quad ([x], [y]) \mapsto [xy].$$

Die Abbildung τ ist wohldefiniert und $\mathbb{Z}$-bilinear. Sei $f\colon \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to A$ eine beliebige $\mathbb{Z}$-bilineare Abbildung. Eine solche Abbildung ist eindeutig durch das Element $a = f([1], [1]) \in A$ bestimmt, denn $f([x], [y]) = xya$. Zudem gilt $na = f([n], [1]) = 0$ und $ma = f([1], [m]) = 0$. Damit gibt es genau eine $\mathbb{Z}$-lineare Abbildung $g\colon \mathbb{Z}/(n, m) \to A$ mit $g([1]) = a$, d.h., $g \circ \tau = f$.

(ii) Ist $n \in \mathbb{N} \setminus \{0\}$, so gilt $\mathbb{Z}/n\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Q} = \{0\}$. Denn für einen reinen Tensor $[x] \otimes q$ gilt

$$[x] \otimes q = [x] \otimes n \cdot \frac{q}{n} = n \cdot [x] \otimes \frac{q}{n} = 0 \otimes \frac{q}{n} = 0.$$

Auf ähnliche Weise gilt $\mathbb{Z}/n\mathbb{Z} \otimes_{\mathbb{Z}} K = \{0\}$ für alle Körper K der Charakteristik 0.

Proposition 10.1.18 (Tensorprodukt und Dualität)**.** *Sei R ein kommutativer Ring, $n \in \mathbb{N}$ und $M_1, \dots, M_n$ Moduln über R. Dann gibt es eine lineare Abbildung*

$$\begin{aligned} M_1^* \otimes_R \cdots \otimes_R M_n^* &\to (M_1 \otimes_R \cdots \otimes_R M_n)^*, \\ \alpha_1 \otimes \cdots \otimes \alpha_n &\mapsto (x_1 \otimes \cdots \otimes x_n \mapsto \alpha_1(x_1) \cdot \ldots \cdot \alpha_n(x_n)). \end{aligned}$$

Sie ist ein Isomorphismus, wenn jedes M_i frei und endlich erzeugt ist.

Beweis. Durch das Exponentialgesetz entspricht diese lineare Abbildung der $2n$-linearen Form

$$\begin{aligned} M_1^* \times \cdots \times M_n^* \times M_1 \times \cdots \times M_n &\to R, \\ (\alpha_1, \dots, \alpha_n, x_1, \dots, x_n) &\mapsto \alpha_1(x_1) \cdot \ldots \cdot \alpha_n(x_n). \end{aligned}$$

Hat jedes M_i eine endliche Basis, so hat $M_1 \otimes_R \cdots \otimes_R M_n$ eine induzierte endliche Basis nach Korollar 10.1.14, und die Dualmoduln dazu haben duale Basen. Die gegebene Abbildung bildet dann eine Basis auf eine Basis ab, und ist damit ein Isomorphismus. □

10.1.1 Alternative Definitionen von Ringen, Moduln und Algebren

Man schreibt oft $\otimes$ anstelle von $\otimes_{\mathbb{Z}}$ für das Tensorprodukt von abelschen Gruppen.

Wie wir bereits bemerkt haben (siehe Bemerkung 8.1.43), ist das Distributivgesetz in der Definition eines Ringes R äquivalent zu der $\mathbb{Z}$-Bilinearität der Multiplikationsabbildung $\cdot\colon R \times R \to R$. Insbesondere induziert die Multiplikation eine $\mathbb{Z}$-lineare Abbildung (d.h., einen Gruppenhomomorphismus) $m\colon R \otimes R \to R$ mit $m(r \otimes s) = r \cdot s$. Es gibt auch genau einen Gruppenhomomorphismus $e\colon \mathbb{Z} \to R$ mit $e(1) = 1$ (da $(\mathbb{Z}, 1)$ das nulläre Tensorprodukt von abelschen Gruppen ist, nach Proposition 10.1.9(i)). Man kann dann alle Axiome für einen Ring durch die Gruppenhomomorphismen m und e ausdrücken, was eine neue aber äquivalente Definition liefert:

Definition 10.1.19 (Ring, Ringhomomorphismus)**.** Ein *Ring* ist ein Tripel (R, m, e) bestehend aus einer abelschen Gruppe R und Gruppenhomomorphismen

$$m\colon R \otimes R \to R \quad \text{und} \quad e\colon \mathbb{Z} \to R,$$

so dass folgende Diagramme kommutativ sind:

(i) (Neutralität)

$$\begin{array}{ccccc} \mathbb{Z} \otimes R & \xrightarrow{e \otimes \mathrm{id}_R} & R \otimes R & \xleftarrow{\mathrm{id}_R \otimes e} & R \otimes \mathbb{Z} \\ & \searrow^{\sim} & \downarrow m & \swarrow_{\sim} & \\ & & R, & & \end{array}$$

wobei die diagonalen Pfeile die Isomorphismen aus Bemerkung 10.1.10 sind.

(ii) (Assoziativität)

$$\begin{array}{ccc} R \otimes R \otimes R & \xrightarrow{\mathrm{id}_R \otimes m} & R \otimes R \\ {\scriptstyle m \otimes \mathrm{id}_R}\downarrow & & \downarrow{\scriptstyle m} \\ R \otimes R & \xrightarrow{\quad m \quad} & R. \end{array}$$

Der Ring (R, m, e) heißt *kommutativ*, wenn außerdem das folgende Diagramm kommutiert:

(iii) (Kommutativität)

$$\begin{array}{ccc} R \otimes R & \xrightarrow{\sim} & R \otimes R \\ & {\scriptstyle m}\searrow \quad \swarrow{\scriptstyle m} & \\ & R, & \end{array}$$

wobei der obere Pfeil der Isomorphismus $r \otimes s \mapsto s \otimes r$ aus Bemerkung 10.1.11 ist.

Ein *Ringhomomorphismus* von (R, m_R, e_R) nach (S, m_S, e_S) ist ein Gruppenhomomorphismus $f\colon R \to S$, so dass beide folgenden Diagramme kommutativ sind:

$$\begin{array}{ccc} R \otimes R & \xrightarrow{f \otimes f} & S \otimes S \\ {\scriptstyle m_R}\downarrow & & \downarrow{\scriptstyle m_S} \\ R & \xrightarrow{f} & S, \end{array} \qquad \begin{array}{ccc} \mathbb{Z} & \xrightarrow{\mathrm{id}_{\mathbb{Z}}} & \mathbb{Z} \\ {\scriptstyle e_R}\downarrow & & \downarrow{\scriptstyle e_S} \\ R & \xrightarrow{f} & S. \end{array}$$

Die Definition eines Moduls kann man auf ähnliche Weise umformulieren:

Definition 10.1.20 (Modul, lineare Abbildung)**.** Sei (R, m, e) ein Ring. Ein *R-Modul* ist ein Paar (M, a) bestehend aus einer abelschen Gruppe M und einem Gruppenhomomorphismus $a\colon R \otimes M \to M$, so dass folgende Diagramme kommutativ sind:

(i) (Neutralität)

$$\begin{array}{ccc} \mathbb{Z} \otimes M & \xrightarrow{e \otimes \mathrm{id}_M} & R \otimes M \\ & \searrow^{\sim} & \downarrow{\scriptstyle a} \\ & & M. \end{array}$$

(ii) (Assoziativität)

$$\begin{array}{ccc} R \otimes R \otimes M & \xrightarrow{\mathrm{id}_R \otimes a} & R \otimes M \\ {\scriptstyle m \otimes \mathrm{id}_M}\downarrow & & \downarrow{\scriptstyle a} \\ R \otimes M & \xrightarrow{a} & M. \end{array}$$

Eine *R-lineare Abbildung* von (M, a_M) nach (N, a_N) ist ein Gruppenhomomorphismus $f\colon M \to N$, so dass folgendes Diagramm kommutativ ist:

$$\begin{array}{ccc} R \otimes M & \xrightarrow{\mathrm{id}_R \otimes f} & R \otimes N \\ {\scriptstyle a_M}\downarrow & & \downarrow{\scriptstyle a_N} \\ M & \xrightarrow{f} & N. \end{array}$$

Bemerkung 10.1.21 (Moduln als Ringhomomorphismen)**.** Nach dem Exponentialgesetz ist ein Gruppenhomomorphismus $a\colon R \otimes M \to M$ äquivalent zu einem Gruppenhomomorphismus $a'\colon R \to \mathrm{End}_{\mathbb{Z}}(M)$. Man beachte dabei, dass $\mathrm{End}_{\mathbb{Z}}(M)$ ein Ring bezüglich der Komposition ist. Man kann dann leicht nachprüfen, dass die Abbildung a genau dann die Axiome (i) und (ii) der Definition 10.1.20 erfüllt, wenn a' ein Ringhomomorphismus ist. Das heißt, eine R-Modulstruktur auf einer abelschen Gruppe M ist äquivalent zu einem Ringhomomorphismus $R \to \mathrm{End}_{\mathbb{Z}}(M)$.

Die Definition einer R-Algebra ist nun genau dieselbe wie die eines Ringes, indem wir abelsche Gruppen durch R-Moduln (und somit $\otimes$ durch $\otimes_R$) ersetzen:

Definition 10.1.22 (Algebra, Algebrenhomomorphismus)**.** Sei R ein kommutativer Ring. Eine *R-Algebra* ist ein Tripel (A, m, e) bestehend aus einem R-Modul A und R-linearen Abbildungen

$$m\colon A \otimes_R A \to A \quad \text{und} \quad e\colon R \to A,$$

so dass folgende Diagramme kommutativ sind:

(i) (Neutralität)

$$\begin{array}{ccccc} R \otimes_R A & \xrightarrow{e\otimes \mathrm{id}_A} & A \otimes_R A & \xleftarrow{\mathrm{id}_A \otimes e} & A \otimes_R R \\ & \searrow^{\sim} & \downarrow{\scriptstyle m} & \swarrow_{\sim} & \\ & & A, & & \end{array}$$

wobei die diagonalen Pfeile die Isomorphismen aus Bemerkung 10.1.10 sind.

(ii) (Assoziativität)

$$\begin{array}{ccc} A \otimes_R A \otimes_R A & \xrightarrow{\mathrm{id}_A \otimes m} & A \otimes_R A \\ {\scriptstyle m\otimes \mathrm{id}_A}\downarrow & & \downarrow{\scriptstyle m} \\ A \otimes_R A & \xrightarrow{\quad m \quad} & A. \end{array}$$

Die R-Algebra (A, m, e) heißt *kommutativ*, wenn außerdem das folgende Diagramm kommutiert:

(iii) (Kommutativität)

$$\begin{array}{ccc} A \otimes_R A & \xrightarrow{\sim} & A \otimes_R A \\ & {\scriptstyle m}\searrow \quad \swarrow{\scriptstyle m} & \\ & A, & \end{array}$$

wobei der obere Pfeil der Isomorphismus $a \otimes b \mapsto b \otimes a$ aus Bemerkung 10.1.11 ist.

Ein *R-Algebrenhomomorphismus* von (A, m_A, e_A) nach (B, m_B, e_B) ist eine R-lineare Abbildung $f\colon A \to B$, so dass beide folgenden Diagramme kommutativ sind:

$$\begin{array}{ccc} A \otimes_R A & \xrightarrow{f\otimes f} & B \otimes_R B \\ {\scriptstyle m_A}\downarrow & & \downarrow{\scriptstyle m_B} \\ A & \xrightarrow{\quad f \quad} & B, \end{array} \qquad \begin{array}{ccc} R & \xrightarrow{\mathrm{id}_R} & R \\ {\scriptstyle e_A}\downarrow & & \downarrow{\scriptstyle e_B} \\ A & \xrightarrow{\quad f \quad} & B. \end{array}$$

Bemerkung 10.1.23. Ist (A, m, e) eine R-Algebra wie in Definition 10.1.22, so ist $e\colon R \to A$ der Ringhomomorphismus entsprechend der R-Algebra A wie in Proposition 8.1.50.

Wir besprechen noch ein paar Anwendungen des Tensorprodukts.

Bemerkung 10.1.24 (Skalarerweiterung)**.** Sei $\varphi\colon R \to S$ ein Ringhomomorphismus. Jeder S-Modul N hat dann eine Struktur von R-Modul mit Skalarmultiplikation $(r, y) \mapsto \varphi(r) \cdot y$. Dieser R-Modul heißt die *Skalareinschränkung* von N entlang φ. Sei nun M ein R-Modul. Eine *Skalarerweiterung* von M entlang φ ist ein Paar (M_φ, u) bestehend aus einem S-Modul M_φ und einer R-linearen Abbildung $u\colon M \to M_\varphi$ mit folgender universellen Eigenschaft: Zu jedem S-Modul N und jeder R-linearen Abbildung $f\colon M \to N$ gibt es genau eine S-lineare Abbildung $g\colon M_\varphi \to P$ mit $g \circ u = f$.

Falls R kommutativ ist und $\varphi(R)$ aus zentralen Elementen von S besteht (d.h., falls S durch φ eine R-Algebra ist), können wir eine Skalarerweiterung M_φ mithilfe des Tensorprodukts konstruieren:

$$M_\varphi := S \otimes_R M, \quad u\colon M \xrightarrow{\sim} R \otimes_R M \xrightarrow{\varphi\otimes \mathrm{id}_M} S \otimes_R M.$$

Die S-Modulstruktur auf $S \otimes_R M$ wird durch $t \cdot (s \otimes x) = ts \otimes x$ definiert. Genauer ist die Skalarmultiplikation die folgende Komposition:

$$S \times (S \otimes_R M) \to S \otimes_R (S \otimes_R M) \xrightarrow{\sim} (S \otimes_R S) \otimes_R M \xrightarrow{m_S \otimes \mathrm{id}_M} S \otimes_R M.$$

Ist N ein S-Modul und ist $f\colon M \to N$ eine R-lineare Abbildung, so erhalten wir eine R-bilineare Abbildung

$$S \times M \to N, \quad (s, x) \mapsto s \cdot f(x)$$

die eine R-lineare Abbildung $g\colon S \otimes_R M \to N$ mit $g \circ u = f$ induziert. Es ist dann klar, dass die Abbildung g S-linear ist, und dass sie die eindeutige S-lineare Abbildung mit $g \circ u = f$ ist, denn jede solche Abbildung muss $1 \otimes x = u(x)$ auf $f(x)$ und somit $s \otimes x$ auf $s \cdot f(x)$ abbilden.

Eine Skalarerweiterung von M entlang φ existiert allgemeiner für einen beliebigen Ringhomomorphismus $\varphi\colon R \to S$. Man definiert nämlich M_φ als den Quotient

$$M_\varphi := (S \otimes_{\mathbb{Z}} M)/\mathrm{Span}_S(E),$$

wobei $E \subset S \otimes_{\mathbb{Z}} M$ die Teilmenge bestehend aus den Elementen $\varphi(r) \otimes x - 1 \otimes rx$ ist. Dies gewährleistet die R-Linearität der Abbildung

$$u\colon M \xrightarrow{\sim} \mathbb{Z} \otimes_{\mathbb{Z}} M \xrightarrow{e_S \otimes \mathrm{id}_M} S \otimes_{\mathbb{Z}} M \twoheadrightarrow M_\varphi.$$

Die gewünschte universelle Eigenschaft kann man dann leicht nachprüfen, indem man die universellen Eigenschaften von $\otimes_{\mathbb{Z}}$ und von dem Quotientenmodul anwendet. Der S-Modul M_φ bezeichnet man auch in diesem Fall mit $S \otimes_R M$, aber die Notation $\otimes_R$ hat jetzt eine andere Bedeutung als oben (selbst wenn R kommutativ ist): Sie stellt nämlich ein Tensorprodukt von *Bimoduln* dar.

Bemerkung 10.1.25 (Skalarerweiterung und Matrizen)**.** Seien R und S kommutative Ringe und sei $\varphi\colon R \to S$ ein Ringhomomorphismus. Nach Proposition 10.1.9(vi) (oder nach Vergleich von universellen Eigenschaften) ist die Skalarerweiterung eines freien R-Moduls M entlang φ wieder frei: Ist genauer $(b_i)_{i \in I}$ eine Basis von M, so ist $(1 \otimes b_i)_{i \in I}$ eine Basis von $S \otimes_R M$. Seien nun N und M endlich erzeugte freie R-Moduln mit Basen B und C und sei $f\colon N \to M$ eine R-lineare Abbildung. Seien B' und C' die entsprechenden Basen von $S \otimes_R N$ und $S \otimes_R M$. Dann gilt

$$[\mathrm{id}_S \otimes f]_{C'}^{B'} = \varphi([f]_C^B), \quad \text{wobei} \quad \varphi\left((a_{ij})_{i,j}\right) := (\varphi(a_{ij}))_{i,j}.$$

Sei zum Beispiel $K \subset L$ eine Körpererweiterung, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$. Ist B eine Basis von V und ist B' die entsprechende Basis von $L \otimes_K V$, so gilt $[f]_B^B = [\mathrm{id}_L \otimes f]_{B'}^{B'}$. Aus Lemma 9.1.10 und Proposition 9.1.17 folgt die Gleichheit von Minimalpolynomen $m_f = m_{\mathrm{id}_L \otimes f}$ in $L[T]$, was die matrixfreie Form dieser Proposition ist.

Beispiel 10.1.26. Die $\mathbb{R}$-lineare Abbildung $\mathbb{C} \to \mathbb{C} \oplus \mathbb{C}$, $w \mapsto (w, \bar{w})$, induziert durch die universelle Eigenschaft der Skalarerweiterung eine $\mathbb{C}$-lineare Abbildung

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \to \mathbb{C} \oplus \mathbb{C}, \quad z \otimes w \mapsto (zw, z\bar{w}),$$

die ein Isomorphismus ist. Denn sie bildet die komplexe Basis $(1 \otimes 1, 1 \otimes i)$ auf die komplexe Basis $((1,1),(i,-i))$ ab.

Bemerkung 10.1.27 (Rang durch das Tensorprodukt)**.** Sei A eine endlich erzeugte abelsche Gruppe des Ranges n. Nach Satz 8.3.30 ist A die direkte Summe von $\mathbb{Z}^n$ und abelschen Gruppen der Form $\mathbb{Z}/d\mathbb{Z}$ mit $d \geq 2$. Nach Proposition 10.1.9(vi) und Beispiel 10.1.17(ii) gilt dann $n = \dim_{\mathbb{Q}}(\mathbb{Q} \otimes_{\mathbb{Z}} A)$. Auf ähnliche Weise kann man den Rang eines Moduls M über einem beliebigen Integritätsring R definieren: Ganz analog zu der Konstruktion von $\mathbb{Q}$ aus $\mathbb{Z}$ (siehe Abschnitt 2.4.1) kann man einen Körper $\mathrm{Quot}(R)$ definieren, den *Quotientenkörper* von R, der R als Unterring enthält und in dem jedes Element ein Bruch zweier Elemente von R ist. Man definiert dann den Rang von M als die Dimension des $\mathrm{Quot}(R)$-Vektorraums $\mathrm{Quot}(R) \otimes_R M$.

Bemerkung 10.1.28 (Primärteiler durch das Tensoprodukt)**.** Sei A eine endlich erzeugte abelsche Gruppe. Für eine Primzahl p und eine natürliche Zahl $e \geq 1$, sei $\alpha_A(p,e)$ die Anzahl der Primärteiler von A, die gleich p^e sind. Nach Beispiel 10.1.17(i,ii) gilt

$$\sum_{e \in \mathbb{N}\setminus\{0\}} \alpha_A(p,e) = \dim_{\mathbb{F}_p}(\mathbb{F}_p \otimes_{\mathbb{Z}} A) - \dim_{\mathbb{Q}}(\mathbb{Q} \otimes_{\mathbb{Z}} A).$$

Zudem gilt $\alpha_{p^i A}(p,e) = \alpha_A(p, e+i)$. Daraus folgern wir die Formel

$$\alpha_A(p,e) = \dim_{\mathbb{F}_p}(\mathbb{F}_p \otimes_{\mathbb{Z}} p^{e-1}A) - \dim_{\mathbb{F}_p}(\mathbb{F}_p \otimes_{\mathbb{Z}} p^e A),$$

die sich auch auf beliebige Hauptidealringe verallgemeinern lässt. Dies gibt einen neuen Beweis der Eindeutigkeit der Primärteiler.

Bemerkung 10.1.29 (Matrixfreie Definition der Spur)**.** Sei R ein kommutativer Ring und M, N Moduln über R. Die Abbildung

$$M^* \times N \to \mathrm{Hom}_R(M,N), \quad (\alpha, y) \mapsto (x \mapsto \alpha(x)y),$$

ist bilinear und induziert damit eine lineare Abbildung

$$\omega_{M,N} \colon M^* \otimes_R N \to \mathrm{Hom}_R(M,N).$$

Man bemerkt, dass $\omega_{R,N}$ ein Isomorphismus ist. Sind zudem $\omega_{M,N}$ und $\omega_{M',N}$ Isomorphismen, so ist auch $\omega_{M\oplus M',N}$ ein Isomorphismus. Daraus folgt induktiv, dass $\omega_{M,N}$ ein Isomorphismus ist, wenn M ein freier R-Modul vom endlichen Rang ist. Insbesondere erhalten wir in diesem Fall einen kanonischen Isomorphismus

$$\omega_{M,M} \colon M^* \otimes_R M \xrightarrow{\sim} \mathrm{End}_R(M).$$

Auf der anderen Seite gibt es eine lineare Abbildung

$$e \colon M^* \otimes_R M \to R, \quad \alpha \otimes x \mapsto \alpha(x).$$

Ist M frei vom endlichen Rang und ist $f \in \mathrm{End}_R(M)$ ein Endomorphismus, so definiert man die *Spur* $\mathrm{tr}(f)$ von f durch

$$\mathrm{tr}(f) = e(\omega_{M,M}^{-1}(f)) \in R.$$

Nach Konstruktion ist die Spur eine Isomorphie-Invariante von Endomorphismen, und falls $M = R^n$ und $f = L_A$ mit $A \in M_n(R)$, kann man leicht nachrechnen, dass $\mathrm{tr}(f) = \mathrm{tr}(A)$. Deswegen stimmt diese neue Definition der Spur mit Definition 6.1.24 überein.

Ein Vorteil dieser matrixfreien Definition ist, dass die Abbildung $\omega_{M,M}$ ein Isomorphismus sein kann, selbst wenn M nicht frei ist (sie ist genau dann ein Isomorphismus, wenn M endlich erzeugt und *projektiv* ist). Ein Endomorphismus f eines solchen M hat keine Matrixdarstellung, aber die Spur von f ist trotzdem definiert.

10.1.2 Tensorpotenzen und die Tensoralgebra

Definition 10.1.30 (Tensorpotenz)**.** Sei R ein kommutativer Ring, M ein R-Modul und $n \in \mathbb{N}$. Ein Tensorprodukt der n-elementigen Familie $(M, \ldots, M)$ heißt eine *n-te Tensorpotenz* von M und wird mit $\mathrm{T}^n_R(M)$ oder $M^{\otimes n}$ bezeichnet.

Nach Proposition 10.1.9(i,ii) gilt $\mathrm{T}^0_R(M) = R$ und $\mathrm{T}^1_R(M) = M$. Nach Proposition 10.1.9(iii) gibt es kanonische Isomorphismen

$$m_{p,q} \colon \mathrm{T}^p_R(M) \otimes_R \mathrm{T}^q_R(M) \xrightarrow{\sim} \mathrm{T}^{p+q}_R(M),$$
$$(x_1 \otimes \cdots \otimes x_p) \otimes (y_1 \otimes \cdots \otimes y_q) \mapsto x_1 \otimes \cdots \otimes x_p \otimes y_1 \otimes \cdots \otimes y_q.$$

Wir setzen

$$\mathrm{T}_R^*(M) = \bigoplus_{n\in\mathbb{N}} \mathrm{T}_R^n(M),$$

und wir bezeichnen mit $\iota_n\colon \mathrm{T}_R^n(M) \hookrightarrow \mathrm{T}_R^*(M)$ die kanonischen Abbildungen. Nach Proposition 10.1.9(v) und der universellen Eigenschaft der direkten Summe induzieren die Abbildungen $m_{p,q}$ eine Verknüpfung

$$m\colon \mathrm{T}_R^*(M) \otimes_R \mathrm{T}_R^*(M) \xrightarrow{\sim} \bigoplus_{p,q\in\mathbb{N}} \mathrm{T}_R^p(M) \otimes_R \mathrm{T}_R^q(M) \to \mathrm{T}_R^*(M).$$

Die obige Formel für reine Tensoren zeigt, dass diese Verknüpfung assoziativ ist, und dass das Element $1 \in \mathrm{T}_R^0(M) = R$ ein neutrales Element ist. Damit haben wir eine R-Algebra definiert:

Definition 10.1.31 (Tensoralgebra)**.** Sei R ein kommutativer Ring und M ein R-Modul. Die *Tensoralgebra* von M ist die R-Algebra $(\mathrm{T}_R^*(M), m, \iota_0)$. Die Multiplikation auf $\mathrm{T}_R^*(M)$ wird auch mit dem Symbol $\otimes$ geschrieben.

Proposition 10.1.32 (universelle Eigenschaft der Tensoralgebra)**.** *Sei R ein kommutativer Ring und M ein R-Modul. Zu jeder R-Algebra A und jeder R-linearen Abbildung $f\colon M \to A$ gibt es* genau einen *R-Algebrenhomomorphismus $\hat{f}\colon \mathrm{T}_R^*(M) \to A$ mit $\hat{f} \circ \iota_1 = f$.*

Beweis. Die Eindeutigkeit von $\hat{f}$ folgt aus Proposition 10.1.15, denn jedes Element von $\mathrm{T}_R^*(M)$ ist eine Linearkombination von reinen Tensoren, die selbst Produkte von Elementen von $\mathrm{T}_R^1(M) = M$ sind. Da $\hat{f}$ Linearkombinationen und Produkte erhält, ist es durch seine Einschränkung auf M eindeutig bestimmt.

Für jedes $n \in \mathbb{N}$ ist die Abbildung

$$M^n \to A, \quad (x_1, \dots, x_n) \mapsto f(x_1) \cdot \ldots \cdot f(x_n),$$

n-linear und induziert eine R-lineare Abbildung $\hat{f}_n\colon \mathrm{T}_R^n(M) \to A$. Sei $\hat{f}\colon \mathrm{T}_R^*(M) \to A$ die R-lineare Abbildung, so dass für alle $n \in \mathbb{N}$ gilt $\hat{f} \circ \iota_n = \hat{f}_n$. Es bleibt zu zeigen, dass $\hat{f}$ ein Ringhomomorphismus ist. Es gilt $\hat{f}(1) = \hat{f}_0(1) = 1$ (das leere Produkt in A). Seien $x, y \in \mathrm{T}_R^*(M)$. Man muss noch $\hat{f}(x \otimes y) = \hat{f}(x) \cdot \hat{f}(y)$ nachprüfen. Nach Proposition 10.1.15 können wir annehmen, dass $x = x_1 \otimes \cdots \otimes x_p$ und $y = y_1 \otimes \cdots \otimes y_q$ reine Tensoren sind (da $\hat{f}$ R-linear ist und die Multiplikationen auf $\mathrm{T}_R^*(M)$ und auf A R-bilinear sind). Das Gewünschte folgt dann aus den Definitionen, denn:

$$\begin{aligned}\hat{f}(x \otimes y) &= \hat{f}_{p+q}(x_1 \otimes \cdots \otimes x_p \otimes y_1 \otimes \cdots \otimes y_q)\\ &= f(x_1) \cdot \ldots \cdot f(x_p) \cdot f(y_1) \cdot \ldots \cdot f(y_q)\\ &= \hat{f}(x) \cdot \hat{f}(y).\end{aligned}$$

□

Insbesondere hat die Tensoralgebra $\mathrm{T}_R^*(R)$ genau dieselbe universelle Eigenschaft wie die Polynomalgebra $R[T]$ (Proposition 8.1.45). Es gibt deshalb einen Isomorphismus von R-Algebren

$$R[T] \xrightarrow{\sim} \mathrm{T}_R^*(R), \quad T \mapsto 1 \in R = \mathrm{T}_R^1(R).$$

10.2 Symmetrische und äußere Potenzen

Wir erinnern an die Definition von symmetrischen, antisymmetrischen und alternierenden multilinearen Abbildungen (Definition 5.3.12):

Definition 10.2.1 (symmetrisch, antisymmetrisch, alternierend)**.** Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$, seien M, N Moduln über R und sei $f\colon M^n \to N$ eine n-lineare Abbildung.

- f heißt *symmetrisch*, wenn für alle $i, j \in \{1, \dots, n\}$ mit $i < j$ und alle $(x_1, \dots, x_n) \in M^n$ gilt
$$f(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = f(x_1, \dots, x_j, \dots, x_i, \dots, x_n).$$
- f heißt *antisymmetrisch*, wenn für alle $i, j \in \{1, \dots, n\}$ mit $i < j$ und alle $(x_1, \dots, x_n) \in M^n$ gilt
$$f(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = -f(x_1, \dots, x_j, \dots, x_i, \dots, x_n).$$
- f heißt *alternierend*, wenn für alle $i, j \in \{1, \dots, n\}$ mit $i < j$ und alle $(x_1, \dots, x_n) \in M^n$ mit $x_i = x_j$ gilt
$$f(x_1, \dots, x_n) = 0.$$

Proposition 10.2.2. *Sei R ein kommutativer Ring, M, N Moduln über R und $n \in \mathbb{N}$. Sei $f\colon M^n \to N$ eine n-lineare Abbildung. Dann:*

(i) *f ist genau dann symmetrisch, wenn für alle $(x_1, \dots, x_n) \in M^n$ und alle $\sigma \in S_n$ gilt*
$$f(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = f(x_1, \dots, x_n).$$

(ii) *f ist genau dann antisymmetrisch, wenn für alle $(x_1, \dots, x_n) \in M^n$ und alle $\sigma \in S_n$ gilt*
$$f(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = \operatorname{sgn}(\sigma) f(x_1, \dots, x_n).$$

(iii) *f ist genau dann alternierend, wenn für alle $i \in \{1, \dots, n-1\}$ und alle $(x_1, \dots, x_n) \in M^n$ mit $x_i = x_{i+1}$ gilt*
$$f(x_1, \dots, x_n) = 0.$$

(iv) *Ist f alternierend, so ist f antisymmetrisch. Die Umkehrung gilt, wenn 2 eine Einheit in R ist (oder allgemeiner, wenn die Multiplikation mit 2 auf N injektiv ist).*

Beweis. Aussagen (i) und (ii) folgen unmittelbar aus Lemma 5.3.4. Aussage (iv) folgt aus der Bemerkung 5.3.14. Wir beweisen nun (iii). Sei $f\colon M^n \to N$ mit der gegebenen Eigenschaft, sei $i < j$ und sei $(x_1, \dots, x_n) \in M^n$ mit $x_i = x_j$. Wir zeigen $f(x_1, \dots, x_n) = 0$ durch Induktion über $j - i$. Falls $j - i = 1$ ist dies die Voraussetzung. Sei dann $j - i \geq 2$. Wir legen alle x_k mit $k \notin \{i, i+1, j\}$ fest und betrachten die resultierende trilineare Abbildung $g\colon M^3 \to N$. Für alle $x, y \in M$ gilt dann $g(x, x, y) = 0$ nach Voraussetzung und $g(y, x, x) = 0$ nach Induktionsvoraussetzung. Daraus folgt
$$g(x, y, x) = g((x+y), x, x) + g(x, y, x) + g(y, y, x) = g((x+y), (x+y), x) = 0,$$
wie gewünscht. □

Definition 10.2.3 (symmetrische/äußere Potenz)**.** Sei R ein kommutativer Ring, M ein R-Modul und $n \in \mathbb{N}$.

- Eine *n-te symmetrische Potenz* von M ist ein Paar (S, σ) bestehend aus einem R-Modul S und eine symmetrische n-lineare Abbildung $\sigma\colon M^n \to S$ mit folgender universellen Eigenschaft: Zu jedem R-Modul N und jeder symmetrischen n-linearen Abbildung $f\colon M^n \to N$ gibt es *genau eine* R-lineare Abbildung $g\colon S \to N$ mit $g \circ \sigma = f$.
- Eine *n-te äußere Potenz* von M ist ein Paar (A, α) bestehend aus einem R-Modul A und eine alternierende n-lineare Abbildung $\alpha\colon M^n \to A$ mit folgender universellen Eigenschaft: Zu jedem R-Modul N und jeder alternierenden n-linearen Abbildung $f\colon M^n \to N$ gibt es *genau eine* R-lineare Abbildung $g\colon A \to N$ mit $g \circ \alpha = f$.

Satz 10.2.4 (Existenz und Eindeutigkeit der symmetrischen/äußeren Potenzen)**.** *Sei R ein kommutativer Ring, M ein R-Modul und $n \in \mathbb{N}$.*

(i) *Eine n-te symmetrische Potenz (S, σ) von M existiert.*

(ii) *Seien (S, σ) und (S', σ') zwei n-te symmetrische Potenzen von M. Dann gibt es genau eine lineare Abbildung $\varphi\colon S \to S'$ mit $\varphi \circ \sigma = \sigma'$. Außerdem ist φ ein Isomorphismus.*

(iii) *Eine n-te äußere Potenz (A, α) von M existiert.*

(iv) *Seien (A, α) und (A', α') zwei n-te äußere Potenzen von M. Dann gibt es genau eine lineare Abbildung $\varphi\colon S \to S'$ mit $\varphi \circ \alpha = \alpha'$. Außerdem ist φ ein Isomorphismus.*

Beweis. Zu (ii) und (iv) verwendet man genau dasselbe Argument wie im Beweis von Satz 10.1.6(ii).

Zu (i). Sei E die Teilmenge von $\mathrm{T}^n_R(M)$ bestehend aus den Elementen

$$x_1 \otimes \cdots \otimes x_i \otimes \cdots \otimes x_j \otimes \cdots \otimes x_n - x_1 \otimes \cdots \otimes x_j \otimes \cdots \otimes x_i \otimes \cdots \otimes x_n$$

mit $x_1, \ldots, x_n \in M$ und $1 \leq i < j \leq n$. Sei $f\colon M^n \to N$ eine n-lineare Abbildung und sei $g\colon \mathrm{T}^n_R(M) \to N$ die induzierte lineare Abbildung. Nach Definition ist f genau dann symmetrisch, wenn $E \subset \ker g$. Sei dann $S = \mathrm{T}^n_R(M)/\mathrm{Span}_R(E)$ und sei $\sigma\colon M^n \to S$ die Komposition der kanonischen Abbildung $M^n \to \mathrm{T}^n_R(M)$ mit der Quotientenabbildung $\mathrm{T}^n_R(M) \to S$. Nach der universellen Eigenschaft des Quotienten ist das Paar (S, σ) eine n-te symmetrische Potenz von M.

Zu (iii). Sei E die Teilmenge von $\mathrm{T}^n_R(M)$ bestehend aus den Elementen

$$x_1 \otimes \cdots \otimes x_i \otimes \cdots \otimes x_j \otimes \cdots \otimes x_n$$

mit $x_1, \ldots, x_n \in M$, $1 \leq i < j \leq n$ und $x_i = x_j$. Sei $f\colon M^n \to N$ eine n-lineare Abbildung und sei $g\colon \mathrm{T}^n_R(M) \to N$ die induzierte lineare Abbildung. Nach Definition ist f genau dann alternierend, wenn $E \subset \ker g$. Sei dann $A = \mathrm{T}^n_R(M)/\mathrm{Span}_R(E)$ und sei $\alpha\colon M^n \to A$ die Komposition der kanonischen Abbildung $M^n \to \mathrm{T}^n_R(M)$ mit der Quotientenabbildung $\mathrm{T}^n_R(M) \to A$. Nach der universellen Eigenschaft des Quotienten ist das Paar (A, α) eine n-te äußere Potenz von M. □

Notation 10.2.5. Wegen der wesentlichen Eindeutigkeit der symmetrischen und äußeren Potenzen schreiben wir

$$\mathrm{Sym}^n_R(M) \quad \text{bzw.} \quad \Lambda^n_R(M)$$

für „die“ n-te symmetrische bzw. äußere Potenz von M. Wir schreiben weiter

$$\begin{aligned} \sigma\colon M^n &\to \mathrm{Sym}^n_R(M), & \alpha\colon M^n &\to \Lambda^n_R(M), \\ (x_1, \ldots, x_n) &\mapsto x_1 \cdot \ldots \cdot x_n, & (x_1, \ldots, x_n) &\mapsto x_1 \wedge \cdots \wedge x_n, \end{aligned}$$

für die universelle symmetrische bzw. alternierende n-lineare Abbildung auf M^n.

Elemente von $\mathrm{Sym}^n_R(M)$ heißen *symmetrische Tensoren*, und die im Bild von σ heißen *rein*. Das Element $x_1 \cdot \ldots \cdot x_n$ heißt das *symmetrische Produkt* von $x_1, \ldots, x_n$.

Elemente von $\Lambda^n_R(M)$ heißen *äußere Tensoren*, und die im Bild von α heißen *rein*. Das Element $x_1 \wedge \cdots \wedge x_n$ heißt das *äußere Produkt* oder das *Dachprodukt* von $x_1, \ldots, x_n$.

Beispiel 10.2.6.

(i) Jede 0-lineare oder 1-lineare Abbildung ist gleichzeitig symmetrisch und alternierend. Damit gilt

$$\mathrm{Sym}^0_R(M) = \Lambda^0_R(M) = \mathrm{T}^0_R(M) = R \quad \text{und} \quad \mathrm{Sym}^1_R(M) = \Lambda^1_R(M) = \mathrm{T}^1_R(M) = M.$$

(ii) Ist M ein zyklischer R-Modul, so ist jede n-lineare Abbildung $M^n \to N$ symmetrisch. In diesem Fall gilt also $\mathrm{Sym}^n_R(M) = \mathrm{T}^n_R(M)$.

Beispiel 10.2.7 (Determinantenfunktionen)**.** Sei M ein freier R-Modul vom Rang $n \in \mathbb{N}$. Eine Determinantenfunktion auf M ist eine alternierende n-lineare Form $M^n \to R$ (Definition 5.3.16). Nach der universellen Eigenschaft der äußeren Potenz ist dies äquivalent zu einer Linearform $\Lambda^n_R(M) \to R$. Das heißt, es gibt einen kanonischen Isomorphismus

$$\mathrm{Det}(M) \cong \Lambda^n_R(M)^*$$

zwischen dem R-Modul von Determinantenfunktionen auf M und dem Dualmodul der n-ten äußeren Potenz von M. Die n-te äußere Potenz $\Lambda^n_R(M)$ wird auch als *Determinante* von M und mit $\det(M)$ bezeichnet.

Bemerkung 10.2.8 (Funktorialität der symmetrischen/äußeren Potenzen)**.** Sei R ein kommutativer Ring, sei $n \in \mathbb{N}$ und sei $f\colon M \to N$ eine R-lineare Abbildung. Die Abbildung

$$M^n \to \mathrm{Sym}^n_R(N), \quad (x_1, \dots, x_n) \mapsto f(x_1) \cdot \ldots \cdot f(x_n),$$

ist n-linear und symmetrisch, und damit induziert eine R-lineare Abbildung

$$\mathrm{Sym}^n(f)\colon \mathrm{Sym}^n_R(M) \to \mathrm{Sym}^n_R(N), \quad x_1 \cdot \ldots \cdot x_n \mapsto f(x_1) \cdot \ldots \cdot f(x_n).$$

Auf ähnliche Weise gibt es eine R-lineare Abbildung

$$\Lambda^n(f)\colon \Lambda^n_R(M) \to \Lambda^n_R(N), \quad x_1 \wedge \cdots \wedge x_n \mapsto f(x_1) \wedge \cdots \wedge f(x_n).$$

Beide Konstruktionen sind zudem kompatibel mit der Komposition von R-linearen Abbildungen, und sie erhalten Identitätsabbildungen und Isomorphismen (siehe Proposition 10.1.8).

Proposition 10.2.9 (Erzeugung durch reine Tensoren)**.** *Sei R ein kommutativer Ring, M ein R-Modul und $n \in \mathbb{N}$. Dann sind die R-Moduln $\mathrm{Sym}^n_R(M)$ und $\Lambda^n_R(M)$ von reinen Tensoren erzeugt. Wenn $n \geq 1$ ist sogar jeder symmetrische/äußere Tensor eine Summe von reinen symmetrischen/äußeren Tensoren.*

Beweis. Nach Proposition 10.1.15 genügt es zu zeigen, dass die kanonischen Abbildungen $\mathrm{T}^n_R(M) \to \mathrm{Sym}^n_R(M)$ und $\mathrm{T}^n_R(M) \to \Lambda^n_R(M)$ surjektiv sind. Dies folgt aus den expliziten Konstruktionen im Satz 10.2.4, aber man kann auch einen direkten Beweis geben. Die Abbildung $\mathrm{T}^n_R(M) \to \mathrm{Sym}^n_R(M)$ ist genau dann surjektiv, wenn ihr Kokern Q gleich $\{0\}$ ist. Die Quotientenabbildung $q\colon \mathrm{Sym}^n_R(M) \twoheadrightarrow Q$ entspricht einer symmetrischen n-linearen Abbildung $f\colon M^n \to Q$, deren zugehörige lineare Abbildung $\mathrm{T}^n_R(M) \to Q$ null ist. Damit ist f null und insbesondere ist f die Komposition der universellen Abbildung $\sigma\colon M^n \to \mathrm{Sym}^n_R(M)$ mit der Nullabbildung $0\colon \mathrm{Sym}^n_R(M) \to Q$. Aus der universellen Eigenschaft von σ folgt, dass $q = 0$, und daher dass $Q = \{0\}$. Der Beweis für die äußere Potenz ist ähnlich. □

Proposition 10.2.10 (Dimension der symmetrischen/äußeren Potenz)**.** *Sei R ein kommutativer Ring, $n \in \mathbb{N}$ und M ein freier R-Modul mit Basis $(b_i)_{i \in I}$. Sei $\leq$ eine beliebige totale Ordnung auf I (die nach Korollar 1.4.23 existiert). Man definiert die folgenden Teilmengen von I^n:*

$$I^n_{\leq} = \{(i_1, \dots, i_n) \in I^n \mid i_1 \leq \cdots \leq i_n\},$$
$$I^n_{<} = \{(i_1, \dots, i_n) \in I^n \mid i_1 < \cdots < i_n\}.$$

(i) *$\mathrm{Sym}^n_R(M)$ ist ein freier R-Modul mit Basis $(b_{i_1} \dots b_{i_n})_{(i_1,\dots,i_n) \in I^n_{\leq}}$. Insbesondere: Ist M frei vom Rang $d \in \mathbb{N}$, so ist $\mathrm{Sym}^n_R(M)$ frei vom Rang $\binom{d+n-1}{n}$.*

(ii) *$\Lambda^n_R(M)$ ist ein freier R-Modul mit Basis $(b_{i_1} \wedge \cdots \wedge b_{i_n})_{(i_1,\dots,i_n) \in I^n_{<}}$. Insbesondere: Ist M frei vom Rang $d \in \mathbb{N}$, so ist $\Lambda^n_R(M)$ frei vom Rang $\binom{d}{n}$.*

Beweis. Zu (i). Nach Proposition 10.2.9 ist $\{b_{i_1} \dots b_{i_n} \mid (i_1, \dots, i_n) \in I^n\}$ ein Erzeugendensystem von $\mathrm{Sym}^n_R(M)$. Das symmetrische Produkt $b_{i_1} \dots b_{i_n}$ ist aber nach Definition unabhängig von der Reihenfolge, so dass die gegebene Familie erzeugend ist. Zur linearen Unabhängigkeit betrachten wir die Abbildung

$$o\colon I^n \to I^n_{\leq}$$

die ein n-Tupel $(i_1, \dots, i_n)$ in aufsteigender Reihenfolge anordnet. Für jedes $\sigma \in S_n$ gilt dann $o(i_1, \dots, i_n) = o(i_{\sigma(1)}, \dots, i_{\sigma(n)})$. Sei $F = R^{(I^n_{\leq})}$ und sei

$$f\colon M^n \to F$$

die n-lineare Abbildung, die $(b_{i_1}, \dots, b_{i_n})$ auf $e_{o(i_1,\dots,i_n)}$ abbildet. Wir behaupten, dass f symmetrisch ist. Seien $x_1, \dots, x_n \in M$ und sei $\sigma \in S_n$ eine Permutation. Ist $x_j = \sum_{i \in I} r_{ji} b_i$, so gilt

$$f(x_1, \dots, x_n) = \sum_{(i_1,\dots,i_n) \in I^n} r_{1i_1} \dots r_{ni_n} e_{o(i_1,\dots,i_n)}.$$

Durch die Permutation der Indexmenge $I^n \xrightarrow{\sim} I^n$, $(i_1, \dots, i_n) \mapsto (i_{\sigma^{-1}(1)}, \cdots, i_{\sigma^{-1}(n)})$, erhalten wir

$$\begin{aligned} f(x_1, \dots, x_n) &= \sum_{(i_1,\dots,i_n) \in I^n} r_{1i_{\sigma^{-1}(1)}} \dots r_{ni_{\sigma^{-1}(n)}} e_{o(i_{\sigma^{-1}(1)},\dots,i_{\sigma^{-1}(n)})} \\ &= \sum_{(i_1,\dots,i_n) \in I^n} r_{\sigma(1)i_1} \dots r_{\sigma(n)i_n} e_{o(i_{\sigma^{-1}(1)},\dots,i_{\sigma^{-1}(n)})} \\ &= \sum_{(i_1,\dots,i_n) \in I^n} r_{\sigma(1)i_1} \dots r_{\sigma(n)i_n} e_{o(i_1,\dots,i_n)} \\ &= f(x_{\sigma(1)}, \dots, x_{\sigma(n)}), \end{aligned}$$

wie behauptet. Damit induziert f eine lineare Abbildung

$$\mathrm{Sym}^n_R(M) \to F,$$

die die gegebene Familie $(b_{i_1} \dots b_{i_n})_{(i_1,\dots,i_n) \in I^n_{\leq}}$ auf die Standardbasis von F abbildet. Nach Proposition 4.1.20(ii) ist die Familie linear unabhängig.

Zu (ii). Nach Proposition 10.2.9 ist $\{b_{i_1} \wedge \dots \wedge b_{i_n} \mid (i_1, \dots, i_n) \in I^n\}$ ein Erzeugendensystem von $\Lambda^n_R(M)$. Das äußere Produkt $b_{i_1} \wedge \dots \wedge b_{i_n}$ ist unabhängig von der Reihenfolge bis auf ein Vorzeichen, und es ist null, wenn zwei Indizes gleich sind, so dass die gegebene Familie erzeugend ist. Sei $o\colon I^n \to I^n_{\leq}$ wie im Beweis von (i) und sei $F = R^{(I^n_{<})}$. Zu jedem $(i_1, \dots, i_n) \in I^n$ definiert man

$$e^{\pm}(i_1, \dots, i_n) \in F$$

wie folgt: Falls das Tupel $(i_1, \dots, i_n)$ aus paarweise verschiedenen Indizes besteht, dann gibt es genau eine Permutation $\sigma \in S_n$ mit $(i_{\sigma(1)}, \dots, i_{\sigma(n)}) \in I^n_{<}$, und man setzt $e^{\pm}(i_1, \dots, i_n) = \mathrm{sgn}(\sigma) e_{o(i_1,\dots,i_n)}$. Sonst setzt man $e^{\pm}(i_1, \dots, i_n) = 0$. Sei nun

$$f\colon M^n \to F$$

die n-lineare Abbildung, die $(b_{i_1}, \dots, b_{i_n})$ auf $e^{\pm}(i_1, \dots, i_n)$ abbildet. Wir behaupten, dass f alternierend ist. Seien $x_1, \dots, x_n \in M$ und seien $k < l$ Indizes mit $x_k = x_l$. Ist $x_j = \sum_{i \in I} r_{ji} b_i$, so gilt

$$f(x_1, \dots, x_n) = \sum_{(i_1,\dots,i_n) \in I^n} r_{1i_1} \dots r_{ni_n} e^{\pm}(i_1, \dots, i_n).$$

In dieser Summe sind alle Summanden mit $i_k = i_l$ gleich Null, und falls $i_k < i_l$ ist der $(i_1, \dots, i_k, \dots, i_l, \dots, i_n)$-indizierte Summand gleich dem $(i_1, \dots, i_l, \dots, i_k, \dots, i_n)$-indizierten Summand aber mit einem anderen Vorzeichen (da $\operatorname{sgn}(k\ l) = -1$). Deswegen ist die ganze Summe gleich Null, wie behauptet. Damit induziert f eine lineare Abbildung

$$\Lambda^n_R(M) \to F,$$

die die gegebene Familie $(b_{i_1} \wedge \dots \wedge b_{i_n})_{(i_1,\dots,i_n)\in I^n_<}$ auf die Standardbasis von F abbildet. Nach Proposition 4.1.20(ii) ist die Familie linear unabhängig. □

Korollar 10.2.11 (Rang freier Moduln über kommutativen Ringen)**.** *Sei $R \neq \{0\}$ ein kommutativer Ring und seien $n, m \in \mathbb{N}$. Sind die R-Moduln R^n und R^m isomorph, so gilt $n = m$.*

Beweis. Sei $n < m$. Nach Proposition 10.2.10(ii) gilt $\Lambda^m_R(R^n) = \{0\}$ und $\Lambda^m_R(R^m) \cong R$. Da $R \neq \{0\}$ sind die R-Moduln R und $\{0\}$ nicht isomorph. Deswegen sind R^n und R^m nicht isomorph. □

Beispiel 10.2.12. Sei M ein freier R-Modul vom Rang $n \in \mathbb{N}$ mit Basis $B = (b_1, \dots, b_n)$. Nach Proposition 10.2.10(ii) ist $\Lambda^n_R(M)$ frei vom Rang $\binom{n}{n} = 1$ mit Basis $(b_1 \wedge \dots \wedge b_n)$. Die duale Basis des Dualmoduls $\Lambda^n_R(M)^* \cong \operatorname{Det}(M)$ besteht aus der Determinantenfunktion Δ_B aus Satz 5.3.21, denn $\Delta_B(b_1, \dots, b_n) = 1$.

Bemerkung 10.2.13 (äußere Potenzen und die Determinante)**.** Sei M ein freier R-Modul vom Rang $n \in \mathbb{N}$ und sei $f \in \operatorname{End}_R(M)$ ein Endomorphismus. Nach Proposition 10.2.10(ii) ist $\Lambda^n_R(M)$ zu R isomorph, so dass die lineare Abbildung

$$R \to \operatorname{End}_R(\Lambda^n_R(M)), \quad r \mapsto (x \mapsto r \cdot x),$$

ein Isomorphismus ist. Dadurch ist der Endomorphismus $\Lambda^n(f)$ von $\Lambda^n_R(M)$ gleich der Multiplikation mit einem Element von R. Dieses Element ist genau die Determinante $\det(f) \in R$, nach Definition 5.3.46 und dem Isomorphismus $\Lambda^n_R(M)^* \cong \operatorname{Det}(M)$ (siehe Beispiel 10.2.7).

Bemerkung 10.2.14 (äußere Potenzen und Minoren)**.** Seien $m, n, r \in \mathbb{N}$ und sei $A \in M_{m\times n}(R)$. Die Darstellungsmatrix von $\Lambda^r(L_A)\colon \Lambda^r_R(R^n) \to \Lambda^r_R(R^m)$ bezüglich der Basen aus Proposition 10.2.10(ii) besteht genau aus den *Minoren r-ter Ordnung* von A, d.h., den Determinanten der $r \times r$-Untermatrizen von A. Denn sei $e_{j_1} \wedge \dots \wedge e_{j_r}$ ein Basisvektor von $\Lambda^r_R(R^n)$, $e_{i_1} \wedge \dots \wedge e_{i_r}$ ein Basisvektor von $\Lambda^r_R(R^m)$ und $B \in M_r(R)$ die Einschränkung von A auf $\{i_1, \dots, i_r\} \times \{j_1, \dots, j_r\}$. Die lineare Abbildung L_B ist dann gleich der Komposition

$$R^r = R^{\{j_1,\dots,j_r\}} \hookrightarrow R^n \xrightarrow{L_A} R^m \twoheadrightarrow R^{\{i_1,\dots,i_r\}} = R^r,$$

und $\Lambda^r(L_B)$ ist gleich Multiplikation mit dem entsprechenden Koeffizient der Darstellungsmatrix von $\Lambda^r(L_A)$. Nach Bemerkung 10.2.13 ist dieser Koeffizient gleich $\det(B)$.

Bemerkung 10.2.15 (äußere Potenzen und Elementerteiler)**.** Eine R-lineare Abbildung $f\colon N \to M$ definiert ein kanonisches Ideal $I_f \subset R$, nämlich das Bild der linearen Abbildung

$$e_f\colon N \otimes_R M^* \to R, \quad x \otimes \alpha \mapsto \alpha(f(x)).$$

Das Ideal I_f ist eine Isomorphie-Invariante von f im folgenden Sinne: Sind $\psi\colon N \xrightarrow{\sim} N'$ und $\varphi\colon M \xrightarrow{\sim} M'$ Isomorphismen, so gilt $e_{\varphi\circ f\circ\psi^{-1}} = e_f \circ (\psi^{-1} \otimes \varphi^*)$ und damit $I_f = I_{\varphi\circ f\circ\psi^{-1}}$. Sind zum Beispiel M und N frei und endlich erzeugt, so wird I_f von den Koeffizienten einer Darstellungsmatrix von f erzeugt. Nach Bemerkung 10.2.14 wird in diesem Fall $I_{\Lambda^i(f)}$ von den Minoren i-ter Ordnung einer Darstellungsmatrix von f erzeugt. Im Allgemeinen heißen die Ideale $I_{\Lambda^i(f)}$ mit $i \in \mathbb{N}$ die *Determinantenideale* oder *Fitting-Ideale* von f.

Die Determinantenideale sind eine Verallgemeinerung der Elementarteiler aus Satz 8.3.21 auf beliebige kommutative Ringe. Denn sei R ein Hauptidealring und seien M und N frei und endlich erzeugt. Seien $d_1|d_2|\dots|d_r$ die Elementarteiler von f. Man setzt ferner $d_i = 0$ für $i > r$. Dann gilt $I_{\Lambda^i(f)} = (d_1 \cdot \ldots \cdot d_i)$ für alle $i \in \mathbb{N}$, da das Produkt $d_1 \cdot \ldots \cdot d_i$ genau der ggT der Minoren i-ter Ordnung einer Smith-Normalform von f ist. Dies gibt einen neuen Beweis der Eindeutigkeit der Elementarteiler und zeigt zudem die folgende Aussage: Das Produkt der ersten i Elementarteiler einer Matrix A ist genau der ggT der Minoren i-ter Ordnung von A.

Proposition 10.2.16 (äußere Potenzen und Dualität)**.** *Sei R ein kommutativer Ring, M ein R-Modul und $n \in \mathbb{N}$. Dann gibt es eine wohldefinierte lineare Abbildung*

$$\begin{aligned}\Theta\colon \Lambda_R^n(M^*) &\to \Lambda_R^n(M)^*,\\ \alpha_1 \wedge \dots \wedge \alpha_n &\mapsto \left(x_1 \wedge \dots \wedge x_n \mapsto \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \alpha_{\sigma(1)}(x_1) \dots \alpha_{\sigma(n)}(x_n)\right).\end{aligned}$$

Sie ist ein Isomorphismus, wenn M frei und endlich erzeugt ist.

Beweis. Die Abbildung Θ haben wir schon im Beweis der Proposition 5.3.48(iv) definiert. Sei nun M frei und endlich erzeugt. Sei $(b_1, \dots, b_d)$ eine Basis von M und sei $(\beta_1, \dots, \beta_d)$ die duale Basis von M^*. Nach Proposition 10.2.10(ii) erhalten wir induzierte Basen von $\Lambda_R^n(M)$ und $\Lambda_R^n(M^*)$. Für $1 \le i_1 < \dots < i_n \le d$ und $1 \le j_1 < \dots < j_n \le d$ gilt

$$\Theta(\beta_{i_1} \wedge \dots \wedge \beta_{i_n})(b_{j_1} \wedge \dots \wedge b_{j_n}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \delta_{i_{\sigma(1)} j_1} \dots \delta_{i_{\sigma(n)} j_n} = \delta_{i_1 j_1} \dots \delta_{i_n j_n}.$$

Das heißt, Θ bildet die Basis von $\Lambda_R^n(M^*)$ auf die duale Basis zur Basis von $\Lambda_R^n(M)$ ab. Insbesondere ist Θ ein Isomorphismus. □

Bemerkung 10.2.17 (symmetrische Potenzen und Dualität)**.** Das Analogon der Proposition 10.2.16 für symmetrische Potenzen ist komplizierter: Es gibt nämlich eine dritte Konstruktion $\Gamma_R^n(M)$, die n-te *dividierte Potenz* von M, und kanonische lineare Abbildungen

$$\operatorname{Sym}_R^n(M^*) \to \Gamma_R^n(M)^* \quad \text{und} \quad \Gamma_R^n(M^*) \to \operatorname{Sym}_R^n(M)^*,$$

die Isomorphismen sind, wenn M frei und endlich erzeugt ist. Der R-Modul $\Gamma_R^n(M)$ wird auch durch eine universelle Eigenschaft definiert, aber er ist kein Quotient der Tensorpotenz.

Bemerkung 10.2.18 (Potenzen von Summen)**.** Seien M und N Moduln über R und sei $n \in \mathbb{N}$. Die Tensorpotenz $\mathrm{T}_R^n(M \oplus N)$ kann man durch n Anwendungen des Distributivgesetzes berechnen: Es ergibt sich den „binomischen Lehrsatz“

$$\mathrm{T}_R^n(M \oplus N) \cong \bigoplus_{p+q=n} (\mathrm{T}_R^p(M) \otimes_R \mathrm{T}_R^q(N))^{\binom{n}{p}}.$$

Für die symmetrischen une äußeren Potenzen haben wir stattdessen die Formeln

$$\begin{aligned}\operatorname{Sym}_R^n(M \oplus N) &\cong \bigoplus_{p+q=n} \operatorname{Sym}_R^p(M) \otimes_R \operatorname{Sym}_R^q(N),\\ \Lambda_R^n(M \oplus N) &\cong \bigoplus_{p+q=n} \Lambda_R^p(M) \otimes_R \Lambda_R^q(N).\end{aligned}$$

Falls M und N frei sind lassen sich diese Isomorphismen aus Proposition 10.2.10 herleiten; im Allgemeinen braucht man eine Abwandlung des Beweises. Diese Formeln verallgemeinern sich auf direkte Summen mit beliebiger Indexmenge I: In diesem Fall haben wir auf der rechten Seite direkte Summen über die Menge $I_{\le}^n$ wie in Proposition 10.2.10.

Bemerkung 10.2.19 (antisymmetrische Potenzen)**.** Man könnte auch antisymmetrische Potenzen definieren. Die haben aber keine guten formalen Eigenschaften und werden deshalb selten benutzt. Es gibt zum Beispiel kein antisymmetrisches Analogon der Proposition 10.2.10, denn: Wenn 2 eine Einheit in R ist, ist die antisymmetrische Potenz gleich der äußeren Potenz, aber wenn $2 = 0$ in R, ist sie gleich der symmetrischen Potenz. Im Allgemeinen, wenn 2 weder eine Einheit noch null in R ist, kann eine antisymmetrische Potenz eines freien R-Moduls nicht einmal frei sein: Die zweite antisymmetrische Potenz des $\mathbb{Z}$-Moduls $\mathbb{Z}$ ist isomorph zu $\mathbb{Z}/2\mathbb{Z}$.

10.2.1 Symmetrische und äußere Algebren

Analog der Tensoralgebra können wir die symmetrische Algebra und die äußere Algebra eines Moduls definieren.

Sei R ein kommutativer Ring und sei M ein R-Modul. Für alle $p, q \in \mathbb{N}$ gibt es eine wohldefinierte lineare Abbildung

$$\begin{aligned} m_{p,q} \colon \operatorname{Sym}_R^p(M) \otimes_R \operatorname{Sym}_R^q(M) &\to \operatorname{Sym}_R^{p+q}(M), \\ (x_1 \dots x_p) \otimes (y_1 \dots y_q) &\mapsto x_1 \dots x_p y_1 \dots y_q. \end{aligned}$$

Denn nach dem Exponentialgesetz ist dies äquivalent zu einer linearen Abbildung

$$\operatorname{Sym}_R^p(M) \to \operatorname{Hom}_R(\operatorname{Sym}_R^q(M), \operatorname{Sym}_R^{p+q}(M)),$$

die man wie folgt definieren kann: Man beginnt mit der Abbildung

$$f \colon M^p \to \operatorname{Abb}(M^q, \operatorname{Sym}_R^{p+q}(M)), \quad (x_1, \dots, x_p) \mapsto ((y_1, \dots, y_q) \mapsto x_1 \dots x_p y_1 \dots y_q).$$

Jeder Wert $f(x_1, \dots, x_p)$ ist eine symmetrische q-lineare Abbildung, und f selbst ist p-linear und symmetrisch. Nach der universellen Eigenschaft der symmetrischen Potenzen erhalten wir aufeinander folgende Faktorisierungen von f:

$$\begin{array}{ccc} M^p & \xrightarrow{\quad f \quad} & \operatorname{Abb}(M^q, \operatorname{Sym}_R^{p+q}(M)) \\ \downarrow & \searrow & \uparrow \\ \operatorname{Sym}_R^p(M) & \dashrightarrow & \operatorname{Hom}_R(\operatorname{Sym}_R^q(M), \operatorname{Sym}_R^{p+q}(M)). \end{array}$$

Setzt man

$$\operatorname{Sym}_R^*(M) := \bigoplus_{n \in \mathbb{N}} \operatorname{Sym}_R^n(M),$$

so definieren die Abbildungen $m_{p,q}$ eine Verknüpfung

$$m \colon \operatorname{Sym}_R^*(M) \otimes_R \operatorname{Sym}_R^*(M) \to \operatorname{Sym}_R^*(M).$$

Zusammen mit der kanonischen Abbildung $\iota_0 \colon R = \operatorname{Sym}_R^0(M) \hookrightarrow \operatorname{Sym}_R^*(M)$ erhalten wir eine R-Algebra $(\operatorname{Sym}_R^*(M), m, \iota_0)$, die als *symmetrische Algebra* von M bezeichnet wird. Im Gegensatz zur Tensoralgebra $\mathrm{T}_R^*(M)$ ist die symmetrische Algebra $\operatorname{Sym}_R^*(M)$ stets kommutativ.

Proposition 10.2.20 (universelle Eigenschaft der symmetrischen Algebra)**.** *Sei R ein kommutativer Ring und sei M ein R-Modul. Zu jeder R-Algebra A und jeder R-linearen Abbildung $f \colon M \to A$, so dass für alle $x, y \in M$ gilt $f(x)f(y) = f(y)f(x)$, gibt es* genau einen *R-Algebrenhomomorphismus $\hat{f} \colon \operatorname{Sym}_R^*(M) \to A$ mit $\hat{f} \circ \iota_1 = f$.*

Beweis. Die Eindeutigkeit ist klar, denn $\hat{f}$ muss das symmetrische Produkt $x_1 \dots x_n \in \operatorname{Sym}_R^n(M)$ auf $f(x_1) \dots f(x_n)$ abbilden. Für jedes $n \in \mathbb{N}$ ist die n-lineare Abbildung

$$M^n \to A, \quad (x_1, \dots, x_n) \mapsto f(x_1) \dots f(x_n)$$

symmetrisch, da die Elemente $f(x_i)$ miteinander kommutieren. Es gibt also eine lineare Abbildung

$$\hat{f}_n \colon \mathrm{Sym}^n_R(M) \to A, \quad x_1 \dots x_n \mapsto f(x_1) \dots f(x_n).$$

Sei $\hat{f} \colon \mathrm{Sym}^*_R(M) \to A$ die lineare Abbildung mit $\hat{f} \circ \iota_n = \hat{f}_n$. Es bleibt zu zeigen, dass $\hat{f}$ ein Ringhomomorphismus ist. Es gilt $\hat{f}(1) = \hat{f}_0(1) = 1$ nach Definition. Die Formel für $\hat{f}_n$ zeigt, dass $\hat{f}(x \cdot y) = \hat{f}(x) \cdot \hat{f}(y)$ gilt, wenn x und y reine symmetrische Tensoren sind. Da diese Tensoren $\mathrm{Sym}^*_R(M)$ erzeugen (nach Proposition 10.2.9), gilt dies auch für beliebige x, y. □

Bemerkung 10.2.21 (Polynomalgebren sind symmetrische Algebren)**.** Für jeden R-Modul M ist die R-Algebra $\mathrm{Sym}^*_R(M)$ kommutativ, und nach Proposition 10.2.20 ist sie sogar die universelle kommutative R-Algebra mit einer R-linearen Abbildung von M. Falls $M = R^n$ ist insbesondere $\mathrm{Sym}^*_R(R^n)$ die universelle kommutative R-Algebra mit n gegebenen Elementen. Die Polynomalgebra $R[T_1, \dots, T_n]$ hat dieselbe universelle Eigenschaft (indem man die universelle Eigenschaft des Polynomringes in einer Variablen n mal anwendet), so dass es einen Isomorphismus

$$R[T_1, \dots, T_n] \xrightarrow{\sim} \mathrm{Sym}^*_R(R^n), \quad T_i \mapsto e_i \in R^n = \mathrm{Sym}^1_R(R^n),$$

gibt. Als Realitätscheck kann man bemerken, dass Polynome in n Variablen vom totalen Grad k einen freien R-Modul mit Basis $(T_{i_1} \dots T_{i_k})_{1 \le i_1 \le \dots \le i_k \le n}$ bilden, in Übereinstimmung mit Proposition 10.2.10(i). Für beliebiges M kann man also die symmetrische Algebra $\mathrm{Sym}^*_R(M)$ als Verallgemeinerung der Polynomalgebra in mehreren Variablen auffassen.

Auf ähnliche Weise gibt es lineare Abbildungen

$$\Lambda^p_R(M) \otimes_R \Lambda^q_R(M) \to \Lambda^{p+q}_R(M),$$
$$(x_1 \wedge \dots \wedge x_p) \otimes (y_1 \wedge \dots \wedge y_q) \mapsto x_1 \wedge \dots \wedge x_p \wedge y_1 \wedge \dots \wedge y_q.$$

Sie definieren auf der direkten Summe

$$\Lambda^*_R(M) := \bigoplus_{n \in \mathbb{N}} \Lambda^n_R(M)$$

eine Struktur von R-Algebra, die als *äußere Algebra* oder *Graßmann-Algebra* von M bezeichnet wird. Die Multiplikation auf $\Lambda^*_R(M)$ wird auch mit dem Symbol $\wedge$ geschrieben.

Proposition 10.2.22 (universelle Eigenschaft der äußeren Algebra)**.** *Sei R ein kommutativer Ring und sei M ein R-Modul. Zu jeder R-Algebra A und jeder R-linearen Abbildung $f \colon M \to A$, so dass für alle $x \in M$ gilt $f(x)^2 = 0$, gibt es* genau einen *R-Algebrenhomomorphismus $\hat{f} \colon \Lambda^*_R(M) \to A$ mit $\hat{f} \circ \iota_1 = f$.*

Beweis. Die Eindeutigkeit ist klar, denn $\hat{f}$ muss das äußere Produkt $x_1 \wedge \dots \wedge x_n \in \Lambda^n_R(M)$ auf $f(x_1) \dots f(x_n)$ abbilden. Für jedes $n \in \mathbb{N}$ ist die n-lineare Abbildung

$$M^n \to A, \quad (x_1, \dots, x_n) \mapsto f(x_1) \dots f(x_n)$$

alternierend, nach Proposition 10.2.2(iii). Es gibt also eine lineare Abbildung

$$\hat{f}_n \colon \Lambda^n_R(M) \to A, \quad x_1 \wedge \dots \wedge x_n \mapsto f(x_1) \dots f(x_n).$$

Sei $\hat{f} \colon \Lambda^*_R(M) \to A$ die lineare Abbildung mit $\hat{f} \circ \iota_n = \hat{f}_n$. Es bleibt zu zeigen, dass $\hat{f}$ ein Ringhomomorphismus ist. Es gilt $\hat{f}(1) = \hat{f}_0(1) = 1$ nach Definition. Die Formel für $\hat{f}_n$ zeigt, dass $\hat{f}(x \wedge y) = \hat{f}(x) \cdot \hat{f}(y)$ gilt, wenn x und y reine äußere Tensoren sind. Da diese Tensoren $\Lambda^*_R(M)$ erzeugen (nach Proposition 10.2.9), gilt dies auch für beliebige x, y. □

Beispiel 10.2.23 (Kreuzprodukt)**.** Sei M ein freier R-Modul mit Basis $B = (b_1, \dots, b_n)$ und sei $k \in \{0, \dots, n\}$. Die Basis B induziert Isomorphismen $\lambda_B \colon \Lambda_R^n(M) \xrightarrow{\sim} R$ (Beispiel 10.2.12) und $\varepsilon_B \colon M \xrightarrow{\sim} M^*$ (Proposition 4.1.56(i)). Die Verknüpfung

$$\Lambda_R^k(M) \otimes_R \Lambda_R^{n-k}(M) \to \Lambda_R^n(M) \xrightarrow{\lambda_B} R$$

induziert nach dem Exponentialgesetz eine lineare Abbildung

$$\Lambda_R^k(M) \to \Lambda_R^{n-k}(M)^*.$$

Im Spezialfall $k = n - 1$ erhalten wir eine lineare Abbildung

$$\Lambda_R^{n-1}(M) \to M^* \xrightarrow{\varepsilon_B^{-1}} M,$$

d.h., eine alternierende $(n-1)$-lineare Abbildung

$$M^{n-1} \to M, \quad (x_1, \dots, x_{n-1}) \mapsto x_1 \times \cdots \times x_{n-1},$$

die als *Kreuzprodukt* bzgl. B bezeichnet wird. Nach Konstruktion gilt

$$\varepsilon_B(x_1 \times \cdots \times x_{n-1})(x_n) = \Delta_B(x_1, \dots, x_n),$$

wobei Δ_B die Determinantenfunktion aus Satz 5.3.21 ist. Das gewöhnliche binäre Kreuzprodukt auf R^3 ist der Sonderfall dieser Konstruktion mit der Standardbasis von R^3.

Anhang A

Einführung in die Kategorientheorie

Zu jeder Art von strukturierter Menge (wie Gruppen, K-Vektorräumen, Ringen, usw.), gibt es eine entsprechende Art von strukturerhaltender Abbildung (Gruppenhomomorphismen, K-lineare Abbildungen, Ringhomomorphismen, usw.). Die Praxis der Mathematik zeigt, dass diese Abbildungen eine entscheidende Rolle spielen, wenn man diese strukturierten Mengen verstehen will. Das heißt, die bilden einen wesentlichen Teil der zugehörigen mathematischen Theorie (wie die Gruppentheorie, die lineare Algebra, usw.).

Die Kombination von strukturierten Mengen und strukturerhaltenden Abbildungen wird zum Begriff der *Kategorie* abstrahiert.

A.1 Kategorien

Wir beginnen mit einem kurzen mengentheoretischen Exkurs. Die Kategorientheorie kann leider nicht einfach mit der gewöhnlichen Mengenlehre entwickelt werden, denn die meisten Kategorien sind „zu groß“, um sie mit Mengen beschreiben zu können. Wegen der Russelschen Antinomie 1.2.2 gibt es zum Beispiel keine Menge aller Mengen, aber wir möchten trotzdem von der Kategorie aller Mengen sprechen. Damit ist die Zermelo-Fraenkel-Mengenlehre, die wir im Kapitel 1 teilweise besprochen haben, nicht hinreichend für eine formale Entwicklung der Kategorientheorie. Man braucht dazu eine Erweiterung der Zermelo-Fraenkel-Mengenlehre, wie zum Beispiel die *Neumann-Bernays-Gödel-Mengenlehre* (NBG) oder die *Zermelo-Fraenkel-Mengenlehre mit Universen* (ZFU).

Hier werden wir solche mengentheoretischen Schwierigkeiten ignorieren, und wir bezeichnen als *Klasse* eine beliebige Zusammenfassung von mathematischen Objekten, selbst wenn diese Zusammenfassung keine Menge bildet (z.B., die Klasse aller Mengen).

Definition A.1.1 (Kategorie)**.** Eine *Kategorie* $\mathcal{C}$ besteht aus folgenden Zutaten:

- Es gibt eine Klasse $\operatorname{Ob}\mathcal{C}$ von *Objekten* von $\mathcal{C}$.
- Zu jeden zwei Objekten $X, Y \in \operatorname{Ob}\mathcal{C}$ gibt es eine Menge $\operatorname{Mor}_{\mathcal{C}}(X,Y)$ oder $\operatorname{Hom}_{\mathcal{C}}(X,Y)$ von *Morphismen* oder *Pfeilen* von X nach Y. Für einen Morphismus $f \in \operatorname{Mor}_{\mathcal{C}}(X,Y)$ schreibt man $f\colon X \to Y$; X heißt die *Quelle* und Y das *Ziel* von f.
- Zu jeden Morphismen $f\colon X \to Y$ und $g\colon Y \to Z$ gibt es einen Morphismus $g \circ f\colon X \to Z$, die *Komposition* von f und g. Die Komposition soll assoziativ sein, d.h.: Sind $f\colon X \to Y$, $g\colon Y \to Z$ und $h\colon Z \to W$ Morphismen von $\mathcal{C}$, so gilt

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

- Zu jedem Object $X \in \operatorname{Ob} \mathcal{C}$ gibt es ein Morphismus $\mathrm{id}_X \in \mathrm{Mor}_{\mathcal{C}}(X, X)$, die als *Identitätsmorphismus* oder *Identität* auf X bezeichnet wird. Die Identitätsmorphismen sollen neutrale Elemente bzgl. Komposition sein, d.h.: Für alle Morphismen $f\colon X \to Y$ gilt

$$f \circ \mathrm{id}_X = f = \mathrm{id}_Y \circ f.$$

Bemerkung A.1.2. Die Assoziativität der Komposition bedeutet wie üblich, dass man die Komposition von drei oder mehr Morphismen ohne Klammern schreiben darf, z.B.: $h \circ g \circ f$.

Beispiel A.1.3. In dieser Vorlesung sind wir bereits auf mehrere Beispiele von Kategorien gestoßen:

(i) Die Kategorie $\mathcal{S}\mathrm{et}$ von Mengen: Die Objekte sind alle Mengen und die Morphismen sind beliebige Abbildungen.

(ii) Die Kategorie $\mathcal{G}\mathrm{rp}$ von Gruppen, mit Gruppenhomomorphismen als Morphismen.

(iii) Die Kategorie $\mathcal{A}\mathrm{b}$ von abelschen Gruppen, mit Gruppenhomomorphismen als Morphismen.

(iv) Die Kategorie $\mathcal{R}\mathrm{ing}$ von Ringen, mit Ringhomomorphismen als Morphismen.

(v) Die Kategorie $\mathcal{CR}\mathrm{ing}$ von kommutativen Ringen, mit Ringhomomorphismen als Morphismen.

(vi) Die Kategorie $\mathcal{M}\mathrm{od}_R$ von Moduln über einem Ring R, mit Modulhomomorphismen als Morphismen. Wenn K ein Körper ist, schreibt man auch $\mathcal{V}\mathrm{ect}_K = \mathcal{M}\mathrm{od}_K$ für die Kategorie von K-Vektorräumen.

(vii) Die Kategorie $\mathcal{A}\mathrm{lg}_R$ von Algebren über einem kommutativen Ring R, mit Algebrenhomomorphismen als Morphismen.

(viii) Die Kategorie $\mathcal{P}\mathrm{os}$ von partiell geordneten Mengen, mit monotonen Abbildungen als Morphismen.

(ix) Es gibt auch mehrere Beispiele aus der Analysis, wie die Kategorie $\mathcal{M}\mathrm{et}$ von metrischen Räumen mit nichtexpansiven Abbildungen (d.h., Abbildungen mit der Lipschitzkonstante 1) oder die Kategorie $\mathcal{T}\mathrm{op}$ von topologischen Räumen mit stetigen Abbildungen.

Beispiel A.1.4 (Gruppen als Kategorien)**.** Sei G eine Gruppe. Man definiert eine Kategorie $\mathcal{B}G$ wie folgt:

- $\mathcal{B}G$ hat ein einziges Objekt $*$: $\operatorname{Ob} \mathcal{B}G = \{*\}$. (Die Beschaffenheit dieses Objekts spielt keine Rolle.)
- Es gilt $\mathrm{Mor}_{\mathcal{B}G}(*, *) = G$.
- Die Komposition $\mathrm{Mor}_{\mathcal{B}G}(*, *) \times \mathrm{Mor}_{\mathcal{B}G}(*, *) \to \mathrm{Mor}_{\mathcal{B}G}(*, *)$ ist die Verknüpfung der Gruppe G: $g \circ h = g \cdot h$.
- Der Identitätsmorphismus id_* in $\mathcal{B}G$ ist das neutrale Element von G.

Beispiel A.1.5 (partiell geordnete Mengen als Kategorien)**.** Sei X eine Menge versehen mit einer partiellen Ordnung $\leq$ (oder nur einer reflexiven und transitiven Relation). Man definiert eine Kategorie $\mathcal{N}X$ wie folgt:

- Es gilt $\operatorname{Ob} \mathcal{N}X = X$.
- Wenn $x \leq y$ besteht die Menge $\mathrm{Mor}_{\mathcal{N}X}(x, y)$ aus genau einem Morphismus, sonst ist sie leer. (Die Beschaffenheit dieser Morphismen spielt keine Rolle.)

- Die Komposition ist eindeutig bestimmt: Sind $\mathrm{Mor}_{\mathcal{N}X}(x,y)$ und $\mathrm{Mor}_{\mathcal{N}X}(y,z)$ nicht leer, so gilt $x \leq y$ und $y \leq z$ und daher $x \leq z$ nach der Transitivität von $\leq$, so dass $\mathrm{Mor}_{\mathcal{N}X}(x,z)$ eine einelementige Menge ist.
- Die Identitätsmorphismen sind eindeutig bestimmt: Für jedes $x \in X$ gilt $x \leq x$ nach der Reflexivität von $\leq$, so dass $\mathrm{Mor}_{\mathcal{N}X}(x,x)$ eine einelementige Menge ist.

Definition A.1.6 (duale Kategorie)**.** Sei $\mathcal{C}$ eine Kategorie. Die *duale Kategorie* $\mathcal{C}^{\mathrm{op}}$ zu $\mathcal{C}$ ist die gleiche Kategorie aber mit umgekehrter Komposition. Genauer:

- $\mathrm{Ob}\,\mathcal{C}^{\mathrm{op}} = \mathrm{Ob}\,\mathcal{C}$.
- $\mathrm{Mor}_{\mathcal{C}^{\mathrm{op}}}(X,Y) = \mathrm{Mor}_{\mathcal{C}}(Y,X)$.
- Für die Komposition $\circ^{\mathrm{op}}$ in $\mathcal{C}^{\mathrm{op}}$ gilt $f \circ^{\mathrm{op}} g = g \circ f$.
- $\mathcal{C}^{\mathrm{op}}$ hat dieselben Identitätsmorphismen wie $\mathcal{C}$.

Definition A.1.7 (Isomorphismus, isomorph)**.** Sei $\mathcal{C}$ eine Kategorie und seien $X, Y \in \mathrm{Ob}\,\mathcal{C}$.

- Ein Morphismus $f\colon X \to Y$ in $\mathcal{C}$ heißt *Isomorphismus*, und man schreibt $f\colon X \xrightarrow{\sim} Y$, wenn ein Morphismus $g\colon Y \to X$ existiert, so dass $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$. Der Morphismus g ist dann eindeutig durch f bestimmt; er heißt der *Umkehrmorphismus* oder *inverse Morphismus* zu f und wird mit f^{-1} bezeichnet.
- Man sagt, dass X zu Y *isomorph* ist, in Zeichen $X \cong Y$, wenn ein Isomorphismus von X nach Y existiert.

Beispiel A.1.8. In den Kategorien $\mathcal{S}\mathrm{et}$, $\mathcal{G}\mathrm{rp}$, $\mathcal{A}\mathrm{b}$, $\mathcal{R}\mathrm{ing}$, $\mathcal{C}\mathcal{R}\mathrm{ing}$ $\mathcal{M}\mathrm{od}_R$, $\mathcal{A}\mathrm{lg}_R$ und $\mathcal{P}\mathrm{os}$ sind die Isomorphismen genau die bijektiven Morphismen. Dies gilt aber nicht im Allgemeinen: In der Kategorie $\mathcal{T}\mathrm{op}$ ist eine bijektive stetige Abbildung nicht unbedingt ein Isomorphismus, da die Umkehrabbildung nicht unbedingt stetig ist. Ein Beispiel davon ist die bijektive stetige Abbildung $[0, 2\pi) \to S^1$, $t \mapsto (\cos t, \sin t)$.

Proposition A.1.9 (Isomorphie ist eine Äquivalenzrelation)**.** *Sei $\mathcal{C}$ eine Kategorie. Dann ist Isomorphie eine Äquivalenzrelation auf* $\mathrm{Ob}\,\mathcal{C}$.

Beweis. Reflexivität. Der Identitätsmorphismus id_X ist ein Isomorphismus von X nach X, denn $\mathrm{id}_X \circ \mathrm{id}_X = \mathrm{id}_X$.

Symmetrie. Ist $f\colon X \to Y$ ein Isomorphismus von X nach Y, so ist f^{-1} ein Isomorphismus von Y nach X.

Transitivität. Sind $f\colon X \to Y$ und $g\colon Y \to Z$ Isomorphismen, so ist $g \circ f\colon X \to Z$ ein Isomorphismus von X nach Z, mit Umkehrmorphismus $f^{-1} \circ g^{-1}$, denn:

$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = f^{-1} \circ (g^{-1} \circ g) \circ f = f^{-1} \circ \mathrm{id}_Y \circ f = f^{-1} \circ f = \mathrm{id}_X$$

und auf ähnliche Weise $(g \circ f) \circ (f^{-1} \circ g^{-1}) = \mathrm{id}_Z$. □

Definition A.1.10 (Monomorphismus, Epimorphismus)**.** Sei $\mathcal{C}$ eine Kategorie und sei $f\colon X \to Y$ ein Morphismus in $\mathcal{C}$.

- f heißt *Monomorphismus*, wenn die folgende Bedingung erfüllt ist: Für alle $Z \in \mathrm{Ob}\,\mathcal{C}$ und alle Morphismen $g, h\colon Z \to X$ in $\mathcal{C}$, ist $f \circ g = f \circ h$, so folgt bereits $g = h$.
- f heißt *Epimorphismus*, wenn die folgende Bedingung erfüllt ist: Für alle $Z \in \mathrm{Ob}\,\mathcal{C}$ und alle Morphismen $g, h\colon Y \to Z$ in $\mathcal{C}$, ist $g \circ f = h \circ f$, so folgt bereits $g = h$.

Beispiel A.1.11. In den Kategorien $\mathcal{S}\mathrm{et}$, $\mathcal{G}\mathrm{rp}$, $\mathcal{A}\mathrm{b}$, $\mathcal{M}\mathrm{od}_R$, $\mathcal{P}\mathrm{os}$ und $\mathcal{T}\mathrm{op}$ sind die Monomorphismen/Epimorphismen genau die injektiven/surjektiven Morphismen. Dies gilt aber nicht im Allgemeinen: In der Kategorie $\mathcal{R}\mathrm{ing}$ ist die Inklusionsabbildung $\mathbb{Z} \hookrightarrow \mathbb{Q}$ ein Epimorphismus, und in der Kategorie $\mathcal{M}\mathrm{et}$ ist die Inklusionsabbildung $\mathbb{Q} \hookrightarrow \mathbb{R}$ ein Epimorphismus.

Bemerkung A.1.12. Ein Isomorphismus ist stets gleichzeitig ein Monomorphismus und ein Epimorphismus. Die Umkehrung gilt in den Kategorien $\mathcal{S}\mathrm{et}$, $\mathcal{G}\mathrm{rp}$, $\mathcal{A}\mathrm{b}$, $\mathcal{M}\mathrm{od}_R$ und $\mathcal{P}\mathrm{os}$, aber nicht im Allgemeinen. Zum Beispiel:

(i) In der Kategorie $\mathcal{R}\mathrm{ing}$ ist die Inklusionsabbildung $\mathbb{Z} \hookrightarrow \mathbb{Q}$ ein Monomorphismus sowie ein Epimorphismus, aber kein Isomorphismus.

(ii) In der Kategorie $\mathcal{T}\mathrm{op}$ ist die stetige Abbildung $[0, 2\pi) \to S^1$, $t \mapsto (\cos t, \sin t)$, ein Monomorphismus sowie ein Epimorphismus, aber kein Isomorphismus.

Definition A.1.13 (Endomorphismus, Automorphismus)**.** Sei $\mathcal{C}$ eine Kategorie und sei $X \in \operatorname{Ob} \mathcal{C}$.

- Ein *Endomorphismus* von X ist ein Morphismus von X nach X. Man bezeichnet mit $\operatorname{End}_{\mathcal{C}}(X) = \operatorname{Mor}_{\mathcal{C}}(X, X)$ die Menge aller Endomorphismen von X.
- Ein *Automorphismus* von X ist ein Endomorphismus von X, der auch ein Isomorphismus ist. Man bezeichnet mit $\operatorname{Aut}_{\mathcal{C}}(X) \subset \operatorname{End}_{\mathcal{C}}(X)$ die Teilmenge bestehend aus den Automorphismen. Das Paar $(\operatorname{Aut}_{\mathcal{C}}(X), \circ)$ ist dann eine Gruppe und heißt die *Automorphismengruppe* des Objekts X.

Beispiel A.1.14 (Kategorien von Morphismen)**.** Sei $\mathcal{C}$ eine Kategorie. Man definiert wie folgt die Kategorie $\mathcal{M}\mathrm{or}(\mathcal{C})$ von Morphismen von $\mathcal{C}$:

- Die Objekte von $\mathcal{M}\mathrm{or}(\mathcal{C})$ sind die Morphismen von $\mathcal{C}$. Genauer gilt:
$$\operatorname{Ob} \mathcal{M}\mathrm{or}(\mathcal{C}) = \{(X, Y, f) \mid X, Y \in \operatorname{Ob} \mathcal{C} \text{ und } f \in \operatorname{Mor}_{\mathcal{C}}(X, Y)\}.$$
- Sind $f\colon X \to Y$ und $g\colon Z \to W$ Morphismen in $\mathcal{C}$, so gilt:
$$\operatorname{Mor}_{\mathcal{M}\mathrm{or}(\mathcal{C})}(f, g) = \{(h, k) \in \operatorname{Mor}_{\mathcal{C}}(X, Z) \times \operatorname{Mor}_{\mathcal{C}}(Y, W) \mid g \circ h = k \circ f\}.$$
Anders gesagt ist ein Morphismus von f nach g in $\mathcal{M}\mathrm{or}(\mathcal{C})$ ein kommutatives Quadrat
$$\begin{array}{ccc} X & \xrightarrow{h} & Z \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ Y & \xrightarrow{k} & W. \end{array}$$
- Komposition wird komponentenweise definiert: $(h', k') \circ (h, k) = (h' \circ h, k' \circ k)$.
- Der Identitätsmorphismus auf $f\colon X \to Y$ ist $(\mathrm{id}_X, \mathrm{id}_Y)$.

Man definiert auf ähnliche Weise die Kategorie $\mathcal{E}\mathrm{nd}(\mathcal{C})$ von Endomorphismen von $\mathcal{C}$: Die Objekte sind jetzt Paare (X, f) mit $X \in \operatorname{Ob} \mathcal{C}$ und $f \in \operatorname{End}_{\mathcal{C}}(X)$, und ein Morphismus von (X, f) nach (Y, g) ist ein Morphismus $h\colon X \to Y$ in $\mathcal{C}$, so dass $g \circ h = h \circ f$.

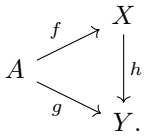
Beispiel A.1.15 (Kommakategorien)**.** Sei $\mathcal{C}$ eine Kategorie und $A \in \operatorname{Ob} \mathcal{C}$. Die Kategorie $\mathcal{C}_{/A}$ von *Objekten über* A ist wie folgt definiert:

- Objekte von $\mathcal{C}_{/A}$ sind Paare (X, f) bestehend aus einem Objekt $X \in \operatorname{Ob} \mathcal{C}$ und einem Morphismus $f\colon X \to A$ in $\mathcal{C}$.
- Ein Morphismus in $\mathcal{C}_{/A}$ von (X, f) nach (Y, g) ist ein Morphismus $h\colon X \to Y$ in $\mathcal{C}$ mit $f = g \circ h$, d.h., so dass folgendes Dreieck kommutiert:
$$\begin{array}{ccc} X & & \\ {\scriptstyle h}\downarrow & \searrow^{f} & \\ Y & \nearrow_{g} & A. \end{array}$$

- Die Komposition und Identitätsmorphismen in $\mathcal{C}_{/A}$ sind wie in $\mathcal{C}$.

Die Kategorie $\mathcal{C}_{A/}$ von *Objekten unter* A ist wie folgt definiert:

- Objekte von $\mathcal{C}_{A/}$ sind Paare (X, f) bestehend aus einem Objekt $X \in \operatorname{Ob} \mathcal{C}$ und einem Morphismus $f\colon A \to X$ in $\mathcal{C}$.
- Ein Morphismus in $\mathcal{C}_{A/}$ von (X, f) nach (Y, g) ist ein Morphismus $h\colon X \to Y$ in $\mathcal{C}$ mit $h \circ f = g$, d.h., so dass folgendes Dreieck kommutiert:



- Die Komposition und Identitätsmorphismen in $\mathcal{C}_{A/}$ sind wie in $\mathcal{C}$.

A.1.1 Produkte und Summen

Definition A.1.16 (Produkt, Summe)**.** Sei $(X_i)_{i \in I}$ eine Familie von Objekten in einer Kategorie $\mathcal{C}$.

- Ein *Produkt* der Familie $(X_i)_{i \in I}$ in $\mathcal{C}$ besteht aus einem Objekt $P \in \operatorname{Ob} \mathcal{C}$ und Morphismen $\pi_i\colon P \to X_i$ für alle $i \in I$, so dass folgende universelle Eigenschaft erfüllt ist: Zu jedem Objekt Y und jeder Familie von Morphismen $(f_i\colon Y \to X_i)$ gibt es *genau einen* Morphismus $f\colon Y \to P$ mit $\pi_i \circ f = f_i$ für alle $i \in I$.
- Eine *Summe* oder ein *Koprodukt* der Familie $(X_i)_{i \in I}$ in $\mathcal{C}$ besteht aus einem Objekt $S \in \operatorname{Ob} \mathcal{C}$ und Morphismen $\iota_i\colon X_i \to S$ für alle $i \in I$, so dass folgende universelle Eigenschaft erfüllt ist: Zu jedem Objekt Y und jeder Familie von Morphismen $(f_i\colon X_i \to Y)$ gibt es *genau einen* Morphismus $f\colon S \to Y$ mit $f \circ \iota_i = f_i$ für alle $i \in I$.

Bemerkung A.1.17 (Dualität zwischen Produkten und Summen)**.** Produkte in $\mathcal{C}$ sind Summen in $\mathcal{C}^{\mathrm{op}}$, und Summen in $\mathcal{C}$ sind Produkten in $\mathcal{C}^{\mathrm{op}}$.

Die nächste Proposition ist ein typisches elementares Argument in der Kategorientheorie: Grob gesagt zeigt es, dass Objekte, die durch eine universelle Eigenschaft charakterisiert werden, stets eindeutig bis auf (eindeutige) Isomorphie sind.

Proposition A.1.18 (Eindeutigkeit des Produkts bzw. der Summe)**.** *Sei $(X_i)_{i \in I}$ eine Familie von Objekten in einer Kategorie $\mathcal{C}$.*

(i) *Sind $(P, (\pi_i)_{i \in I})$ und $(P', (\pi'_i)_{i \in I})$ zwei Produkte der Familie $(X_i)_{i \in I}$, so gibt es* genau einen *Morphismus $\varphi\colon P \to P'$, so dass für alle $i \in I$ gilt $\pi'_i \circ \varphi = \pi_i$. Außerdem ist φ ein Isomorphismus.*

(ii) *Sind $(S, (\iota_i)_{i \in I})$ und $(S', (\iota'_i)_{i \in I})$ zwei Summen der Familie $(X_i)_{i \in I}$, so gibt es* genau einen *Morphismus $\varphi\colon S \to S'$, so dass für alle $i \in I$ gilt $\varphi \circ \iota_i = \iota'_i$. Außerdem ist φ ein Isomorphismus.*

Beweis. Wir behandeln nur (i), da der Beweis von (ii) ganz ähnlich ist. Die Existenz und Eindeutigkeit von φ ist ein Sonderfall der universellen Eigenschaft von P'. Nach der universellen Eigenschaft von P gibt es auch genau einem Morphismus $\psi\colon P' \to P$, so dass für alle $i \in I$ gilt $\pi_i \circ \psi = \pi'_i$. Wir behaupten, dass φ und ψ zueinander invers sind. Die Komposition $\psi \circ \varphi\colon P \to P$ erfüllt

$$\pi_i \circ (\psi \circ \varphi) = (\pi_i \circ \psi) \circ \varphi = \pi'_i \circ \varphi = \pi_i = \pi_i \circ \mathrm{id}_P$$

für alle $i \in I$. Aus der universellen Eigenschaft von P folgt, dass $\psi \circ \varphi = \mathrm{id}_P$ gilt. Aus der universellen Eigenschaft von P' folgt auf ähnliche Weise, dass $\varphi \circ \psi = \mathrm{id}_{P'}$ gilt. □

Notation A.1.19. Wegen der Eindeutigkeit bis auf Isomorphie von Produkten und Summen spricht man oft von *dem* Produkt oder *der* Summe einer Familie $(X_i)_{i\in I}$ (wenn es existiert). Man schreibt

$$\prod_{i\in I} X_i \quad \text{bzw.} \quad \coprod_{i\in I} X_i$$

für das Produkt bzw. die Summe der Familie $(X_i)_{i\in I}$ in einer Kategorie $\mathcal{C}$.

Beispiel A.1.20 (Produkte und Summen von Mengen)**.** Sei $(X_i)_{i\in I}$ eine Mengenfamilie. Das kartesische Produkt $\prod_{i\in I} X_i$ aus Definition 1.2.15(iv) ist ein Produkt der Familie $(X_i)_{i\in I}$ in der Kategorie $\mathcal{S}\mathrm{et}$. Die disjunkte Vereinigung $\coprod_{i\in I} X_i$ aus Definition 1.2.15(iii) ist eine Summe der Familie $(X_i)_{i\in I}$ in der Kategorie $\mathcal{S}\mathrm{et}$.

Beispiel A.1.21 (Produkte und Summen von Moduln)**.** Sei R ein Ring und $(M_i)_{i\in I}$ eine Familie von R-Moduln. Das Produkt $\prod_{i\in I} M_i$ ist ein Produkt der Familie $(M_i)_{i\in I}$ in der Kategorie $\mathcal{M}\mathrm{od}_R$. Die direkte Summe $\bigoplus_{i\in I} M_i$ ist eine Summe der Familie $(M_i)_{i\in I}$ in der Kategorie $\mathcal{M}\mathrm{od}_R$. Siehe dazu Proposition 6.1.3.

Bemerkung A.1.22. In „algebraischen“ Kategorien wie $\mathcal{S}\mathrm{et}$, $\mathcal{G}\mathrm{rp}$, $\mathcal{A}\mathrm{b}$, $\mathcal{R}\mathrm{ing}$ und $\mathcal{M}\mathrm{od}_R$ ist das kategorische Produkt immer das kartesische Produkt der unterliegenden Mengen, versehen mit den komponentenweisen Verknüpfungen. Summen sind aber komplizierter: Zum Beispiel ist die unterliegende Menge einer Summe von R-Moduln *nicht* zu der disjunkten Vereinigung der unterliegenden Mengen isomorph (vgl. Beispiele A.1.20 und A.1.21). Summen existieren auch in den Kategorien $\mathcal{G}\mathrm{rp}$ und $\mathcal{R}\mathrm{ing}$, aber können nicht so einfach beschrieben werden. Zum Beispiel ist die Summe von $\mathbb{Z}/2\mathbb{Z}$ und $\mathbb{Z}/3\mathbb{Z}$ in $\mathcal{G}\mathrm{rp}$ eine unendliche nicht-abelsche Gruppe.

A.2 Funktoren

Funktoren sind „Morphismen zwischen Kategorien“:

Definition A.2.1 (Funktor)**.** Seien $\mathcal{C}$ und $\mathcal{D}$ Kategorien. Ein *Funktor* $F\colon \mathcal{C} \to \mathcal{D}$ besteht aus

- einer Abbildung $F\colon \operatorname{Ob}\mathcal{C} \to \operatorname{Ob}\mathcal{D}$, und
- zu je zwei Objekten $X, Y \in \mathcal{C}$ einer Abbildung $F\colon \operatorname{Mor}_{\mathcal{C}}(X,Y) \to \operatorname{Mor}_{\mathcal{D}}(F(X),F(Y))$,

so dass die folgenden Bedingungen erfüllt sind:

- Für alle Morphismen $f\colon X \to Y$ und $g\colon Y \to Z$ in $\mathcal{C}$ gilt $F(g\circ f) = F(g)\circ F(f)$.
- Für alle Objekte $X \in \operatorname{Ob}\mathcal{C}$ gilt $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

Beispiel A.2.2 (Identitätsfunktor)**.** Jede Kategorie $\mathcal{C}$ besitzt einen *Identitätsfunktor*

$$\mathrm{id}_{\mathcal{C}}\colon \mathcal{C} \to \mathcal{C},$$

bestehend aus den Identitätsabbildungen auf $\operatorname{Ob}\mathcal{C}$ und auf $\operatorname{Mor}_{\mathcal{C}}(X,Y)$.

Beispiel A.2.3 (Vergissfunktoren)**.** Jede Gruppe $(G,\cdot)$ hat eine unterliegende Menge G, une jeder Gruppenhomomorphismus ist insbesondere eine Abbildung zwischen den unterliegenden Mengen. Damit gibt es einen Funktor

$$\begin{aligned} \mathcal{G}\mathrm{rp} &\to \mathcal{S}\mathrm{et}, \\ (G,\cdot) &\mapsto G, \\ f &\mapsto f, \end{aligned}$$

den man als *Vergissfunktor* von Gruppen nach Mengen bezeichnet, denn dieser Funktor „vergisst“ die Verknüpfung $\cdot$ der Gruppe $(G, \cdot)$. Auf ähnliche Weise gibt es Vergissfunktoren

$$\begin{aligned} \mathcal{M}\mathrm{od}_R &\to \mathcal{A}\mathrm{b}, & \mathcal{A}\mathrm{b} &\to \mathcal{G}\mathrm{rp}, \\ (M, +, \cdot) &\mapsto (M, +), & (A, +) &\mapsto (A, +) \\ f &\mapsto f, & f &\mapsto f, \end{aligned}$$

die die Skalarmultiplikation eines R-Moduls und die Kommutativität einer abelschen Gruppe vergessen. Die Kategorie $\mathcal{A}\mathrm{lg}_R$ von R-Algebren besitzt zwei Vergissfunktoren $\mathcal{A}\mathrm{lg}_R \to \mathcal{M}\mathrm{od}_R$ und $\mathcal{A}\mathrm{lg}_R \to \mathcal{R}\mathrm{ing}$, die die Ringmultiplikation und die Skalarmultiplikation vergessen.

Beispiel A.2.4 (additive und multiplikative Gruppe eines Ringes)**.** Jeder Ring hat eine zugehörige additive Gruppe sowie eine zugehörige multiplikative Gruppe. Es gibt zwei entsprechende Funktoren

$$\begin{aligned} \mathcal{R}\mathrm{ing} &\to \mathcal{G}\mathrm{rp}, & \mathcal{R}\mathrm{ing} &\to \mathcal{G}\mathrm{rp}, \\ (R, +, \cdot) &\mapsto (R, +), & (R, +, \cdot) &\mapsto (R^\times, \cdot) \\ f &\mapsto f, & f &\mapsto \text{Einschränkung von } f. \end{aligned}$$

Der erste ist der Vergissfunktor und kann auch als Funktor $\mathcal{R}\mathrm{ing} \to \mathcal{A}\mathrm{b}$ betrachtet werden. Zu dem zweiten Funktor siehe Bemerkung 8.1.23.

Beispiel A.2.5 (freie Moduln)**.** Sei R ein Ring. Die Konstruktion $I \mapsto R^{(I)}$ kann zu einem Funktor $\mathcal{S}\mathrm{et} \to \mathcal{M}\mathrm{od}_R$ befördert werden: Zu jeder Abbildung $f\colon I \to J$ gibt es genau eine R-lineare Abbildung

$$f_*\colon R^{(I)} \to R^{(J)}, \quad e_i \mapsto e_{f(i)},$$

nach der universellen Eigenschaft von Basen. Man kann dann leicht nachprüfen, dass $(\mathrm{id}_I)_* = \mathrm{id}_{R^{(I)}}$ und $(g \circ f)_* = g_* \circ f_*$.

Beispiel A.2.6 (Tensorprodukt)**.** Sei R ein kommutativer Ring. Das binäre Tensorprodukt von R-Moduln ist ein Funktor

$$\otimes_R\colon \mathcal{M}\mathrm{od}_R \times \mathcal{M}\mathrm{od}_R \to \mathcal{M}\mathrm{od}_R, \quad (M, N) \mapsto M \otimes_R N.$$

Hier wird das Produkt $\mathcal{C} \times \mathcal{D}$ von zwei Kategorien folgendermaße definiert:

- Es gilt $\mathrm{Ob}(\mathcal{C} \times \mathcal{D}) = \mathrm{Ob}\,\mathcal{C} \times \mathrm{Ob}\,\mathcal{D}$.
- Für Objekte $X, X' \in \mathcal{C}$ und $Y, Y' \in \mathcal{D}$ gilt

$$\mathrm{Mor}_{\mathcal{C}\times\mathcal{D}}((X, Y), (X', Y')) = \mathrm{Mor}_{\mathcal{C}}(X, X') \times \mathrm{Mor}_{\mathcal{D}}(Y, Y').$$

- Komposition von Morphismen und Identitätsmorphismen werden komponentenweise definiert.

Beispiel A.2.7 (Skalareinschränkung und Skalarerweiterung)**.** Sei $f\colon R \to S$ ein Ringhomomorphismus. Jeder S-Modul M hat dann auch eine R-Modulstruktur mit der Skalarmultiplikation $R \times M \to M$, $(r, x) \mapsto f(r) \cdot x$. Dies definiert einen Funktor $\mathcal{M}\mathrm{od}_S \to \mathcal{M}\mathrm{od}_R$, der als *Skalareinschränkung* entlang f bezeichnet wird. In der anderen Richtung definiert das Tensorprodukt einen Funktor

$$\mathcal{M}\mathrm{od}_R \to \mathcal{M}\mathrm{od}_S, \quad M \mapsto S \otimes_R M,$$

der als *Skalarerweiterung* entlang f bezeichnet wird.

Beispiel A.2.8 (Gruppenhomomorphismen und monotone Abbildungen als Funktoren)**.** Man kann die folgenden Aussagen leicht nachprüfen:

(i) Zu jedem Gruppenhomomorphismus $f\colon G \to H$ gibt es genau einen Funktor $\mathcal{B}f\colon \mathcal{B}G \to \mathcal{B}H$, deren Morphismenabbildung $\mathrm{Mor}_{\mathcal{B}G}(*,*) \to \mathrm{Mor}_{\mathcal{B}H}(*,*)$ gleich f ist.

(ii) Zu jeder monotonen Abbildung $f\colon X \to Y$ zwischen partiell geordneten Mengen gibt es genau einen Funktor $\mathcal{N}f\colon \mathcal{N}X \to \mathcal{N}Y$, deren Objektabbildung $\mathrm{Ob}\,\mathcal{N}X \to \mathrm{Ob}\,\mathcal{N}Y$ gleich f ist.

Proposition A.2.9 (Funktoren erhalten Isomorphie)**.** *Sei $F\colon \mathcal{C} \to \mathcal{D}$ ein Funktor. Ist $f\colon X \to Y$ ein Isomorphismus in $\mathcal{C}$, so ist $F(f)\colon F(X) \to F(Y)$ ein Isomorphismus in $\mathcal{D}$. Insbesondere: Aus $X \cong Y$ folgt $F(X) \cong F(Y)$.*

Beweis. Sei $g\colon Y \to X$ der Umkehrmorphismus zu f. Dann ist $F(g)$ ein Umkehrmorphismus zu $F(f)$, denn:

$$\begin{aligned} F(g) \circ F(f) &= F(g \circ f) = F(\mathrm{id}_X) = \mathrm{id}_{F(X)}, \\ F(f) \circ F(g) &= F(f \circ g) = F(\mathrm{id}_Y) = \mathrm{id}_{F(Y)}\,. \end{aligned}$$ □

Bemerkung A.2.10 (Erhaltung von Produkten bzw. Summen)**.** Nicht alle kategorische Konstruktionen werden automatisch von Funktoren erhalten. Sei zum Beispiel $F\colon \mathcal{C} \to \mathcal{D}$ ein Funktor und $(X_i)_{i\in I}$ eine Familie von Objekten von $\mathcal{C}$. Mit der universellen Eigenschaft des Produkts bzw. der Summe erhalten wir kanonische Morphismen

$$F\left(\prod_{i\in I} X_i\right) \to \prod_{i\in I} F(X_i) \quad \text{und} \quad \coprod_{i\in I} F(X_i) \to F\left(\coprod_{i\in I} X_i\right)$$

in $\mathcal{D}$, sofern beide Produkte und beide Summen existieren. Im Allgemeinen sind aber diese Morphismen keine Isomorphismen. Man sagt, dass ein Funktor F Produkte bzw. Summen erhält, wenn die obigen Morphismen Isomorphismen sind. Zum Beispiel:

(i) Die Vergissfunktoren $\mathcal{A}\mathrm{b} \to \mathcal{G}\mathrm{rp} \to \mathcal{S}\mathrm{et}$ erhalten Produkte aber nicht Summen.

(ii) Der Vergissfunktor $\mathcal{M}\mathrm{od}_R \to \mathcal{A}\mathrm{b}$ erhält beide Produkte und Summen.

(iii) Der „freier Modul“ Funktor $\mathcal{S}\mathrm{et} \to \mathcal{M}\mathrm{od}_R$, $I \mapsto R^{(I)}$, erhält Summen aber nicht Produkte.

Beispiel A.2.11 (Gruppenoperationen)**.** Sei G eine Gruppe und $\mathcal{C}$ eine Kategorie. Eine *Operation* der Gruppe G auf einem Objekt $X \in \mathrm{Ob}\,\mathcal{C}$ ist ein Funktor $F\colon \mathcal{B}G \to \mathcal{C}$ mit $F(*) = X$. Ein Objekt von $\mathcal{C}$ versehen mit einer Operation von G heißt auch *G-Objekt* von $\mathcal{C}$. Nach Proposition A.2.9 liefert der Funktor F eine Abbildung

$$G = \mathrm{Aut}_{\mathcal{B}G}(*) \to \mathrm{Aut}_{\mathcal{C}}(X),$$

und die Kompatibiltät mit Komposition impliziert, dass diese Abbildung ein Gruppenhomomorphismus ist. Umgekehrt bestimmt jeder Gruppenhomomorphimsus $G \to \mathrm{Aut}_{\mathcal{C}}(X)$ eine Operation von G auf X. Zum Beispiel:

(i) Eine Operation von G auf einer Menge X ist ein Gruppenhomorphismus $\rho\colon G \to S_X$, wobei $S_X = \mathrm{Aut}_{\mathcal{S}\mathrm{et}}(X)$ die symmetrische Gruppe von X ist. Die Abbildung ρ ist auch durch die Verknüpfung

$$G \times X \to X, \quad (g,x) \mapsto g \cdot x := \rho(g)(x)$$

bestimmt, für die gilt $e \cdot x = x$ und $g \cdot (h \cdot x) = (g \cdot h) \cdot x$. Umgekehrt bestimmt jede solche Verknüpfung einen Gruppenhomomorphismus von G nach S_X, und damit eine Operation von G auf X.

(ii) Eine Operation von G auf einem K-Vektorraum heißt eine *lineare Darstellung* von G über K. Zum Beispiel ist eine Operation von G auf K^n ein Gruppenhomorphismus $G \to \mathrm{Aut}_K(K^n) \cong \mathrm{GL}_n(K)$.

Definition A.2.12 (kontravarianter Funktor)**.** Seien $\mathcal{C}$ und $\mathcal{D}$ Kategorien. Ein Funktor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$ heißt *kontravarianter Funktor* von $\mathcal{C}$ nach $\mathcal{D}$. Zur Unterscheidung wird ein Funktor $\mathcal{C} \to \mathcal{D}$ auch als *kovarianter Funktor* von $\mathcal{C}$ nach $\mathcal{D}$ bezeichnet.

Ein kontravarianter Funktor F von $\mathcal{C}$ nach $\mathcal{D}$ dreht die Richtung der Pfeile um: Für einen Morphismus $f\colon X \to Y$ in $\mathcal{C}$ ist $F(f)$ ein Morphismus $F(Y) \to F(X)$ in $\mathcal{D}$.

Beispiel A.2.13 (Dualraum)**.** Sei K ein Körper. Die Konstruktion des Dualraums definiert einen kontravarianten Funktor von $\mathcal{V}\mathrm{ect}_K$ nach $\mathcal{V}\mathrm{ect}_K$, indem wir V auf V^* und eine lineare Abbildung $f\colon V \to W$ auf ihre duale Abbildung $f^*\colon W^* \to V^*$ abbilden. Dabei muss man nachprüfen:

- $\mathrm{id}_V^*\colon V^* \to V^*$ ist die Identität auf V^*: Es gilt nach Definition $\mathrm{id}_V^*(\alpha) = \alpha \circ \mathrm{id}_V = \alpha$.
- Für lineare Abbildungen $f\colon U \to V$ und $g\colon V \to W$ gilt $(g \circ f)^* = f^* \circ g^*$: Siehe Bemerkung 4.1.50

Beispiel A.2.14 (Potenzen von Moduln)**.** Sei R ein Ring. Die Konstruktion $I \mapsto R^I$ lässt sich zu einem *kontravarianten* Funktor von $\mathcal{S}\mathrm{et}$ nach $\mathcal{M}\mathrm{od}_R$ befördern: Zu jeder Abbildung $f\colon I \to J$ definiert man eine R-lineare Abbildung

$$f^*\colon R^J \to R^I, \quad (h\colon J \to R) \mapsto (h \circ f\colon I \to R).$$

Dieses Beispiel und Beispiel A.2.5 zeigen, dass die ähnlichen Konstruktionen $I \mapsto R^{(I)}$ und $I \mapsto R^I$ ein ganz gegensätzliches funktorielles Verhalten haben.

Beispiel A.2.15 (dargestellte Funktoren)**.** Sei $\mathcal{C}$ eine Kategorie und sei $X \in \mathrm{Ob}\,\mathcal{C}$. Der *von X dargestellte kovariante Funktor* $\mathrm{Mor}_{\mathcal{C}}(X, -)\colon \mathcal{C} \to \mathcal{S}\mathrm{et}$ ist wie folgt definiert:

- Ein Objekt $Y \in \mathrm{Ob}\,\mathcal{C}$ wird auf die Menge $\mathrm{Mor}_{\mathcal{C}}(X, Y)$ abgebildet.
- Ein Morphismus $f\colon Y \to Z$ in $\mathcal{C}$ wird auf den Morphismus

  $$\mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Mor}_{\mathcal{C}}(X, Z), \quad g \mapsto f \circ g,$$

  in $\mathcal{S}\mathrm{et}$ abgebildet.

Der *von X dargestellte kontravariante Funktor* $\mathrm{Mor}_{\mathcal{C}}(-, X)\colon \mathcal{C}^{\mathrm{op}} \to \mathcal{S}\mathrm{et}$ ist wie folgt definiert:

- Ein Objekt $Y \in \mathrm{Ob}\,\mathcal{C}$ wird auf die Menge $\mathrm{Mor}_{\mathcal{C}}(Y, X)$ abgebildet.
- Ein Morphismus $f\colon Y \to Z$ in $\mathcal{C}$ wird auf den Morphismus

  $$\mathrm{Mor}_{\mathcal{C}}(Z, X) \to \mathrm{Mor}_{\mathcal{C}}(Y, X), \quad g \mapsto g \circ f,$$

  in $\mathcal{S}\mathrm{et}$ abgebildet.

Definition A.2.16 (volltreuer Funktor)**.** Ein Funktor $F\colon \mathcal{C} \to \mathcal{D}$ heißt *volltreu*, wenn folgendes gilt: Für alle $X, Y \in \mathrm{Ob}\,\mathcal{C}$ ist die Abbildung

$$F\colon \mathrm{Mor}_{\mathcal{C}}(X, Y) \to \mathrm{Mor}_{\mathcal{D}}(FX, FY)$$

bijektiv.

Beispiel A.2.17. Der Vergissfunktor $\mathcal{A}\mathrm{b} \to \mathcal{G}\mathrm{rp}$ bzw. $\mathcal{C}\mathcal{R}\mathrm{ing} \to \mathcal{R}\mathrm{ing}$ ist volltreu, da Morphismen auf beiden Seiten Gruppenhomomorphismen bzw. Ringhomomorphismen sind. Der Vergissfunktor $\mathcal{G}\mathrm{rp} \to \mathcal{S}\mathrm{et}$ ist nicht volltreu: Nicht alle Abbildungen zwischen Gruppen sind Gruppenhomomorphismen.

Beispiel A.2.18 (volle Unterkategorie)**.** Sei $\mathcal{C}$ eine Kategorie und sei $D \subset \operatorname{Ob}\mathcal{C}$. Dann können wir eine Kategorie $\mathcal{D}$ wie folgt definieren:

- Es gilt $\operatorname{Ob}\mathcal{D} = D$.
- Für alle $X, Y \in D$ gilt $\operatorname{Mor}_{\mathcal{D}}(X,Y) = \operatorname{Mor}_{\mathcal{C}}(X,Y)$.
- Die Komposition und die Identitätsmorphismen in $\mathcal{D}$ sind die von $\mathcal{C}$.

Es gibt dann eine offenbare Inklusionsfunktor $\mathcal{D} \to \mathcal{C}$, der volltreu ist. Man nennt eine solche Kategorie $\mathcal{D}$ eine *volle Unterkategorie* von $\mathcal{C}$. Zum Beispiel: $\mathcal{A}\mathrm{b}$ ist eine volle Unterkategorie von $\mathcal{G}\mathrm{rp}$, und $\mathcal{C}\mathcal{R}\mathrm{ing}$ ist eine volle Unterkategorie von $\mathcal{R}\mathrm{ing}$.

Für volltreue Funktoren gilt die Umkehrung der Proposition A.2.9:

Proposition A.2.19 (volltreue Funktoren sind konservativ)**.** *Sei $F\colon \mathcal{C} \to \mathcal{D}$ ein volltreuer Funktor und $f\colon X \to Y$ ein Morphismus in $\mathcal{C}$. Ist $F(f)$ ein Isomorphismus, so ist f bereits ein Isomorphismus.*

Beweis. Sei $h\colon F(Y) \to F(X)$ ein Umkehrmorphismus zu $F(f)$. Da $F\colon \operatorname{Mor}_{\mathcal{C}}(Y,X) \to \operatorname{Mor}_{\mathcal{D}}(F(Y),F(X))$ surjektiv ist, gibt es einen Morphismus $g\colon Y \to X$ mit $F(g) = h$. Dann gilt $F(g \circ f) = h \circ F(f) = \mathrm{id}_{F(X)} = F(\mathrm{id}_X)$ und $F(f \circ g) = f(f) \circ h = \mathrm{id}_{F(Y)} = F(\mathrm{id}_Y)$. Da die Abbildungen $F\colon \operatorname{End}_{\mathcal{C}}(X) \to \operatorname{End}_{\mathcal{D}}(F(X))$ und $F\colon \operatorname{End}_{\mathcal{C}}(Y) \to \operatorname{End}_{\mathcal{D}}(F(Y))$ injektiv sind, gilt $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$, d.h., g ist ein Umkehrmorphismus zu f. □

Definition A.2.20 (wesentliches Bild, wesentlich surjektiv)**.** Sei $F\colon \mathcal{C} \to \mathcal{D}$ ein Funktor.

- Das *wesentliche Bild* von F ist die volle Unterkategorie von $\mathcal{D}$, deren Objekte diejenigen Objekte von $\mathcal{D}$ sind, die zu einem Objekt der Gestalt $F(X)$ mit $X \in \operatorname{Ob}\mathcal{C}$ isomorph sind.
- F heißt *wesentlich surjektiv*, wenn sein wesentliches Bild gleich $\mathcal{D}$ ist.

Beispiel A.2.21. Der Vergissfunktor $\mathcal{G}\mathrm{rp} \to \mathcal{S}\mathrm{et}$ ist nicht wesentlich surjektiv, da die leere Menge keine Gruppenstruktur besitzt. Sei aber $\mathcal{S}\mathrm{et}_{\neq\varnothing}$ die volle Unterkategorie von $\mathcal{S}\mathrm{et}$, deren Objekte die nicht-leeren Mengen sind. Dann kann man zeigen, dass der Vergissfunktor $\mathcal{G}\mathrm{rp} \to \mathcal{S}\mathrm{et}_{\neq\varnothing}$ wesentlich surjektiv ist.

A.3 Natürliche Transformationen

Natürliche Transformationen sind „Morphismen zwischen Funktoren“:

Definition A.3.1 (natürliche Transformation)**.** Seien $\mathcal{C}, \mathcal{D}$ Kategorien und seien $F, G\colon \mathcal{C} \to \mathcal{D}$ Funktoren. Eine *natürliche Transformation* $\alpha\colon F \to G$ von F nach G ist eine Familie $(\alpha_X\colon F(X) \to G(X))_{X \in \operatorname{Ob}\mathcal{C}}$ von Morphismen in $\mathcal{D}$ mit folgender Eigenschaft: Für jeden Morphismus $f\colon X \to Y$ in $\mathcal{C}$ ist das folgende Quadrat kommutativ:

$$\begin{array}{ccc} F(X) & \xrightarrow{\alpha_X} & G(X) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(Y) & \xrightarrow{\alpha_Y} & G(Y). \end{array}$$

Man bezeichnet mit $\operatorname{Nat}(F,G)$ die Menge (oder Klasse) aller natürlichen Transformationen von F nach G.

Definition A.3.2 (Kategorie von Funktoren)**.** Seien $\mathcal{C}$ und $\mathcal{D}$ Kategorien. Man definiert die *Kategorie von Funktoren* $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ von $\mathcal{C}$ nach $\mathcal{D}$ wie folgt:

- $\mathrm{Ob}\,\mathrm{Fun}(\mathcal{C},\mathcal{D})$ ist die Klasse aller Funktoren von $\mathcal{C}$ nach $\mathcal{D}$.
- Sind F und G zwei Funktoren, so setzt man
$$\mathrm{Mor}_{\mathrm{Fun}(\mathcal{C},\mathcal{D})}(F,G) = \mathrm{Nat}(F,G).$$
- Zwei natürliche Transformationen $\alpha\colon F \to G$ und $\beta\colon G \to H$ können punktweise komponiert werden:
$$(\beta\circ\alpha)_X = \beta_X \circ \alpha_X,$$
was eine natürliche Transformation $\beta\circ\alpha\colon F \to H$ definiert.
- Die Identitätstransformation eines Funktors F ist $\mathrm{id}_F = (\mathrm{id}_{F(X)})_{X\in\mathrm{Ob}\,\mathcal{C}}$.

Beispiel A.3.3 (Kategorie von G-Objekten)**.** Sei $\mathcal{C}$ eine Katgeorie und G eine Gruppe. Die *Kategorie von G-Objekten* in $\mathcal{C}$ ist die Kategorie $\mathrm{Fun}(\mathcal{B}G,\mathcal{C})$ (siehe Beispiel A.2.11).

Definition A.3.4 (natürlicher Isomorphismus)**.** Seien $F,G\colon \mathcal{C}\to\mathcal{D}$ Funktoren. Ein *natürlicher Isomorphismus* $\alpha\colon F \xrightarrow{\sim} G$ ist ein Isomorphismus in der Kategorie $\mathrm{Fun}(\mathcal{C},\mathcal{D})$.

Bemerkung A.3.5 (Kriterium für natürliche Isomorphismen)**.** Seien $F,G\colon \mathcal{C}\to\mathcal{D}$ Funktoren und sei $\alpha\colon F\to G$ eine natürliche Transformation. Dann ist α genau dann ein natürlicher Isomorphismus, wenn für alle $X\in\mathrm{Ob}\,\mathcal{C}$ der Morphismus $\alpha_X\colon F(X)\to G(X)$ in $\mathcal{D}$ ein Isomorphismus ist. Denn die Umkehrmorphismen $(\alpha_X)^{-1}\colon G(X)\to F(X)$ bilden automatisch eine natürliche Transformation von G nach F, die invers zu α ist.

Beispiel A.3.6 (Doppeldualraum)**.** Sei K ein Körper und sei $(-)^{**} = (-)^*\circ(-)^*\colon \mathcal{V}\mathrm{ect}_K \to \mathcal{V}\mathrm{ect}_K$ der Doppeldualraumfunktor. Die linearen Abbildungen $\mathrm{ev}\colon V\hookrightarrow V^{**}$ aus Proposition 4.1.59 bilden eine natürliche Transformation
$$\mathrm{ev}\colon \mathrm{id}_{\mathcal{V}\mathrm{ect}_K} \to (-)^{**}.$$
Wenn man beide Funktoren auf die Kategorie $\mathcal{V}\mathrm{ect}_K^{<\infty}$ von endlich-dimensionalen K-Vektorräumen einschränkt, wird ev zu einem natürlichen Isomorphismus zwischen $\mathrm{id}_{\mathcal{V}\mathrm{ect}_K^{<\infty}}$ und $(-)^{**}\colon \mathcal{V}\mathrm{ect}_K^{<\infty}\to\mathcal{V}\mathrm{ect}_K^{<\infty}$. Deswegen sagt man, dass ein endlich-dimensionaler K-Vektorraum V zu seinem Doppeldualraum V^{**} *natürlich isomorph* ist.

Definition A.3.7 (darstellbarer Funktor)**.** Sei $\mathcal{C}$ eine Kategorie.

- Ein Funktor $F\colon \mathcal{C}\to\mathcal{S}\mathrm{et}$ heißt *darstellbar*, wenn es ein Objekt $X\in\mathrm{Ob}\,\mathcal{C}$ und einen natürlichen Isomorphismus $\mathrm{Mor}_{\mathcal{C}}(X,-)\cong F$ gibt.
- Ein Funktor $F\colon \mathcal{C}^{\mathrm{op}}\to\mathcal{S}\mathrm{et}$ heißt *darstellbar*, wenn es ein Objekt $X\in\mathrm{Ob}\,\mathcal{C}$ und einen natürlichen Isomorphismus $\mathrm{Mor}_{\mathcal{C}}(-,X)\cong F$ gibt.

Bemerkung A.3.8. Ist $F\colon \mathcal{C}\to\mathcal{S}\mathrm{et}$ durch ein Objekt $X\in\mathrm{Ob}\,\mathcal{C}$ darstellbar, so ist X eindeutig bis auf Isomorphie durch F bestimmt. Denn seien $X,Y\in\mathrm{Ob}\,\mathcal{C}$ mit einem natürlichen Isomorphismus $\alpha\colon \mathrm{Mor}_{\mathcal{C}}(X,-)\xrightarrow{\sim}\mathrm{Mor}_{\mathcal{C}}(Y,-)$. Dann haben wir Morphismen $\alpha_X(\mathrm{id}_X)\in\mathrm{Mor}_{\mathcal{C}}(Y,X)$ und $\alpha_Y^{-1}(\mathrm{id}_Y)\in\mathrm{Mor}_{\mathcal{C}}(X,Y)$, und aus der Natürlichkeit von α folgt leicht, dass sie zueinander invers sind.

Beispiel A.3.9 (darstellbare Vergissfunktoren)**.**

(i) Sei R ein kommutativer Ring und sei $V\colon \mathcal{A}\mathrm{lg}_R\to\mathcal{S}\mathrm{et}$ der Vergissfunktor. Auswertung in T definiert eine natürliche Transformation
$$\alpha\colon \mathrm{Mor}_{\mathcal{A}\mathrm{lg}_R}(R[T],-)\to V,\quad \alpha_A(f\colon R[T]\to A) = f(T).$$
Die universelle Eigenschaft der R-Algebra $R[T]$ (Proposition 8.1.45) besagt, dass α ein natürlicher Isomorphismus ist. Insbesondere ist der Vergissfunktor $\mathcal{A}\mathrm{lg}_R\to\mathcal{S}\mathrm{et}$ durch die Polynomalgebra $R[T]$ darstellbar.

(ii) Auf ähnliche Weise ist der Vergissfunktor $\mathcal{R}\mathrm{ing} \to \mathcal{S}\mathrm{et}$ durch den Polynomring $\mathbb{Z}[T]$ darstellbar.

(iii) Der Vergissfuntor $\mathcal{M}\mathrm{od}_R \to \mathcal{S}\mathrm{et}$ ist durch den R-Modul R darstellbar. Dies folgt aus der universellen Eigenschaft der Standardbasis (1) von R: Für alle R-Moduln M ist die Auswertungsabbildung

$$\mathrm{Hom}_R(R, M) \to M, \quad f \mapsto f(1),$$

bijektiv.

(iv) Auf ähnliche Weise ist der Vergissfunktor $\mathcal{A}\mathrm{b} \to \mathcal{S}\mathrm{et}$ durch die abelsche Gruppe $\mathbb{Z}$ darstellbar. In der Tat ist sogar der Vergissfunktor $\mathcal{G}\mathrm{rp} \to \mathcal{S}\mathrm{et}$ durch die Gruppe $\mathbb{Z}$ darstellbar.

(v) Der Vergissfunktor $\mathcal{P}\mathrm{os} \to \mathcal{S}\mathrm{et}$ ist durch eine einelementige partiell geordnete Menge darstellbar, denn die Auswertungsabbilung

$$\mathrm{Mor}_{\mathcal{P}\mathrm{os}}((\{*\}, \leq), (X, \leq)) \to X, \quad f \mapsto f(*),$$

ist bijektiv.

Beispiel A.3.10 (Potenzmenge)**.** Die Konstruktion $X \mapsto \mathcal{P}(X)$ lässt sich zu einem kontravarianten Funktor $\mathcal{P}$ von $\mathcal{S}\mathrm{et}$ nach $\mathcal{S}\mathrm{et}$ befördern: Eine Abbildung $f\colon X \to Y$ wird auf die Urbildsabbildung $\mathcal{P}(Y) \to \mathcal{P}(X)$, $B \mapsto f^{-1}(B)$, abgebildet. Die bijektiven Abbildungen

$$\mathrm{Abb}(X, \{0,1\}) \to \mathcal{P}(X), \quad \chi \mapsto \chi^{-1}(\{1\}),$$

(siehe Beispiel 1.3.24) bilden einen natürlichen Isomorphismus $\mathrm{Mor}_{\mathcal{S}\mathrm{et}}(-, \{0,1\}) \cong \mathcal{P}$. Insbesondere ist der kontravariante Funktor $\mathcal{P}$ durch die zweielementige Menge $\{0,1\}$ darstellbar.

Beispiel A.3.11 (Tensorprodukt als darstellendes Objekt)**.** Sei R ein kommutativer Ring und seien M, N Moduln über R. Es gibt einen Funktor

$$\mathrm{Hom}^2_R(M, N, -)\colon \mathcal{M}\mathrm{od}_R \to \mathcal{S}\mathrm{et}, \quad P \mapsto \{R\text{-bilineare Abbildungen } M \times N \to P\}.$$

Nach Definition des Tensorprodukts gibt es zu jedem R-Modul P eine Bijektion

$$\alpha_P\colon \mathrm{Hom}_R(M \otimes_R N, P) \xrightarrow{\sim} \mathrm{Hom}^2_R(M, N, P), \quad \alpha_P(f) = f \circ \tau,$$

wobei $\tau\colon M \times N \to M \otimes_R N$ die universelle R-bilineare Abbildung ist. Man kann dann leicht nachprüfen, dass die Abbildungen α_P einen natürlichen Isomorphismus

$$\alpha\colon \mathrm{Hom}_R(M \otimes_R N, -) \xrightarrow{\sim} \mathrm{Hom}^2_R(M, N, -)$$

bilden. Damit ist der Funktor $\mathrm{Hom}^2_R(M, N, -)$ durch den R-Modul $M \otimes_R N$ darstellbar.

A.3.1 Äquivalenzen von Kategorien

Definition A.3.12 (Äquivalenz von Kategorien)**.** Ein Funktor $F\colon \mathcal{C} \to \mathcal{D}$ heißt *Äquivalenz von Kategorien*, wenn ein Funktor $G\colon \mathcal{D} \to \mathcal{C}$ mit natürlichen Isomorphismen $G \circ F \cong \mathrm{id}_{\mathcal{C}}$ und $F \circ G \cong \mathrm{id}_{\mathcal{D}}$ existiert.

Diese Definition ist ähnlich wie die von Isomorphismus (Definition A.1.7), aber es wird nicht erfordert, dass $G \circ F$ und $F \circ G$ *gleich* der Identität sind, sondern nur zu ihr *isomorph* sind. Trotzdem spielen Äquivalenzen von Kategorien die gleiche Rolle wie Isomorphismen innerhalb einer Kategorie: Zwei äquivalente Kategorien haben genau dieselben kategorischen Eigenschaften (z.B., die Existenz von Produkten oder Summen), und allgemeiner sind völlig durcheinander ersetzbar.

Satz A.3.13 (Charakterisierung von Äquivalenzen)**.** *Sei $F\colon \mathcal{C} \to \mathcal{D}$ ein Funktor. Die folgenden Aussagen sind äquivalent:*

(i) *F ist eine Äquivalenz von Kategorien.*

(ii) *F ist volltreu und wesentlich surjektiv.*

Beweis. Zu (i) $\Rightarrow$ (ii). Sei $G\colon \mathcal{D} \to \mathcal{C}$ ein Funktor mit natürliche Isomorphismen $\alpha\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} G \circ F$ und $\beta\colon F \circ G \xrightarrow{\sim} \mathrm{id}_{\mathcal{D}}$. Ist $Y \in \mathrm{Ob}\,\mathcal{D}$, so ist $\beta_Y\colon Y \xrightarrow{\sim} F(G(Y))$ ein Isomorphismus zwischen Y und einem Objekt der Gestalt $F(X)$. Damit ist F wesentlich surjektiv. Seien nun $X, Y \in \mathrm{Ob}\,\mathcal{C}$. Da α eine natürliche Transformation ist, kommutiert für jedes $f \in \mathrm{Mor}_{\mathcal{C}}(X, Y)$ das Quadrat

$$\begin{array}{ccc} X & \xrightarrow{\alpha_X} & GFX \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle G(F(f))} \\ Y & \xrightarrow{\alpha_Y} & GFY. \end{array}$$

Anders gesagt ist die Komposition

$$\mathrm{Mor}_{\mathcal{C}}(X, Y) \xrightarrow{F} \mathrm{Mor}_{\mathcal{D}}(FX, FY) \xrightarrow{G} \mathrm{Mor}_{\mathcal{C}}(GFX, GFY) \xrightarrow{\sim} \mathrm{Mor}_{\mathcal{C}}(X, Y)$$

gleich der Identität, wobei die dritte Abbildung die Bijektion $g \mapsto \alpha_X^{-1} \circ g \circ \alpha_Y$ ist. Daraus folgt, dass die erste Abbildung injektiv und die zweite surjektiv ist. Aber G ist auch eine Äquivalenz von Kategorien, und somit ist die zweite Abbildung injektiv nach demselben Argument. Also ist die zweite Abbildung bijektiv und damit ist auch die erste Abbildung bijektiv, d.h., F ist volltreu.

Zu (ii) $\Rightarrow$ (i). Sei $F\colon \mathcal{C} \to \mathcal{D}$ volltreu und wesentlich surjektiv. Wegen der wesentlichen Surjektivität kann man zu jedem Objekt $Z \in \mathrm{Ob}\,\mathcal{D}$ ein Objekt $G(Z) \in \mathrm{Ob}\,\mathcal{D}$ mit einem Isomorphismus $\beta_Z\colon F(G(Z)) \xrightarrow{\sim} Z$ auswählen. Ist $g\colon Z \to W$ ein Morphismus in $\mathcal{D}$, so gibt es genau einen Morphismus $G(g)\colon G(Z) \to G(W)$ mit

$$F(G(g)) = \beta_W^{-1} \circ g \circ \beta_Z,$$

denn F ist volltreu. Aus der Eindeutigkeit von $G(g)$ folgt leicht, dass für zwei Morphismen $g\colon Z \to W$ und $h\colon W \to V$ gilt $G(h \circ g) = G(h) \circ G(g)$. Es gilt auch $G(\mathrm{id}_Z) = \mathrm{id}_{G(Z)}$, denn $F(\mathrm{id}_{G(Z)}) = \mathrm{id}_{F(G(Z))} = \beta_Z^{-1} \circ \mathrm{id}_Z \circ \beta_Z$. Das heißt, wir haben einen Funktor $G\colon \mathcal{D} \to \mathcal{C}$ definiert. Außerdem ist nach Konstruktion die Familie $(\beta_Z)_{Z \in \mathrm{Ob}\,\mathcal{D}}$ ein natürlicher Isomorphismus $\beta\colon F \circ G \xrightarrow{\sim} \mathrm{id}_{\mathcal{D}}$.

Es bleibt zu zeigen, dass $G \circ F$ zu dem Identitätsfunktor isomorph ist. Sei $X \in \mathrm{Ob}\,\mathcal{C}$. Da F volltreu ist, gibt es genau einen Morphismus $\alpha_X\colon G(F(X)) \to X$, so dass $F(\alpha_X) = \beta_{F(X)}\colon F(G(F(X))) \to F(X)$. Da $\beta_{F(X)}$ ein Isomorphismus ist, ist auch α_X ein Isomorphismus nach Proposition A.2.19. Für jeden Morphismus $f\colon X \to Y$ in $\mathcal{C}$ kommutiert das Quadrat

$$\begin{array}{ccc} GFX & \xrightarrow{\alpha_X} & X \\ {\scriptstyle G(F(f))}\downarrow & & \downarrow{\scriptstyle f} \\ GFY & \xrightarrow{\alpha_Y} & Y, \end{array}$$

denn es kommutiert nach Anwendung des volltreuen Funktors F wegen der Natürlichkeit von β bzgl. $F(f)$. Damit ist die Familie $(\alpha_X)_{X \in \mathrm{Ob}\,\mathcal{C}}$ ein natürlicher Isomorphismus $\alpha\colon G \circ F \xrightarrow{\sim} \mathrm{id}_{\mathcal{C}}$, wie gewünscht. □

Bemerkung A.3.14. Nach Satz A.3.13 induziert jeder volltreue Funktor $F\colon \mathcal{C} \to \mathcal{D}$ eine Äquivalenz zwischen $\mathcal{C}$ und dem wesentlichen Bild von F.

Beispiel A.3.15 (kodiskrete Kategorien)**.** Für X eine beliebige Menge, sei $X^{\natural}$ die wie folgt definierte Kategorie:

- Es gilt $\operatorname{Ob} X^\natural = X$.
- Für alle $x, y \in X$ gilt $\operatorname{Mor}_{X^\natural}(x, y) = \{*\}$.
- Die Komposition und die Identitätsmorphismen sind eindeutig bestimmt.

In der Kategorie $X^\natural$ sind alle Objekte zueinander isomorph. Ist $f\colon X \to Y$ eine beliebige Abbildung zwischen nicht-leeren Mengen, so ist der induzierte Funktor $f^\natural\colon X^\natural \to Y^\natural$ volltreu und wesentlich surjektiv, und damit eine Äquivalenz von Kategorien. Dieses Beispiel zeigt, dass die Objektmengen von äquivalenten Kategorien ganz unterschiedlich sein können.

Beispiel A.3.16 (Matrizen und Vektorräume)**.** Sei K ein Körper und sei $\mathcal{V}\mathrm{ect}_K^{<\infty}$ die Kategorie von endlich erzeugten K-Vektorräumen. Man definiert eine Kategorie $\mathcal{M}\mathrm{at}_K$ wie folgt:

- Es gilt $\operatorname{Ob} \mathcal{M}\mathrm{at}_K = \mathbb{N}$.
- Für alle $n, m \in \mathbb{N}$ gilt $\operatorname{Mor}_{\mathcal{M}\mathrm{at}_K}(n, m) = M_{m\times n}(K)$.
- Komposition von Morphismen ist die Matrixmultiplikation, und der Identitätsmorphismus id_n ist die Einheitsmatrix I_n.

Es gibt dann einen Funktor

$$\begin{aligned} \mathcal{M}\mathrm{at}_K &\to \mathcal{V}\mathrm{ect}_K^{<\infty}, \\ n &\mapsto K^n, \\ A &\mapsto L_A. \end{aligned}$$

Nach Satz 4.2.29 ist dieser Funktor volltreu, und nach Korollar 3.3.23(ii) ist er wesentlich surjektiv. Damit ist er eine Äquivalenz von Kategorien.

Beispiel A.3.17 (Vergissäquivalenzen)**.**

(i) Der Vergissfunktor $\mathcal{M}\mathrm{od}_\mathbb{Z} \to \mathcal{A}\mathrm{b}$ ist nach Proposition 8.1.36 volltreu und wesentlich surjektiv. Damit ist er eine Äquivalenz von Kategorien. Auf ähnliche Weise ist der Vergissfunktor $\mathcal{A}\mathrm{lg}_\mathbb{Z} \to \mathcal{R}\mathrm{ing}$ eine Äquivalenz von Kategorien.

(ii) Sei R ein Ring. Die Kategorie $\mathcal{E}\mathrm{nd}(\mathcal{M}\mathrm{od}_R)$ von Endomorphismen von R-Moduln wurde im Beispiel A.1.14 definiert. Nach Proposition 8.1.37 ist der Funktor

$$\mathcal{M}\mathrm{od}_{R[T]} \to \mathcal{E}\mathrm{nd}(\mathcal{M}\mathrm{od}_R), \quad M \mapsto (M, x \mapsto T \cdot x),$$

volltreu und wesentlich surjektiv. Damit ist er eine Äquivalenz von Kategorien.

Beispiel A.3.18 (Algebren als Ringhomomorphismen)**.** Sei R ein kommutativer Ring. Zu jedem R-Algebra A gibt es einen Ringhomomorphismus $e_A\colon R \to A$, $r \mapsto r \cdot 1$, und für jeden R-Algebrenhomomorphismus $f\colon A \to B$ gilt offensichtlich $e_A \circ f = e_B$. Anders gesagt gibt es einen Funktor

$$\begin{aligned} \mathcal{A}\mathrm{lg}_R &\to \mathcal{R}\mathrm{ing}_{R/}, \\ A &\mapsto (A, e_A), \\ f &\mapsto f. \end{aligned}$$

Dieser Funktor ist volltreu, denn ein Ringhomomorphismus $f\colon A \to B$ zwischen R-Algebren ist genau dann ein R-Modulhomomorphismus, wenn es gilt $e_A \circ f = e_B$. Nach Proposition 8.1.50 besteht das wesentliche Bild dieses Funktors aus allen Ringhomomorphismen $e\colon R \to S$, so dass jedes Element von $e(R)$ zentral in S ist. Damit gibt es eine Äquivalenz zwischen $\mathcal{A}\mathrm{lg}_R$ und dieser vollen Unterkategorie von $\mathcal{R}\mathrm{ing}_{R/}$. Ist $\mathcal{CA}\mathrm{lg}_R$ die volle Unterkategorie von $\mathcal{A}\mathrm{lg}_R$ bestehend aus den kommutativen R-Algebren, so erhalten wir sogar eine Äquivalenz von Kategorien $\mathcal{CA}\mathrm{lg}_R \xrightarrow{\sim} \mathcal{CR}\mathrm{ing}_{R/}$.

Beispiel A.3.19 (Morphismen als Funktoren)**.** Sei $\mathcal{C}$ eine beliebige Kategorie und sei $[1]$ die Kategorie mit zwei Objekten 0 und 1 und einem einzigen Morphismus $0 \to 1$ (außer den Identitätsmorphismen id_0 und id_1). Jeder Funktor $F\colon [1] \to \mathcal{C}$ induziert insbesondere einen Morphismus $F(0) \to F(1)$ in $\mathcal{C}$, und jede natürliche Transformation zwischen solchen Funktoren induziert ein kommutatives Quadrat in $\mathcal{C}$. Dies definiert einen Funktor

$$\mathrm{Fun}([1], \mathcal{C}) \to \mathcal{M}\mathrm{or}(\mathcal{C}),$$

und man kann leicht nachprüfen, dass er eine Äquivalenz von Kategorien ist.

Beispiel A.3.20 (Monomorphismen und Unterobjekte)**.** Sei $\mathcal{C}$ eine Kategorie. Ein Monomorphismus $i\colon Y \to X$ in $\mathcal{C}$ heißt auch ein *Unterobjekt* von X. Unterobjekte von X bilden eine Kategorie $\mathcal{S}\mathrm{ub}(X)$, nämlich die volle Unterkategorie der Kommakategorie $\mathcal{C}_{/X}$ bestehend aus den Monomorphismen. Nach Definition von Monomorphismus gibt es in $\mathcal{S}\mathrm{ub}(X)$ höchstens einen Morphismus zwischen je zwei Objekten. Der kategorische Begriff des Unterobjekts stimmt nicht genau mit den gewöhnlichen Begriffen (wie Teilmenge, Untervektorraum, Unterring, usw.) überein, aber aus der Sichtweise der Kategorientheorie sind sie trotzdem äquivalent.

Sei zum Beispiel X eine Menge und sei $\mathcal{P}(X)$ die Potenzmenge von X, versehen mit der partiellen Ordnung $\subset$. Für eine Teilmenge $A \subset B$ bezeichnen wir mit $i_{A,B}\colon A \hookrightarrow B$ die Inklusionsabbildung. Es gibt dann einen Funktor

$$\begin{aligned} \mathcal{N}(\mathcal{P}(X)) &\to \mathcal{S}\mathrm{ub}(X), \\ A &\mapsto i_{A,X}, \\ (A \subset B) &\mapsto i_{A,B}. \end{aligned}$$

Dieser Funktor ist volltreu, da es in $\mathcal{S}\mathrm{ub}(X)$ genau dann einen Morphismus von $i_{A,X}$ nach $i_{B,X}$ gibt, wenn $A \subset B$. Er ist auch wesentlich surjektiv und damit eine Äquivalenz, denn in $\mathcal{S}\mathrm{ub}(X)$ ist jeder Monomorphismus $i\colon Y \hookrightarrow X$ zu der Inklusionsabbildung $i(Y) \hookrightarrow X$ isomorph.

Als weiteres Beispiel sei M ein Modul über einem Ring R. Dann gibt es ebenso eine Äquivalenz von Kategorien

$$\mathcal{N}(\{\text{Untermoduln von } M\}, \subset) \xrightarrow{\sim} \mathcal{S}\mathrm{ub}(M).$$

Beispiel A.3.21 (Epimorphismen und Quotientenobjekte)**.** Auf ähnliche Weise definiert man die Kategorie $\mathcal{Q}\mathrm{uot}(X)$ der *Quotientenobjekte* von $X \in \mathrm{Ob}\,\mathcal{C}$ als die volle Unterkategorie von $\mathcal{C}_{X/}$ bestehend aus den Epimorphismen. Für eine Menge X gibt es dann eine Äquivalenz von Kategorien

$$\mathcal{N}(\{\text{Äquivalenzrelationen auf } X\}, \subset) \xrightarrow{\sim} \mathcal{Q}\mathrm{uot}(X), \quad \sim \mapsto (X \twoheadrightarrow X/\!\sim).$$

Die Kategorie $\mathcal{M}\mathrm{od}_R$ ist insofern besonders, als es stets eine Äquivalenz zwischen Unterobjekten und Quotientenobjekten eines R-Moduls M gibt:

$$\begin{aligned} \mathcal{S}\mathrm{ub}(M) &\xrightarrow{\sim} \mathcal{Q}\mathrm{uot}(M), \\ (i\colon N \hookrightarrow M) &\mapsto \mathrm{coker}(i) = M/\operatorname{im} i, \\ \ker p &\mapsfrom (p\colon M \twoheadrightarrow P). \end{aligned}$$

Beispiel A.3.22 (metrische vs. topologische Räume)**.** Sei $\mathcal{C}$ die Kategorie von metrischen Räumen und *stetigen* Abbildungen, und sei $\mathcal{T}\mathrm{op}$ die Kategorie von topologischen Räumen und stetigen Abbildungen. Es gibt einen Funktor $\mathcal{C} \to \mathcal{T}\mathrm{op}$, der jeden metrischen Raum (X, d) auf den topologischen Raum $(X, \mathcal{T}_d)$ abbildet, wobei $\mathcal{T}_d$ die von d erzeugte Topologie ist. Da die Morphismen auf beiden Seiten stetige Abbildungen sind, ist dieser Funktor volltreu. Damit gibt es eine Äquivalenz zwischen $\mathcal{C}$ und der vollen Unterkategorie von $\mathcal{T}\mathrm{op}$ bestehend aus den metrisierbaren topologischen Räumen.

Index