Lineare Algebra — WS 2012/13

Sebastian Goette

Inhaltsverzeichnis

Einleitung 1

Kapitel 1. Zahlen 3
1.1. Mengen und Abbildungen 3
1.2. Natürliche Zahlen 9
1.3. Ganze und Rationale Zahlen 14
1.4. Etwas Euklidische Geometrie 20
1.5. Komplexe Zahlen und die Geometrie der Ebene 23
1.6. Geometrie des Raumes und Quaternionen 28

Kapitel 2. Vektorräume und Moduln 35
2.1. Gruppen, Ringe, Körper 35
2.2. Moduln und Vektorräume 45
2.3. Lineare Abbildungen 52
2.4. Unterräume und Quotienten 57
2.5. Matrizen 66

Kapitel 3. Vektorräume über Körpern und Schiefkörpern 81
3.1. Basen 81
3.2. Dimension und Rang 85
3.3. Lineare Gleichungssysteme 92

Kapitel 4. Determinanten 101
4.1. Volumina und Determinantenfunktionen 101
4.2. Die Determinante 106
4.3. Orientierung reeller Vektorräume 118

Kapitel 5. Eigenwerte 121
5.1. Eigenvektoren 121
5.2. Polynome 126
5.3. Das Charakteristische Polynom und das Minimalpolynom 136

Kapitel 6. Endomorphismen und Normalformen 147
6.1. Euklidische Ringe und Hauptidealringe 147
6.2. Die Smith-Normalform und invariante Faktoren 151
6.3. Primfaktorzerlegung in Hauptidealringen 160
6.4. Der chinesische Restsatz und der Elementarteilersatz 170
6.5. Allgemeine und Jordan-Normalform 177

Kapitel 7. Vektorräume mit Skalarprodukt 189

7.1. Skalarprodukte 189
7.2. Skalarprodukte als Matrizen 197
7.3. Dualräume und adjungierte Abbildungen 205
7.4. Normale Endomorphismen 214
7.5. Affine Räume 224
7.6. Bilinearformen und quadratische Funktionen 231

Notation 241

Einleitung

Die Lineare Algebra ist die Lehre von Vektorräumen und linearen Abbildungen. Was das ist und warum man sich das anschauen sollte, wird im Laufe der Vorlesung hoffentlich klarer. Jedenfalls werden Ihnen in vielen weiterführenden Vorlesungen immer wieder Vektorräume und lineare Abbildungen begegnen, so dass es sicher sinnvoll ist, sie bereits am Anfang des Studiums kennenzulernen.

Wir beginnen im ersten Kapitel mit einer allgemeinen Einführung, bei der wir Grundlagen und erste Beispiele kennenlernen. Dazu wiederholen wir die Zahlbereiche $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$, die Sie aus der Schule kennen. Dann führen wir die komplexen Zahlen $\mathbb{C}$ und die Quaternionen $\mathbb{H}$ ein. Als ersten Vorgeschmack auf den Inhalt der Vorlesung beschäftigen wir uns mit der Euklidischen Geometrie der Ebene $\mathbb{R}^2$ und des Raumes $\mathbb{R}^3$ und stellen eine Beziehung zu $\mathbb{C}$ beziehungsweise $\mathbb{H}$ her.

Im zweiten Kapitel führen wir systematisch die Grundbegriffe ein. An die Stelle konkreter Zahlbereiche treten Ringe (wie $\mathbb{Z}$), Körper (wie $\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$) und Schiefkörper (wie $\mathbb{H}$). Die Ebene $\mathbb{R}^2$ und der Raum $\mathbb{R}^3$ sind die einfachsten Beispiele von Vektorräumen. Abbildungen, die mit der Vektorraum-Struktur verträglich sind, heißen linear. Wir werden allgemeiner mit dem Begriff eines Moduls über einem Ring beginnen, und wir werden viele der folgenden Überlegungen für bestimmte Klassen von Moduln durchführen. Beispielsweise lernen wir, wie man Elemente in freien Moduln durch Koordinaten und lineare Abbildungen zwischen solchen Moduln durch Matrizen beschreibt. Wir werden aber immer nur dann allgemeinere Objekte als Vektorräume über Körpern betrachten, wenn das ohne zusätzlichen technischen Aufwand möglich ist.

Im dritten Kapitel konzentrieren wir uns auf Vektorräume über Körpern und Schiefkörpern. Wir zeigen, dass jeder Vektorraum eine Basis besitzt, und dass die Dimension eine Invariante des Vektorraums ist, die ihn bis auf Isomorphie bestimmt. Außerdem betrachten wir die Struktur einer allgemeinen linearen Abbildung und lernen ein universelles Verfahren zum Lösen linearer Gleichungssysteme.

Im vierten Kapitel beschäftigen wir uns mit Endomorphismen freier Moduln über kommutativen Ringen und lernen die Determinante als wichtige Invariante kennen. Anschließend betrachten wir Eigenwerte und das charakteristische Polynom, und lernen erste Strukturaussagen über lineare Abbildungen von einem festen Vektorraum in sich selbst kennen.

KAPITEL 1

Zahlen

In diesem ersten Kapitel legen wir dazu die Grundlagen. Zuerst führen wir Sprechweisen für Mengen, Abbildungen und natürliche Zahlen ein. Danach konstruieren wir ganze und rationale Zahlen, wohingegen wir die reellen Zahlen als gegeben annehmen werden — ihre Konstruktion fällt in den Bereich der Analysis. Aus den reellen Zahlen konstruieren wir die komplexen Zahlen und die Quaternionen. Zum einen sind beides wichtige Beispiele für Körper beziehungsweise Schiefkörper. Auf der anderen Seite besteht ein enger Zusammenhang zur Euklidischen Geometrie in den Dimensionen 2 und 3, und euklidische Geometrie ist sicher einer der wichtigsten Vorläufer für den Vektorraum-Kalkül, um den es in dieser Vorlesung schwerpunktmäßig gehen wird.

1.1. Mengen und Abbildungen

Wenn man möchte, kann man fast die gesamte Mathematik auf das Studium von Mengen und ihren Elementen zurückführen. Das ist aber leider recht mühsam, und man muss sehr sorgfältig sein, um nicht in Widersprüche zu geraten. Wenn Sie wissen möchten, wie das geht, sollten Sie später im Verlauf Ihres Studiums eine Vorlesung über Mengenlehre besuchen. Wir wollen die Mengenlehre als eine Sprache benutzen, in der man sehr elegant über mathematische Sachverhalte sprechen kann. Dazu lernen wir jetzt die ersten Vokabeln und grammatikalischen Regeln.

Georg Cantor hat den Mengenbegriff als erster eingeführt.

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

1.1. Beispiel. Zahlen sind Objekte unserer Anschauung, also ist $\{1, 2, 3\}$ eine Menge. Die Menge $\mathbb{N} = \{0, 1, 2, \dots\}$ der natürlichen Zahlen lernen wir im Abschnitt 1.2 kennen.

Die „Objekte" in einer Menge heißen *Elemente*. Wenn ein Objekt a in einer Menge M enthalten ist, schreiben wir

$$a \in M,$$

ansonsten $a \notin M$.

1.2. Definition. Zwei Mengen heißen gleich, wenn sie die gleichen Elemente enthalten.

1.3. Bemerkung. Wenn man Mengen als Aufzählung $M = \{a_1, \dots, a_n\}$ angibt, kann es passieren, dass $a_i = a_j$ für zwei Indizes i und j. Trotzdem ist a_i dadurch nicht „zweimal“ in M enthalten. Also zum Beispiel

$$\{1, 1, 2\} = \{2, 1\} = \{1, 2\},$$

denn alle drei Mengen enthalten die gleichen Elemente, nämlich 1 und 2. Aber natürlich gilt

$$\{1, 2\} \neq \{1, 2, 3\}.$$

1.4. Beispiel. Besonders wichtig ist die *leere Menge*, die gar kein Element enthält. Wir schreiben

$$\emptyset = \{\,\}.$$

Inzwischen sind auch Mengen „Objekte unseres Denkens oder unserer Anschauung“ geworden. Also kann man auch Mengen betrachten, deren Elemente selbst wieder Mengen sind. In der Tat kann man ausgehend von der leeren Menge bereits sehr viele andere Mengen konstruieren, etwa

$$\emptyset = \{\,\}, \qquad \{\emptyset\}, \qquad \{\{\emptyset\}, \emptyset\} \qquad \text{usw.}\dots,$$

genug, um alle Objekte dieser Vorlesung zu beschreiben.

Wir stoßen jetzt auf das erste Problem mit Cantors Mengenbegriff.

1.5. Satz (Russellsche Antinomie)**.** *Es gibt keine Menge M, deren Elemente genau diejenigen Mengen sind, die sich nicht selbst enthalten.*

Wir formulieren die Russellsche Antinomie hier wie selbstverständlich als einen *Satz*, also als eine bewiesene mathematische Aussage. Zu ihrer Zeit war die Russellsche Antinomie ein Widerspruch im mathematischen Denkgebäude — so etwas darf es nicht geben, denn aus einem Widerspruch lässt sich alles folgern, man könnte als Mathematiker nicht mehr zwischen „richtig“ und „falsch“ unterscheiden, und dadurch würde Mathematik als Ganzes bedeutungslos. Man hat einige Zeit gebraucht, um eine handhabbare Version der Mengenlehre zu formulieren, in der aus dem fatalen Widerspruch ein harmloser Satz wird.

Beweis. Würde es eine solche Menge M geben, dann müsste entweder $M \in M$ oder $M \notin M$ gelten. Aber nach Definition von M gilt $M \in M$ genau dann, wenn $M \notin M$, und das ist ein Widerspruch. Also gibt es keine Menge M. $\square$

1.6. Bemerkung. Wir haben gerade unseren ersten *indirekten Beweis* kennengelernt. Bei einem indirekten Beweis nimmt man an, dass die Aussage, die man beweisen möchte, falsch ist, und leitet daraus einen Widerspruch her. Manchmal ist das die einfachste Weise, einen Satz zu beweisen. Der Nachteil ist aber, dass man — wie im obigen Beweis — nicht auf Anhieb versteht, warum der Satz gilt. Wenn möglich, wollen wir daher indirekte Beweise vermeiden.

Zurück zu Cantors Mengenbegriff und zur Russellschen Antinomie. Wir sehen, dass nicht jede „Zusammenfassung von Objekten unseres Denkens und unserer Anschauung“ eine Menge sein kann. Wir werden daher die Existenz einiger nützlicher Mengen annehmen, und wir werden einige Konstruktionen angeben, die neue Mengen aus alten erzeugen. Die gesamte Mathematik basiert auf der Annahme, dass man das ohne Widersprüche machen kann — aber aus prinzipiellen Gründen lässt sich die Widerspruchsfreiheit der Axiome der Mengenlehre nicht beweisen.

1.7. Definition. Seien M und N Mengen, dann heißt M eine *Teilmenge* von N, wenn alle Elemente a von M auch in N enthalten sind. Dafür schreiben wir

$$M \subset N .$$

1.8. Bemerkung. (1) Die leere Menge ist Teilmenge jeder Menge M.
(2) Es gilt $\{x\} \subset M$ genau dann, wenn $x \in M$.
(3) Es gilt immer $M \subset M$.
(4) Wenn $M \subset N$ und $M \neq N$ gilt, heißt M auch *echte Teilmenge* von N.

In den meisten Mathebüchern wird das Symbol „$\subset$“ so verwendet wie hier. Es gibt zwar eine internationale Norm, nach der nur echte Teilmengen mit „$\subset$“ bezeichnet werden sollen, aber in der Mathematik benötigt man das Symbol für beliebige Teilmengen weitaus häufiger, und schreibt daher „$\subset$“. Für echte Teilmengen verwenden wir das Symbol „$\subsetneqq$“. Falls Sie ein Mathebuch zur Hand nehmen, in dem das Symbol „$\subset$“ vorkommt, sollten Sie zur Sicherheit trotzdem herausfinden, ob der Autor damit beliebige oder nur echte Teilmengen bezeichnet. Genauso vorsichtig sollten Sie eigentlich mit allen Definitionen und Bezeichnungen verfahren.

Kommen wir jetzt zur Konstruktion neuer Mengen aus alten.

1.9. Definition. Seien M und N Mengen.

(1) Der *Durchschnitt* $M \cap N$ enthält genau die Elemente, die sowohl in M als auch in N enthalten sind.
(2) Die *Vereinigung* $M \cup N$ enthält genau die Elemente, die in M oder in N enthalten sind.
(3) Wenn $M \cap N = \emptyset$ gilt, heissen M und N *disjunkt*, und $M \cup N$ ist eine *disjunkte Vereinigung*. Um zu zeigen, dass eine Vereinigung disjunkt ist, schreiben wir $M \dot\cup N$.
(4) Die (*Mengen-*) *Differenz* $N \setminus M$ enthält genau die Elemente, die in N, aber nicht in M enthalten sind. Ist M Teilmenge von N, so nennt man $N \setminus M$ auch das *Komplement* von M in N.
(5) Das *kartesische Produkt* $M \times N$ besteht aus allen Paaren (x, y) von Elementen $x \in M$ und $y \in N$. Für das Produkt einer Menge M mit sich selbst schreibt man auch $M^2 = M \times M$.

Insbesondere sind $M \cap N$, $M \cup N$, $N \setminus M$ und $M \times N$ auch wieder Mengen.Für den Anfang reichen uns diese Konstruktionen. Später werden wir Vereinigungen und Durchschnitte beliebig vieler Mengen benötigen.

1.10. Bemerkung. Die Notation (x, y) bezeichnet ein (geordnetes) *Paar*, allgemeiner bezeichnet $(x_1, \dots, x_n)$ ein *n-Tupel*. Hierbei kommt es auf die Reihenfolge der Einträge (nicht „Elemente“!) an, und ein und derselbe Eintrag kann mehrfach auftreten. Zum Beispiel:

$$(1,1) \in \{1,2\} \times \{1,2,3\}$$

und

$$(1,2) \neq (2,1) \neq (2,1,1).$$

1.11. Definition. Die Menge aller Teilmengen von M heißt *Potenzmenge* $\mathcal{P}(M)$. Auch die Potenzmenge einer Menge ist wieder eine Menge.

1.12. Beispiel. Sei $M = \{1,2\}$, dann gilt

$$\mathcal{P}(M) = \big\{\emptyset, \{1\}, \{2\}, \{1,2\}\big\}.$$

Es sei M eine Menge. Man betrachtet oft die Teilmenge aller Elemente z von M, die eine bestimmte Eigenschaft E haben, und schreibt dafür

$$\{z \in M \mid z \text{ hat die Eigenschaft } E\}.$$

Wenn E eine mathematisch wohldefinierte Eigenschaft ist, dann erhalten wir wieder eine Menge.

1.13. Folgerung (aus der Russellschen Antinomie 1.5)**.** *Die Gesamtheit aller Mengen ist keine Menge.*

Beweis. Wäre die Gesamtheit aller Mengen selbst eine Menge N, dann wäre auch

$$M = \{X \in N \mid X \notin X\}$$

wieder eine Menge, was nach Satz 1.5 aber nicht sein kann. □

1.14. Definition. Es seien M und N Mengen. Eine *Abbildung* $F \colon M \to N$ (lies „F von M nach N“) ordnet jedem Element $x \in M$ ein Element $F(x) \in N$ zu.

Formal fassen wir F als Tripel $(M, N, \Gamma(F))$ auf. Hierbei ist $\Gamma(F) \subset M \times N$ der *Graph* von F. Wir fordern, dass zu jedem $x \in M$ genau ein $y \in N$ mit $(x,y) \in \Gamma(F)$ existiert, und setzen $F(x) = y$.

1.15. Definition. Es sei $F \colon M \to N$ eine Abbildung. Dann heißt M der *Definitionsbereich* von M und N der *Wertebereich*. Die Menge aller Abbildungen von M nach N wird mit $\operatorname{Abb}(M,N)$ bezeichnet .

Zwei Abbildungen sind gleich, wenn sie den gleichen Definitions- und den gleichen Wertebereich haben, und jedem Element des Definitionsbereichs jeweils dasselbe Element des Bildbereichs zuordnen.

1.16. Definition. Es sei $F \colon M \to N$ eine Abbildung. Dann heißt die Teilmenge

$$\operatorname{im} F = \{\, y \in N \mid \text{ Es gibt } x \in M \text{ mit } F(x) = y \,\} = \{\, F(x) \mid x \in M \,\}$$

das *Bild* von F.

Sei $V \subset N$ eine Teilmenge, dann heißt

$$F^{-1}(V) = \{x \in M \mid F(x) \in V\}$$

das *Urbild* von V unter F.

Für das Urbild der einelementigen Menge $\{y\}$ schreibt man manchmal kurz $F^{-1}(y)$ statt $F^{-1}(\{y\})$. Da das zu Missverständnissen führen kann, bleiben wir erst einmal bei $F^{-1}(\{y\})$.

1.17. Definition. Eine Abbildung $F\colon M \to N$ heißt

(1) *injektiv*, wenn für alle $x_1, x_2 \in M$ aus $F(x_1) = F(x_2)$ schon $x_1 = x_2$ folgt,
(2) *surjektiv*, wenn für alle $y \in N$ ein $x \in M$ existiert mit $F(x) = y$, und
(3) *bijektiv*, wenn sie injektiv und surjektiv ist.

1.18. Beispiel. (1) Für alle Mengen M ist die Abbildung $\mathrm{id}_M\colon M \to M$ mit $\mathrm{id}_M(x) = x$ definiert. Sie heißt die *Identität* und ist stets bijektiv.
(2) Die Abbildung $F\colon \mathbb{R} \to \mathbb{R}$ mit $F(x) = x^2$ ist weder injektiv noch surjektiv, denn

$$F(-2) = F(2) = 4 \qquad \text{und} \qquad -1 \notin \operatorname{im}(F).$$

(3) Die Abbildung $F\colon \mathbb{N} \to \mathbb{N}$ mit $F(x) = x^2$ ist injektiv. Die Abbildung $G\colon \mathbb{N} \to \{\, x^2 \mid x \in \mathbb{N} \,\}$ mit $G(x) = x^2$ ist bijektiv. Diese Abbildungen sind verschieden, da sie andere Wertebereiche haben.

Trotzdem werden wir später manchmal beide Abbildungen mit dem gleichen Symbol bezeichnen.

1.19. Definition. Seien L, M, N Mengen und $F\colon M \to N$, $G\colon L \to M$ Abbildungen. Die *Verkettung* $F \circ G\colon L \to N$ (lies „F nach G“) ist die durch

$$(F \circ G)(x) = F(G(x))$$

definierte Abbildung.

1.20. Bemerkung. Die Buchstaben in „$F \circ G$“ scheinen „falsch herum“ zu stehen, denn die Abbildungen verlaufen von links nach rechts geschrieben so:

$$\begin{array}{ccccc} L & \xrightarrow{G} & M & \xrightarrow{F} & N \\ x & \longmapsto & G(x) & \longrightarrow & F(G(x)) \,. \end{array}$$

Aber in „$(F \circ G)(x) = F(G(x))$“ stimmt die Reihenfolge wieder. Beispielsweise seien $F, G\colon \mathbb{R} \to \mathbb{R}$ definiert durch

$$F(x) = x^2 \qquad \text{und} \qquad G(x) = x + 1\,,$$

dann ist

$$(F \circ G)(x) = (x+1)^2 \quad \text{und} \quad (G \circ F) = x^2 + 1\,.$$

Insbesondere gilt $G \circ F \neq F \circ G$.

1.21. Bemerkung. Sei $F\colon M \to N$ eine Abbildung, und sei $U \subset M$ eine Teilmenge. Die Abbildung $G\colon U \to M$ mit $G(x) = x$ für alle $x \in U$ heißt *Inklusion*. Sie ist stets injektiv. Die Verkettung

$$F|_U = F \circ G\colon U \to N$$

(lies „F eingeschränkt auf U“) heißt *Einschränkung* von F auf U.

1.22. Satz. *Seien L, M, N Mengen und $F, F' \colon M \to N$, $G, G' \colon L \to M$ Abbildungen. Dann gilt*

(1) *Sind F, G injektiv, so ist auch $F \circ G$ injektiv.*
(2) *Sind F, G surjektiv, so ist auch $F \circ G$ surjektiv.*
(3) *Sind F, G bijektiv, so ist auch $F \circ G$ bijektiv.*
(4) *Ist $F \circ G$ injektiv, so auch G.*
(5) *Ist $F \circ G$ surjektiv, so auch F.*
(6) *Ist F injektiv, so folgt aus $F \circ G = F \circ G'$ bereits $G = G'$.*
(7) *Ist G surjektiv, so folgt aus $F \circ G = F' \circ G$ bereits $F = F'$.*

Hierbei bezeichnen F' und G' beliebige Abbildungen und nicht die „Ableitungen“ von F und G.

Beweis. Zu (1) seien $x, y \in L$. Aus $(F \circ G)(x) = (F \circ G)(y)$ folgt $F(G(x)) = F(G(y))$, also $G(x) = G(y)$ wegen Injektivität von F, also $x = y$ wegen Injektivität von G, also ist $F \circ G$ ebenfalls injektiv. Der Beweis von (2) verläuft ähnlich wie (1), und (3) folgt sofort aus (1) und (2).

Die Punkte (4), (5) sind Übungsaufgaben zur Vorlesung “Analysis I” und werden hier daher nicht bewiesen.

Aussage (6) folgt ähnlich wie (7). Zu (7) sei $y \in M$. Wegen Surjektivität von G existiert $x \in L$ mit $G(x) = y$. Aus $F \circ G = F' \circ G$ folgt

$$F(y) = (F \circ G)(x) = (F' \circ G)(x) = F'(y).$$

Da das für alle $y \in M$ gilt, folgt $F = F'$. □

1.23. Satz. *Sei $F \colon M \to N$ bijektiv. Dann existiert genau eine Abbildung $G \colon N \to M$ mit $G \circ F = \mathrm{id}_M$ und $F \circ G = \mathrm{id}_N$.*

1.24. Definition. Die Abbildung G aus Satz 1.23 heißt die *Umkehrabbildung* von F.

Die Umkehrabbildung von F wird manchmal mit F^{-1} bezeichnet. Auch das kann zu Missverständnissen führen, so dass wir auf diese Bezeichnung verzichten wollen.

Beweis von Satz 1.23. Wir müssen zeigen, dass G *existiert*, und dass G *eindeutig* ist.

Zur Eindeutigkeit nehmen wir an, dass $G \colon N \to M$ eine Umkehrfunktion ist. Dann sei $y \in N$ beliebig, und sei $x \in M$ das eindeutige Element mit $F(x) = y$. Aus $G \circ F = \mathrm{id}_M$ folgt

$$G(y) = G(F(x)) = x .$$

Wenn eine Umkehrfunktion existiert, sind ihre Werte durch diese Gleichung eindeutig bestimmt. Also ist die Umkehrfunktion eindeutig.

Zur Existenz sei $\Gamma(F)$ der Graph von F. Gemäß der obigen Überlegung betrachten wir

$$X = \left\{ (y, x) \in N \times M \;\middle|\; (x, y) \in \Gamma(F) \right\} ,$$

das ist eine Menge, da M, N und $\Gamma(F)$ auch Mengen sind. Zu jedem $y \in N$ existiert genau ein $x \in M$ mit $F(x) = y$, also mit $(x, y) \in \Gamma(F)$, also auch mit $(y, x) \in X$. Also ist X der Graph einer Funktion $G\colon N \to M$.

Für alle $x \in M$ ist $(F(x), x) \in X = \Gamma(G)$, also $G(F(x)) = x$, und somit $G \circ F = \mathrm{id}_M$. Umgekehrt sei $y \in N$, und sei $x \in M$ das eindeutige Element mit $F(x) = y$, also $G(y) = x$ und $F(G(y)) = F(x) = y$. Somit gilt auch $F \circ G = \mathrm{id}_N$. Also existiert eine Umkehrfunktion, nämlich G. □

1.25. Definition. Zwei Mengen M und N heissen *gleichmächtig*, wenn es eine bijektive Abbildung $F\colon M \to N$ gibt.

1.26. Bemerkung. Gleichmächtige Mengen haben „gleich viele" Elemente. Für alle Mengen L, M und N gilt (Übung):

(1) M ist gleichmächtig zu M;
(2) N ist genau dann gleichmächtig zu M, wenn M zu N gleichmächtig ist;
(3) sind L zu M und M zu N gleichmächtig, so ist auch L zu N gleichmächtig.

Das heißt, Gleichmächtigkeit verhält sich wie eine Äquivalenzrelation. Allerdings sollte eine Relation immer auf einer Menge definiert sein, und die Menge aller Mengen gibt es nach Folgerung 1.13 nicht.

1.27. Beispiel. (1) Die Mengen $M = \{1, 2, 3\}$ und $N = \{4, 7, 15\}$ sind gleichmächtig. Definiere z.B. $F\colon M \to N$ durch

$$F(1) = 7, \quad F(2) = 4, \quad F(3) = 15.$$

(2) Sei $M = \{n^2 \mid n \in \mathbb{N}\} \subset \mathbb{N}$ die Menge der Quadratzahlen. Da $F\colon \mathbb{N} \to M$ mit $F(n) = n^2$ bijektiv ist, sind M und $\mathbb{N}$ gleichmächtig, obwohl M eine echte Teilmenge von $\mathbb{N}$ ist.

1.2. Natürliche Zahlen

Die natürlichen Zahlen sind uns bereits seit unserer Kindheit vertraut — wir benutzen sie zum Zählen. Für den Fall, dass es nichts zu zählen gibt, haben wir die Zahl 0. Es ist erstaunlich, dass die Zahl 0 selbst erst spät als eigenständige Zahl eingeführt wurde. Wenn wir schon ein Stück weit gezählt haben, etwa bis zu einer Zahl n, und weiterzählen wollen, brauchen wir die nächste Zahl. Wir nennen Sie den Nachfolger von n und schreiben $n + 1$. Schließlich wollen wir, dass die natürlichen Zahlen eine Menge bilden, die sonst keine weiteren Objekte enthält.

1.28. Annahme (Peano-Axiome)**.** *Wir nehmen an, dass es eine Menge* $\mathbb{N}$ *mit einem ausgezeichneten Element* $0 \in \mathbb{N}$ *und einer Abbildung* $+1\colon \mathbb{N} \to \mathbb{N}$ *gibt, die die folgenden* Peano-Axiome *erfüllt:*

(1) *Die Nachfolger-Abbildung ist bijektiv als Abbildung* $+1\colon \mathbb{N} \to \mathbb{N} \setminus \{0\}$.

(2) *Sei $M \subset N$ eine Teilmenge mit $0 \in M$, so dass für alle $m \in M$ auch $m+1 \in M$ gilt, dann ist bereits $M = N$.*

Axiom (1) besagt, dass jede Zahl genau einen Nachfolger hat, und jede Zahl außer 0 selbst Nachfolger genau einer anderen Zahl ist. Axiom (2) besagt, dass die Menge N die "kleinste" Menge ist, die (1) erfüllt. Trotzdem bestimmen die Peano-Axiome die natürlichen Zahlen nicht eindeutig — warum das so ist, lernen Sie aber erst in einer Vorlesung über Logik. Wir wollen immerhin annehmen, dass $\mathbb{N}$ nur die Zahlen 0, 1, 2, … enthält, aber keine weiteren Elemente. Übrigens gibt es Autoren, für die $0 \notin \mathbb{N}$. Zur Sicherheit können Sie beide Versionen mit $\mathbb{N}_0$ und $\mathbb{N}_>$ bezeichnen.

1.29. Bemerkung. Wir können natürliche Zahlen als Mengen $\underline{0}$, $\underline{1}$, $\underline{2}$, … konstruieren. Dazu setzen wir $\underline{0} = \emptyset$ und konstruieren Nachfolger als

$$\underline{n} + 1 = \underline{n+1} = \{\underline{0}, \dots, \underline{n}\} = \underline{n} \cup \{\underline{n}\}.$$

Diese Definition ist *rekursiv*, das heißt, man muss alle Zahlen bis $\underline{n}$ kennen, um den Nachfolger $\underline{n+1}$ zu konstruieren. Wir schreiben $\underline{\mathbb{N}} = \{\underline{0}, \underline{1}, \underline{2}, \dots\}$.

Die ersten „Zahlen“ sehen so aus:

$$\begin{aligned}
\underline{0} &= \emptyset \\
\underline{1} &= \{\underline{0}\} = \{\emptyset\} \\
\underline{2} &= \{\underline{0}, \underline{1}\} = \{\emptyset, \{\emptyset\}\} \\
\underline{3} &= \{\underline{0}, \underline{1}, \underline{2}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}
\end{aligned}$$

Auf diese Weise erhalten wir alle Zahlen mit elementaren Konstruktionen aus der leeren Menge. Da das recht mühselig ist, werden wir natürliche Zahlen meistens als Zahlen und nicht als Mengen betrachten. Nur in diesem Abschnitt werden wir die obigen Mengen manchmal benutzen.

1.30. Definition. Eine Menge M heißt *endlich*, wenn sie zu einer Menge $\underline{n}$ gleichmächtig ist. In diesem Fall heißt die Zahl n die *Mächtigkeit* von M, geschrieben $n = \#M$. Ansonsten heißt M *unendlich*.

1.31. Bemerkung.

(1) Man kann sich überlegen, dass zwei Mengen $\underline{n}$ und $\underline{m}$ genau dann gleichmächtig sind, wenn $\underline{n} = \underline{m}$. Wegen Bemerkung 1.26 kann jede Menge M zu höchstens einer Menge $\underline{n}$ gleichmächtig sein. Die Schreibweise $\#M$ für endliche Mengen ist also sinnvoll.

(2) Endliche Mengen kann man immer als Aufzählung angeben. Sei etwa $F \colon \underline{n} \to M$ bijektiv, dann schreibe

$$M = \{F(0), \dots, F(n-1)\}$$

Ist M umgekehrt als $\{x_1, \dots, x_n\}$ gegeben, dann hat M höchstens n Elemente, ist also endlich.

(3) Für unendliche Mengen M führen wir die Schreibweise „$\#M = \infty$“ nicht ein, da nicht alle unendlichen Mengen gleichmächtig sind.

1.32. Definition. Es seien $m, n \in \mathbb{N}$, dann gilt m *kleiner oder gleich* n, kurz $m \leq n$, genau dann, wenn $\underline{m} \subset \underline{n}$. Es ist m *kleiner* als n, kurz $m < n$, wenn $m \leq n$ und $m \neq n$ gilt.

1.33. Bemerkung. Aus Bemerkung 1.29 folgt auch, dass $m < n$ genau dann gilt, wenn $\underline{m} \in \underline{n}$. Man beachte den Unterschied in der Notation. Bei „$\subset$“ ist Gleichheit erlaubt, bei „$<$“ jedoch ausgeschlossen.

Der Vergleich von Zahlen führt uns auf den Begriff der Ordnung. Eine Ordnung einer Menge M ist eine *Relation*, das heißt, eine Teilmenge $R \subset M \times M$, die einige zusätzliche Eigenschaften besitzt. Wir sagen „es gilt xRy“ für $x, y \in M$, wenn $(x, y) \in R$.

1.34. Definition. Eine Relation R auf eine Menge M heißt *Halbordnung*, wenn für alle $x, y, z \in M$ gilt:

(O1) $\qquad xRx \qquad$ (*Reflexivität*),

(O2) $\qquad xRy$ und $yRx \implies x = y \qquad$ (*Antisymmetrie*),

(O3) $\qquad xRy$ und $yRz \implies xRz \qquad$ (*Transitivität*).

Eine Halbordnung heißt *Ordnung*, wenn ausserdem für alle $x, y \in M$ gilt:

(O4) $\qquad xRy$ oder $yRx \qquad$ (*Totalität*).

Die Eigenschaften (1)–(4) heißen auch *Ordnungsaxiome.*

1.35. Beispiel.
(1) Sei M eine Menge, dann definiert „$\subset$“ eine Halbordnung auf der Potenzmenge $\mathcal{P}(M)$.
(2) Die Relation „$\in$“ ist nicht transitiv und daher keine Halbordnung, denn es gilt zum Beispiel $a \in \{a, b\}$ und $\{a, b\} \in \{\{a\}, \{a, b\}\}$, aber nicht $a \in \{\{a\}, \{a, b\}\}$.
(3) Die Relation „$\leq$“ auf $\mathbb{N}$ ist eine Ordnung, denn für alle $\ell, m, n \in \mathbb{N}$ gilt
$$\begin{gathered} n \leq n, \\ m \leq n \text{ und } n \leq m \Longrightarrow m = n, \\ \ell \leq m \text{ und } m \leq n \Longrightarrow \ell \leq n, \\ m \leq n \text{ oder } n \leq m. \end{gathered}$$
Auch hier lassen wir den Beweis aus.
(4) Sei M eine Menge. Die Relation „hat höchstens so viele Elemente wie“ ist keine Ordnung auf der Potenzmenge $\mathcal{P}(M)$, denn sei $M = \{1, 2, 3\}$, dann hat $\{1, 2\}$ höchstens so viele Elemente wie $\{2, 3\}$ und umgekehrt, aber beide Mengen sind nicht gleich. Also ist die Antisymmetrie verletzt.

Das zweite Peano-Axiom 1.28 (2) führt uns zur Beweismethode durch vollständige Induktion.

1.36. Satz (Prinzip der *vollständigen Induktion*)**.** *Für jedes $n \in \mathbb{N}$ sei $A(n)$ eine Aussage. Wenn gilt*

(1) *$A(0)$ ist wahr, und*
(2) *aus $A(n)$ folgt $A(n+1)$ für alle $n \in \mathbb{N}$,*

dann ist $A(n)$ für alle $n \in \mathbb{N}$ wahr.

Beweis. Betrachte

$$M = \{\, n \in \mathbb{N} \mid \text{die Aussage } A(n) \text{ ist wahr} \,\}.$$

Nach unseren Annahmen in Abschnitt 1.1 ist das wieder eine Menge, also $M \subset \mathbb{N}$. Aus den Voraussetzungen folgt

(1) $0 \in M$, und
(2) für alle $n \in M$ gilt $n + 1 \in M$.

Aus dem Axiom 1.28 (2) folgt dann $M = \mathbb{N}$. Nach Definition von M gilt $A(n)$ also für alle $n \in \mathbb{N}$. □

Eine andere Art der vollständigen Induktion funktioniert so: Wenn gilt

(1) $A(0)$ ist wahr, und
(2) aus $A(0) \wedge \cdots \wedge A(n)$ folgt $A(n+1)$ für alle $n \in \mathbb{N}$,

dann gilt $A(n)$ für alle $n \in \mathbb{N}$. Das zeigt man, indem man die Aussage

$$B(n) = A(0) \wedge \cdots \wedge A(n)$$

induktiv mit Satz 1.36 beweist.

Beispiele für diese Beweistechnik finden Sie im Analysis-Skript.

Wir haben in Bemerkung 1.29 Zahlen als Mengen rekursiv eingeführt. *Rekursive Definitionen* funktionieren ähnlich wie vollständige Induktion: um eine Abbildung F von $\mathbb{N}$ in eine Menge M anzugeben, reicht es $F(0) \in M$ festzulegen und eine Vorschrift anzugeben, die $F(n+1)$ aus $F(0)$, …, $F(n)$ bestimmt.

Wir führen jetzt die Grundrechenarten auf $\mathbb{N}$ rekursiv ein. Hierbei handelt es sich um *Verknüpfungen*, das heißt, um Abbildungen $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$, etwa

$$+ \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \text{mit} \quad +(m,n) = m + n.$$

1.37. Definition. Die *Addition, Multiplikation* und *Potenzierung* sind für m, $n \in \mathbb{N}$ definiert durch

$$\begin{array}{llcr} (1) & m + 0 = m & \text{und} & m + (n+1) = (m+n) + 1\,, \\ (2) & m \cdot 0 = 0 & \text{und} & m \cdot (n+1) = m \cdot n + m\,, \\ (3) & m^0 = 1 & \text{und} & m^{(n+1)} = m^n \cdot m\,. \end{array}$$

Beispiel. Zwei einfache Rechnungen:

$$3 + 2 = (3+1) + 1 = ((3+0)+1) + 1 = (3+1) + 1 = 4 + 1 = 5\,,$$
$$3 \cdot 2 = (3 \cdot 1) + 3 = ((3 \cdot 0) + 3) + 3 = \cdots = 6\,.$$

1.38. Proposition. *Seien M, N endliche Mengen.*

(1) *Falls* $M \cap N = \emptyset$ *ist, gilt* $\#(M \dot\cup N) = \#M + \#N$.
(2) *Es gilt* $\#(M \times N) = \#M \cdot \#N$.
(3) *Es gilt* $\#\mathrm{Abb}(N, M) = \#M^{\#N}$.

Beweis. Wir beweisen (1) zur Illustration durch vollständige Induktion über die Mächtigkeit $n = \#N$. Es sei $m = \#M$.

Induktionsanfang: Es sei $n = 0$. Nach den Definitionen 1.25 und 1.30 existiert eine bijektive Abbildung von $\emptyset = \underline{0}$ nach N, also gilt $N = \emptyset$. Somit

$$\#(M \dot\cup N) = \#(M \dot\cup \emptyset) = \#M = m = m + 0 = \#M + \#N .$$

Induktionsschritt: Es sei $\#N = n + 1$. Dann existiert eine bijektive Abbildung $F \colon \underline{n+1} = \underline{n} \dot\cup \{\underline{n}\} \to N$. Setze

$$N' = \mathrm{im}(F|_{\underline{n}}) = \{F(\underline{0}), \dots, F(\underline{n-1})\} \qquad \text{und} \qquad x = F(\underline{n}) ,$$

so dass $\#N' = n$. Nach Induktionsvoraussetzung gilt $\#(M \dot\cup N') = m + n$, also existiert eine bijektive Abbildung $G' \colon \underline{m+n} \to M \dot\cup N'$. Wir definieren $G \colon \underline{(m+n)+1} \to M \dot\cup N$ durch

$$G(\underline{k}) = \begin{cases} G'(\underline{k}) & \text{falls } \underline{k} \in \underline{m+n}, \text{ also } k < m+n, \text{ und} \\ x & \text{falls } \underline{k} = \underline{m+n}, \text{ also } k = m+n. \end{cases}$$

Man überzeugt sich leicht, dass G bijektiv ist. Mit Definition 1.37 (1) folgt

$$\#(M \dot\cup N) = (m+n) + 1 = m + (n+1) = \#M + \#N . \qquad \square$$

1.39. Bemerkung. Die Grundrechenarten hätten wir auch über die Eigenschaften (1)–(3) definieren können. Außerdem folgt aus (1), dass $m \le \ell$ genau dann gilt, wenn ein $n \in \mathbb{N}$ mit $m + n = \ell$ existiert.

Bevor wir das Assoziativgesetz kennenlernen, überlegen wir uns, was „Klammern“ eigentlich bewirken. Fassen wir $+\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ als Abbildung auf, dann bedeutet $(\ell + m) + n$ gerade $+(+(\ell, m), n)$, $\ell + (m + n)$ bedeutet $+(\ell, +(m, n))$.

1.40. Satz. *Für* $\ell, m, n \in \mathbb{N}$ *gelten die Rechenregeln*

(1) Assoziativgesetze

$$(\ell + m) + n = \ell + (m + n)$$
$$(\ell \cdot m) \cdot n = \ell \cdot (m \cdot n)$$

(2) Neutrale Elemente

$$n + 0 = n$$
$$n \cdot 1 = n$$

(3) Kommutativgesetze

$$n + m = m + n$$
$$n \cdot m = m \cdot n$$

(4) Distributivgesetz

$$\ell \cdot (m+n) = \ell \cdot m + \ell \cdot n$$

(5) Kürzungsregeln

$$\begin{aligned} \ell + n = m + n \qquad &\Longrightarrow \qquad \ell = m \\ \ell \cdot n = m \cdot n \qquad &\Longrightarrow \qquad \ell = m \textit{ oder } n = 0 \,. \end{aligned}$$

Beweis. Die Aussagen (2) folgen leicht aus Definition 1.37. Alle anderen lassen sich durch vollständige Induktion beweisen. Der Beweis von (5) ist Übung. □

1.3. Ganze und Rationale Zahlen

In diesem Abschnitt „lösen" wir zwei Probleme: man kann in $\mathbb{N}$ nicht subtrahieren, und man kann in $\mathbb{N}$ auch nicht durch Zahlen $n \neq 0$ dividieren. Um diese „Grundrechenarten" einführen zu können, werden wir $\mathbb{N}$ erst zu den ganzen Zahlen $\mathbb{Z}$, und dann zu den rationalen Zahlen $\mathbb{Q}$ erweitern. Dazu ist zunächst etwas Vorarbeit nötig. Wir erinnern uns an die Definition 1.34 einer Halbordung.

1.41. Definition. Eine Relation R auf einer Menge M heißt *Äquivalenzrelation*, wenn für alle x, y, z gilt:

(Ä1) $\quad xRx \quad$ (*Reflexivität*),

(Ä2) $\quad xRy \Longrightarrow yRx \quad$ (*Symmetrie*),

(Ä3) $\quad xRy$ und $yRz \Longrightarrow xRz \quad$ (*Transitivität*).

Im Unterschied zu Halbordnungen sind Äquivalenzrelationen symmetrisch und nicht antisymmetrisch. Das erlaubt uns, Äquivalenzklassen und Quotientenmengen zu definieren. Wir erinnern uns an die Potenzmenge $\mathcal{P}(M)$ von M aus Definition 1.11.

1.42. Definition. Es sei R eine Äquivalenzrelation auf M. Für alle $x \in M$ definieren wir die *Äquivalenzklasse* $[x]$ von x als

$$[x] = \{\, y \in M \mid xRy \,\} \,.$$

Die Gesamtheit aller Äquivalenzklassen bildet die *Quotientenmenge* (kurz: den *Quotienten*) M/R, also

$$M/R = \{\, [x] \mid x \in M \,\} = \{\, N \in \mathcal{P}(M) \mid \text{es gibt ein } x \in M \text{ mit } N = [x] \,\} \,,$$

und alle Elemente $y \in [x]$ heißen *Repräsentanten* von $[x] \in M/R$. Die Abbildung $p\colon M \to M/R$ mit $p(x) = [x]$ heißt *Quotientenabbildung.*

Das einfachste Beispiel für eine Äquivalenzrelation ist die Gleichheit „=" auf einer beliebigen Menge M. Die Axiome (Ä1)–(Ä3) gelten offensichtlich. In diesem Fall ist die Äquivalenzklasse von $x \in M$ gerade $[x] = \{x\}$, und die Quotientenabbildung $p\colon M \to M/{=}$ ist bijektiv mit $x \mapsto \{x\}$. Allerdings gilt strenggenommen nicht $M = M/{=}$, zum Beispiel ist

$$\{1,2,3\}/{=} \;=\; \{\{1\},\{2\},\{3\}\} \,.$$

1.43. Proposition. *Es sei R eine Äquivalenzrelation auf M.*

(1) *Für alle $x \in M$ und alle $y \in [x]$ gilt $[x] = [y]$, insbesondere liegt jedes $x \in M$ in genau einer Äquivalenzklasse von R.*
(2) *Die Abbildung $p\colon M \to M/R$ ist surjektiv, und es gilt $p(x) = p(y)$ genau dann, wenn xRy gilt.*
(3) *Es sei $F\colon M \to N$ eine Abbildung. Dann existiert genau dann eine Abbildung $\bar{F}\colon M/R \to N$ mit $F = \bar{F} \circ p$, wenn für alle x, $y \in M$ aus xRy folgt, dass $F(x) = F(y)$. In diesem Fall ist $\bar{F}$ eindeutig.*

Die Aussage (3) heißt auch die *universelle Eigenschaft des Quotienten.* Wir nennen $\bar{F}$ die *von F induzierte Abbildung.* Wir stellen (3) als Diagramm dar:

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "p"'] & N \\ M/R \, . \arrow[ur, dashed, "\bar{F}"'] & \end{tikzcd}$$

Beweis. Zu (1) seien $y \in [x]$ und $z \in [y]$ beliebig, dann gilt xRy und yRz. Aus Transitivität folgt xRz, also gilt $z \in [x]$ für alle $z \in [y]$, es folgt $[y] \subset [x]$.

Aus xRy folgt yRx wegen der Symmetrie von R, also folgt $x \in [y]$ aus $[y] \in x$. Nach obigem Argument gilt also auch $[x] \subset [y]$, und somit $[x] = [y]$.

Die Surjektivität von p ist klar nach Definition von M/R, und aus (1) folgt, dass $p(x) = [x] = [y] = p(y)$ genau dann, wenn xRy gilt. Also stimmt (2).

In (3) beginnen wir mit „$\Longrightarrow$“. Sei also $\bar{F}\colon M/R \to N$ gegeben mit $F = \bar{F} \circ p$, und seien x, $y \in M$ gegeben mit xRy. Aus (2) folgt $p(x) = p(y)$, also erst recht

$$F(x) = \bar{F}(p(x)) = \bar{F}(p(y)) = F(y) \, .$$

Zu „$\Longleftarrow$“ gelte $F(x) = F(y)$ für alle x, $y \in M$ mit xRy, also für alle $x \in M$ und alle $y \in [x]$. Seien also $[x] \in M/R$ und $y \in [x]$ beliebig, dann dürfen wir $\bar{F}([x]) = F(y)$ setzen. Diese Konstruktion hängt nach Voraussetzung nicht von der Wahl von $y \in [x]$ ab. Dazu sagen wir, $\bar{F}$ ist *wohldefiniert.*

Die Eindeutigkeit von $\bar{F}$ folgt mit Satz 1.22 (7) aus der Surjektivität von p. $\square$

In der Schule definiert man $\mathbb{Z}$, indem man zu $\mathbb{N}$ noch negative Zahlen hinzunimmt:

$$\mathbb{Z} = \mathbb{N} \,\dot{\cup}\, \{\, -n \mid n \in \mathbb{N} \setminus \{0\} \,\}.$$

Anschließend definiert man Addition, Subtraktion und Multiplikation. Dabei muss man immer einige Fälle unterscheiden. Wir beschreiben ganze Zahlen stattdessen als Differenzen natürlicher Zahlen, also als $m - n$ für $m, n \in \mathbb{N}$.

1.44. Bemerkung. Um die folgenden Konstruktionen zu verstehen, hier ein paar Vorüberlegungen. Für alle m, n, p, $q \in \mathbb{N}$ gilt in $\mathbb{Z}$:

$$(m-n) = (p-q) \in \mathbb{Z} \quad \Longleftrightarrow \quad m+q = n+p \in \mathbb{N}\,, \tag{1}$$

$$(m-n) + (p-q) = (m+p) - (n+q)\,, \tag{2}$$

$$-(m-n) = n - m\,, \tag{3}$$

$$(m-n)\cdot(p-q) = (m\cdot p + n\cdot q) - (m\cdot q + n\cdot p)\,, \tag{4}$$

$$(m-n) \le (p-q) \quad \Longleftrightarrow \quad m+q \le n+p\,. \tag{5}$$

Anstelle von $m - n \in \mathbb{Z}$ betrachten wir das Paar $(m,n) \in \mathbb{N} \times \mathbb{N}$. Gemäß Bemerkung 1.44 (1) definieren wir eine Relation $\sim$ auf der Menge $\mathbb{N} \times \mathbb{N}$ durch

$$(m,n) \sim (p,q) \quad \Longleftrightarrow \quad m+q = n+p \in \mathbb{N}.$$

Außerdem definieren wir Addition, Negatives, Multiplikation und eine Relation $\le$ gemäß Bemerkung 1.44 (2)–(5) durch

$$\begin{aligned}
(m,n) + (p,q) &= (m+p, n+q),\\
-(m,n) &= (n,m),\\
(m,n)\cdot(p,q) &= (m\cdot p + n\cdot q, m\cdot q + n\cdot p),\\
(m,n) \le (p,q) \quad &\Longleftrightarrow \quad m+q \le n+p.
\end{aligned}$$

1.45. Proposition. *Es seien m, n, p, q, r, s, t, $u \in \mathbb{N}$. Dann gilt*

(1) *„$\sim$“ ist eine Äquivalenzrelation.*
(2) *Aus $(m,n) \sim (p,q)$ und $(r,s) \sim (t,u)$ folgt*

$$\begin{aligned}
(m,n) + (r,s) &\sim (p,q) + (t,u),\\
(m,n)\cdot(r,s) &\sim (p,q)\cdot(t,u)\\
\text{und } -(m,n) &\sim -(p,q).
\end{aligned}$$

(3) *Aus $(m,n) \sim (p,q)$ und $(r,s) \sim (t,u)$ folgt*

$$(m,n) \le (r,s) \quad \Longrightarrow \quad (p,q) \le (t,u).$$

Beweis. Zu (1): „$\sim$“ ist reflexiv und symmetrisch nach Konstruktion und dem Kommutativgesetz 1.40 (3). Zur Transitivität benutzen wir zusätzlich die Kürzungsregel 1.40 (5):

$$\begin{aligned}
&(m,n) \sim (p,q) \text{ und } (p,q) \sim (r,s)\\
\Longrightarrow\quad &m+q = n+p \text{ und } p+s = q+r\\
\Longrightarrow\quad &m+q+p+s = n+p+q+r\\
\Longrightarrow\quad &m+s = n+r\\
\Longrightarrow\quad &(m,n) \sim (r,s).
\end{aligned}$$

Zu (2): Seien $(m,n) \sim (p,q)$ und $(r,s) \sim (t,u)$, also $m+q = n+p$ und $r+u = s+t$. Wegen $m+q+r+u = n+p+s+t$ folgt

$$(m,n) + (r,s) = (m+r, n+s) \sim (p+t, q+u) = (p,q) + (t,u).$$

Außerdem gilt

$$\begin{aligned} mr + ns + ps + qr &= (m+q)\cdot r + (n+p)\cdot s \\ =&(n+p)\cdot r + (m+q)\cdot s = pr + qs + ms + nr, \end{aligned}$$

also

$$(m,n)(r,s) = (mr+ns, ms+nr) \sim (pr+qs, ps+qr) = (p,q)(r,s).$$

Genauso zeigt man $(p,q)(r,s) \sim (p,q)(t,u)$, und wegen Transitivität gilt $(m,n)(r,s) \sim (p,q)(t,u)$. Die Behauptung $-(m,n) = (n,m) \sim (q,p) = -(p,q)$ ist leicht einzusehen.

Zu (3): Mit $(m,n) \sim (p,q)$ und $(r,s) \sim (t,u)$ wie oben: Aus $(m,n) \leq (r,s)$ folgt $m+s \leq n+r$, also existiert nach Bemerkung 1.39 ein $k \in \mathbb{N}$ mit

$$\begin{aligned} &m+s+k = n+r \\ &\qquad\Longrightarrow\quad m+p+s+u+k = n+p+r+u = m+q+s+t \\ &\qquad\Longrightarrow\quad p+u+k = q+t \quad\Longrightarrow\quad p+u \leq q+t \\ &\qquad\Longrightarrow\quad (p,q) \leq (t,u). \end{aligned}$$ $\square$

Wir definieren also $\mathbb{Z}$ als Quotienten

$$\mathbb{Z} = \mathbb{N}\times\mathbb{N} \,/ \sim\, = \{\, [(m,n)] \mid (m,n) \in \mathbb{N}\times\mathbb{N} \,\} \,.$$

Proposition 1.45 garantiert wegen der universellen Eigenschaft aus 1.43 (3), dass wir mit Äquivalenzklassen rechnen dürfen:

$$\begin{aligned} [(m,n)] + [(p,q)] &= [(m+p, n+q)], \\ [(m,n)] \cdot [(p,q)] &= [(mp+nq, mq+np)], \\ -[(m,n)] &= [(n,m)], \end{aligned}$$

unabhängig von den Repräsentanten $(m,n) \in [(m,n)]$, $(p,q) \in [(p,q)]$. Auch $[(m,n)] \leq [(p,q)]$ ist wohldefiniert.

Konkreter sei $p\colon \mathbb{N}\times\mathbb{N} \to \mathbb{Z}$ die Quotientenabbildung. Wir halten zunächst das Paar $(r,s) \in \mathbb{N}\times\mathbb{N}$ fest und betrachten die Abbildung $F = p \circ (\,\cdot\, + (r,s))$ wie im folgenden Diagramm:

$$\begin{tikzcd} \mathbb{N}\times\mathbb{N} \arrow[r, "\cdot+(r,s)"] \arrow[d, "p"'] \arrow[dr, "F"] & \mathbb{N}\times\mathbb{N} \arrow[d, "p"] \\ \mathbb{Z} \arrow[r, dashed, "\bar F"'] & \mathbb{Z}\,. \end{tikzcd}$$

Also können wir zu einer ganzen Zahl ein festes Paar (r,s) addieren. Jetzt halten wir die ganze Zahl $[(m,n)]$ fest und betrachten die Abbildung $G = [(m,n)] + \,\cdot\, \colon \mathbb{N}\times\mathbb{N} \to \mathbb{Z}$ wie im folgenden Diagramm:

$$\begin{tikzcd} \mathbb{N}\times\mathbb{N} \arrow[d, "p"'] \arrow[dr, "{[(m,n)]+\cdot}"] & \\ \mathbb{Z} \arrow[r, dashed] & \mathbb{Z}\,. \end{tikzcd}$$

Also dürfen wir zwei ganze Zahlen addieren. Mit den analogen zwei Diagrammen erhalten wir auch die Multiplikation.

1.46. Definition. Die Menge $\mathbb{Z} = \mathbb{N} \times \mathbb{N}/\sim$ heißt Menge der *ganzen Zahlen.*

Wir identifizieren $n \in \mathbb{N}$ mit $[(n,0)] \in \mathbb{Z}$ und schreiben $-n$ für $[(0,n)] \in \mathbb{Z}$. Insbsondere schreiben wir $0 = [(0,0)]$ und $1 = [(1,0)]$.

1.47. Satz. *In $\mathbb{Z}$ gelten Assoziativ- und Kommutativgesetz sowohl für die Addition als auch für die Multiplikation. Neutrale Elemente sind 0 für die Addition und 1 für die Multiplikation. Es gilt das Distributivgesetz. Jedes Element $[(m,n)]$ besitzt ein* additives Inverses $-[(m,n)] = [(n,m)]$, *das heißt, es gilt*

$$[(m,n)] + \bigl(-[(m,n)]\bigr) = [(m,n)] + [(n,m)] = [(0,0)].$$

Es gilt die Kürzungsregel für die Multiplikation.

Die Relation „$\le$“ auf $\mathbb{Z}$ ist eine Ordnung, und für alle a, b, $c \in \mathbb{Z}$ gilt:

$$\begin{aligned} a \le b \qquad &\Longrightarrow \qquad a + c \le b + c\,, \\ 0 \le a \text{ und } 0 \le b \qquad &\Longrightarrow \qquad 0 \le ab\,. \end{aligned}$$

Beweis. Das meiste folgt direkt aus Satz 1.40 und den obigen Definitionen. Die neue Gleichung

$$[(m,n)] + \bigl(-[(m,n)]\bigr) = [(m,n)] + [(n,m)] = [(0,0)]$$

ergibt sich aus

$$(m,n) + (n,m) = (m+n, n+m) \sim (0,0).$$

Ähnlich zeigt man die Eigenschaften von „$\le$“. □

Wir haben die natürlichen Zahlen $\mathbb{N}$ zu den ganzen Zahlen $\mathbb{Z}$ erweitert, um additive Inverse zu finden, also Zahlen $-n$ mit $n + (-n) = 0$. Dazu haben wir natürliche Zahlen durch Paare $(m,n) \in \mathbb{N} \times \mathbb{N}$ ersetzt, die für die Zahl $m - n \in \mathbb{Z}$ stehen. Die Zahlen $-n = [(0,n)]$ sind gerade die negativen Zahlen aus der Schule. Der Einfachheit halber schreiben wir ab sofort $a, b, c, \dots \in \mathbb{Z}$, nicht mehr $[(m,n)]$.

Um nun auch multiplikative Inverse $\frac{1}{n}$ mit $n \cdot \frac{1}{n} = 1$ für alle $n \in \mathbb{Z} \setminus \{0\}$ zu erhalten, ersetzen wir ganze Zahlen durch Paare $(p,q) \in \mathbb{Z} \times (\mathbb{N} \setminus \{0\})$, die für Brüche $\frac{p}{q}$ stehen. Das ist die Bruchrechnung, wie wir sie aus der Schule kennen.

Dazu definieren wir für (p,q), $(r,s) \in \mathbb{Z} \times (\mathbb{N} \setminus \{0\})$:

$$(p,q) \sim (r,s) \iff p \cdot s = q \cdot r \qquad \left(\iff \frac{p}{q} = \frac{r}{s}\right),$$

$$(p,q) + (r,s) = (ps+qr, qs) \qquad \left(\text{da } \frac{p}{q} + \frac{r}{s} = \frac{ps+qr}{qs}\right),$$

$$(p,q) \cdot (r,s) = (pr, qs) \qquad \left(\text{da } \frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs}\right),$$

$$-(p,q) = (-p,q) \qquad \left(\text{da } -\frac{p}{q} = \frac{-p}{q}\right),$$

$$(p,q) \le (r,s) \iff p \cdot s \le q \cdot r \qquad \left(\iff \frac{p}{q} \le \frac{r}{s},\ \text{da } q,s > 0\right).$$

Beachte, dass $qs \in \mathbb{N} \setminus \{0\}$, denn aus $qs = 0 = 0 \cdot s$ würde mit der Kürzungsregel entweder $q = 0$ oder $s = 0$ folgen. Für $p \neq 0$ definieren wir:

$$(p,q)^{-1} = \begin{cases} (q,p) & \text{falls } p > 0, \\ (-q,-p) & \text{falls } p < 0. \end{cases}$$

Beachte: die rechte Seite $(\pm q, \pm p)$ liegt immer in $\mathbb{Z} \times (\mathbb{N} \setminus \{0\})$.

1.48. Proposition. (1) *Die Relation „$\sim$“ ist eine Äquivalenzrelation.*

(2) *Es seiem $(m,n), (p,q), (r,s), (t,u) \in \mathbb{Z} \times (\mathbb{N} \setminus \{0\})$ mit $(m,n) \sim (p,q)$ und $(r,s) \sim (t,u)$ gegeben, dann gilt*

$$(m,n) + (r,s) \sim (p,q) + (t,u),$$
$$(m,n) \cdot (r,s) \sim (p,q) \cdot (t,u),$$

und es gilt $m \neq 0 \Rightarrow p \neq 0$, und in diesem Fall

$$(m,n)^{-1} \sim (p,q)^{-1}.$$

(3) *Unter den gleichen Voraussetzungen wie in (2) gilt*

$$(m,n) \le (r,s) \Rightarrow (p,q) \le (t,u).$$

Beweis. Die Beweismethode ist die gleiche wie bei Proposition 1.45, wir lassen den Beweis daher aus, ein Teil ist Übung. □

1.49. Definition. Der Quotient $\mathbb{Z} \times (\mathbb{N} \setminus \{0\})/\sim$ heißt Menge der *rationalen Zahlen* und wird mit $\mathbb{Q}$ bezeichnet. Für die Äquivalenzklasse $[(p,q)]$ schreiben wir $\frac{p}{q}$.

Wie zuvor schließen wir aus Proposition 1.48 (2), dass wir mit Brüchen so rechnen dürfen, wie wir es aus der Schule kennen. Proposition 1.48 (3) besagt, dass wir zwei Brüche vergleichen können.

Wir identifizieren eine ganze Zahl $n \in \mathbb{Z}$ mit dem Bruch $\frac{n}{1} \in \mathbb{Q}$ und fassen $\mathbb{Z}$ als Teilmenge von $\mathbb{Q}$ auf. Insbesondere liegen $0 = \frac{0}{1}$ und $1 = \frac{1}{1}$ in $\mathbb{Q}$.

1.50. Satz. *In $\mathbb{Q}$ gelten die folgenden Rechenregeln:*

(1) *Assoziativgesetz für Addition und Multiplikation*

(2) *neutrale Elemente:* $\frac{p}{q} + 0 = \frac{p}{q}$, $\frac{p}{q} \cdot 1 = \frac{p}{q}$ *für alle* $\frac{p}{q} \in \mathbb{Q}$;
(3) *inverse Elemente:* $\frac{p}{q} + \frac{-p}{q} = 0$ *für alle* $\frac{p}{q} \in \mathbb{Q}$, $\frac{p}{q} \cdot \left(\frac{p}{q}\right)^{-1} = 1$ *für alle* $\frac{p}{q} \in \mathbb{Q} \setminus \{0\}$;
(4) *Kommutativgesetz für Addition und Multiplikation;*
(5) *Distributivgesetz;*
(6) *Die Relation „$\le$" ist eine Ordnung;*
(7) *Aus* $\frac{p}{q} \le \frac{r}{s}$ *folgt* $\frac{p}{q} + \frac{t}{u} \le \frac{r}{s} + \frac{t}{u}$;
(8) *Aus* $0 \le \frac{p}{q}$ *und* $0 \le \frac{r}{s}$ *folgt* $0 \le \frac{p}{q} \cdot \frac{r}{s}$.

Beweis. Diese Aussagen folgen aus den Sätzen 1.40 und 1.47, und aus der Konstruktion von $\mathbb{Q}$. Seien etwa $p, r, t \in \mathbb{Z}$, $q, s, u \in \mathbb{N} \setminus \{0\}$, dann ergibt sich das Assoziativgesetz für die Addition aus

$$\left(\frac{p}{q} + \frac{r}{s}\right) + \frac{t}{u} = \frac{ps + qr}{qs} + \frac{t}{u} = \frac{(ps + qr) \cdot u + qst}{qsu} = \frac{psu + qru + qst}{qsu}$$
$$= \frac{psu + q(ru + st)}{qsu} = \frac{p}{q} + \frac{ru + st}{su} = \frac{p}{q} + \left(\frac{r}{s} + \frac{t}{u}\right) .$$

Betrachten wir das *multiplikative Inverse* $\left(\frac{p}{q}\right)^{-1}$ von $\frac{p}{q} \in \mathbb{Q} \setminus \{0\}$. Wir unterscheiden zwei Fälle:

Falls $0 < p$, gilt $\left(\frac{p}{q}\right)^{-1} = \frac{q}{p}$ und $\frac{p}{q} \cdot \left(\frac{p}{q}\right)^{-1} = \frac{pq}{qp} = 1$.

Falls $p < 0$, gilt $\left(\frac{p}{q}\right)^{-1} = \frac{-q}{-p}$ und $\frac{p}{q} \cdot \left(\frac{p}{q}\right)^{-1} = \frac{p(-q)}{q(-p)} = \frac{-pq}{-qp} = 1$.

Alle anderen Aussagen lassen sich ähnlich beweisen. □

1.4. Etwas Euklidische Geometrie

Der nächste Schritt wäre jetzt die Einführung der reellen Zahlen $\mathbb{R}$. In der Schule definiert man reelle Zahlen als Dezimalbrüche. Diese Konstruktion hat einige Probleme, eines davon ist $0{,}99\ldots = 1$. In der Analysis lernen Sie eine andere Konstruktion kennen. Die reellen Zahlen haben folgende Eigenschaften.

(1) Die reellen Zahlen bilden einen *angeordneten Körper,* das heißt, es gelten alle Rechenregeln aus Satz 1.50.
(2) Die reellen Zahlen sind *archimedisch angeordnet,* das heißt, die natürlichen Zahlen $\mathbb{N}$ sind in $\mathbb{R}$ enthalten, und zu jeder reellen Zahl $r \in \mathbb{R}$ gibt es eine natürliche Zahl $n \in \mathbb{N}$ mit $r \le n$.
(3) Die reellen Zahlen sind *vollständig,* das heißt, er ist der größte Körper, für den (1) und (2) gelten. Genauer: wenn es einen anderen Körper $\Bbbk$ gibt, der (1) und (2) erfüllt und $\mathbb{R}$ enthält, dann gilt bereits $\mathbb{R} = \Bbbk$.
(4) Die rationalen Zahlen $\mathbb{Q}$ liegen *dicht* in $\mathbb{R}$, das heißt, zu r, $s \in \mathbb{R}$ mit $r < s$ existiert $\frac{p}{q} \in \mathbb{Q}$ mit $r \le \frac{p}{q} \le s$.
(5) Addition, Subtraktion, Multiplikation und Division sind *stetig.*

Die Eigenschaften (1)–(3) definieren $\mathbb{R}$ eindeutig (modulo der Probleme, die wir mit der Eindeutigkeit von $\mathbb{N}$ hatten). Es ist nicht offensichtlich, dass Eigenschaft (3) zu der Definition von Vollständigkeit aus der Analysis äquivalent ist. Aber es ist die einfachste Art, Vollständigkeit zu definieren, ohne analytische Begriffe zu verwenden.

In der Schule haben Sie Vektorrechnung wie folgt kennengelernt. Es sei

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ Faktoren}} = \left\{ x = (x_1, \ldots, x_n) \;\middle|\; x_1, \ldots, x_n \in \mathbb{R} \right\} ,$$

dann definiert man eine Vektoraddition $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, eine skalare Multiplikation $\mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ für $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ und $a \in \mathbb{R}$ und einen Nullvektor 0 durch

$$\begin{aligned} x + y &= (x_1 + y_1, \ldots, x_n + y_n) , \\ ax &= (ax_1, \ldots, ax_n) , \\ 0 &= (0, \ldots, 0) . \end{aligned}$$

Um Euklidische Geometrie zu betreiben, definiert man ein Skalarprodukt. Daraus kann man Längen von Vektoren und Winkel zwischen Vektoren ableiten. Für die folgende Definition erinnern wir uns daran, dass die Cosinus-Funktion invertierbar ist als Funktion $\cos\colon [0, \pi] \to [-1, 1]$ mit Umkehrfunktion $\arccos\colon [-1, 1] \to [0, \pi]$. Hierbei messen wir Winkel grundsätzlich in Bogenmaß. Insbesondere gilt

$$1^\circ = \frac{\pi}{180} .$$

1.51. Definition. Wir definieren das *Standard-Skalarprodukt* auf $\mathbb{R}^n$ als Abbildung $\langle\,\cdot\,,\,\cdot\,\rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ für Vektoren x und $y \in \mathbb{R}^n$ durch

(1) $$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n .$$

Die *Euklidische Norm* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ auf dem $\mathbb{R}^n$ ist definiert durch

(2) $$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_n^2} .$$

Für zwei Vektoren x, $y \in \mathbb{R}^n \setminus \{0\}$ definieren wir den *Winkel* durch

(3) $$\measuredangle(x, y) = \arccos \frac{\langle x, y \rangle}{\|x\|\,\|y\|} \in [0, \pi] .$$

Wir sammeln einige wichtige Eigenschaften und Rechenregeln.

1.52. Bemerkung. Seien x, y, $z \in \mathbb{R}^n$ sowie a, $b \in \mathbb{R}$, dann gilt

(1) $$\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle ;$$

(2) $$\langle x, y \rangle = \langle y, x \rangle ;$$

(3) $$\langle x, x \rangle \geq 0 \qquad \text{und} \qquad \langle x, x \rangle = 0 \Longleftrightarrow x = 0 .$$

All das rechnet man leicht nach; für (3) nutzen wir aus, dass $x_1^2, \ldots, x_n^2 \geq 0$. Man sagt, das Skalarprodukt ist *linear* in der ersten Variablen (1), *symmetrisch* (2) und *positiv definit*(3). Aus (1) und (2) folgt, dass das Skalarprodukt auch in der zweiten Variable linear ist, denn

$$\text{(1')} \qquad \langle x, ay + bz \rangle = \langle ay + bz, x \rangle = a\langle y, x \rangle + b\langle z, x \rangle = a\langle x, y \rangle + b\langle x, z \rangle \ .$$

Für den folgenden Satz benötigen wir den reellen *Absolutbetrag* $|\cdot| : \mathbb{R} \to \mathbb{R}$, definiert durch

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

Insbesondere gilt immer $|r| \geq 0$, und $|r| = \sqrt{r^2}$.

1.53. Satz (Cauchy-Schwarz-Ungleichung)**.** *Für alle Vektoren x, $y \in \mathbb{R}^n$ gilt*

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\| \ .$$

Gleichheit gilt genau dann, wenn Zahlen a, $b \in \mathbb{R}$ existieren, die nicht beide Null sind, so dass

$$ax + by = 0 \ .$$

Beweis. Wir machen eine Fallunterscheidung.

Fall 1: Es sei $x = 0$. Dann folgt sofort

$$\langle x, y \rangle = 0 = 0 \cdot \|y\| = \|x\| \cdot \|y\| \ .$$

Also gilt sogar Gleichheit, und mit $a = 1$ und $b = 0$ gilt ebenfalls

$$ax + by = 1 \cdot 0 + 0 \cdot y = 0 \ .$$

Fall 2: Es sei $x \neq 0$, dann ist auch $\|x\|^2 \neq 0$, und wir berechnen

$$\begin{aligned} 0 \leq \left\| y - \frac{\langle x, y \rangle}{\|x\|^2}\, x \right\|^2 &= \left\langle y - \frac{\langle x, y \rangle}{\|x\|^2}\, x, y - \frac{\langle x, y \rangle}{\|x\|^2}\, x \right\rangle \\ &= \|y\|^2 - 2\frac{\langle x, y \rangle}{\|x\|^2}\, \langle x, y \rangle + \frac{\langle x, y \rangle^2}{\|x\|^4}\, \|x\|^2 = \|y\|^2 - \frac{\langle x, y \rangle^2}{\|x\|^2} \ . \end{aligned}$$

Da $\|x\|^2 > 0$, folgt mit elementaren Umformungen

$$\langle x, y \rangle^2 \leq \|x\|^2 \cdot \|y\|^2 \ .$$

Wurzelziehen liefert die Behauptung.

Wegen $x \neq 0$ ist Gleichheit in der Cauchy-Schwarz-Ungleichung äquivalent zu

$$y - \frac{\langle x, y \rangle}{\|x\|^2}\, x = 0 \ .$$

Daraus folgt $ax + by = 0$ mit $b = \|x\|^2 \neq 0$ und $a = -\langle x, y \rangle$.

Umgekehrt sei $ax + by = 0$. Wäre $b = 0$, so würde aus $ax = 0$ und $x \neq 0$ bereits $a = 0$ folgen, aber a und b dürfen nicht beide verschwinden. Also folgt $b \neq 0$ und

$$y = -\frac{a}{b}\, x = -\frac{\left\langle x, \frac{a}{b}\, x \right\rangle}{\|x\|^2}\, x = \frac{\langle x, y \rangle}{\|x\|^2}\, x \ ,$$

und es gilt Gleichheit in der Cauchy-Schwarz-Ungleichung. $\square$

Der Vektor $y - \frac{\langle x,y\rangle}{\|x\|^2}\, x$ im obigen Beweis entspricht dem Lot vom Punkt y auf die Gerade durch 0 mit Richtung x. Insbesondere gilt Gleichheit, wenn der Punkt y auf dieser Geraden liegt.

1.54. Bemerkung. Aus der Cauchy-Schwarz-Ungleichung 1.53 folgt

$$\frac{\langle x,y\rangle}{\|x\|\,\|y\|} \in [-1,1] \subset \mathbb{R} ,$$

also ist der Arcuscosinus in Definition 1.51 (3) erklärt und der Winkel wohldefiniert. Umgekehrt gilt also

$$\langle x,y\rangle = \|x\|\,\|y\| \cos\measuredangle(x,y) . \tag{1}$$

Zur geometrischen Interpretation betrachten wir das Dreieck mit den Endpunkten 0, x und y. Die dritte Seite ist $x - y$, und wir erhalten den Cosinussatz der Euklidischen Geometrie:

$$\|x-y\|^2 = \|x\|^2 - 2\langle x,y\rangle + \|y\|^2 = \|x\|^2 + \|y\|^2 - 2\,\|x\|\,\|y\| \cos\measuredangle(x,y) . \tag{2}$$

1.5. Komplexe Zahlen und die Geometrie der Ebene

In den reellen Zahlen können wir Wurzeln aus positiven Zahlen ziehen, beispielsweise aus 2, was in $\mathbb{Q}$ nicht möglich ist. Man kann aber keine Wurzeln aus negativen Zahlen ziehen. Diesen Missstand wollen wir jetzt beheben, indem wir die reellen Zahlen zu den komplexen Zahlen erweitern.

Die Idee ist, eine neue Zahl i einzuführen, deren Quadrat -1 ist. Wir möchten mit Zahlen $a + bi$ mit a, $b \in \mathbb{R}$ rechnen, und alle von $\mathbb{R}$ vertrauten Rechenregeln sollen gelten. Zum Beispiel sollten die folgenden Rechnungen richtig sein:

$$\begin{aligned} &(a+bi)+(c+di) = a+c+bi+di = (a+c)+(b+d)i ,\\ \text{und}\qquad &(a+bi)\cdot(c+di) = ac+adi+bci+bdi^2 = (ac-bd)+(ad+bc)i . \end{aligned}$$

Um das rigoros zu machen, betrachten wir eine komplexe Zahl als Paar aus zwei reellen Zahlen, und definieren Addition und Multiplikation wie oben. Für eine Menge M und $n \in \mathbb{N}$ bezeichne M^n das n-fache kartesische Produkt von M mit sich selbst, etwa $M^2 = M \times M$.

1.55. Definition. Die *komplexen Zahlen* sind definiert als $\mathbb{C} = \mathbb{R}^2$, mit

$$\begin{aligned} (a,b)+(c,d) &= (a+c,b+d)\\ \text{und}\qquad (a,b)\cdot(c,d) &= (ac-bd,ad+bc) \end{aligned}$$

für alle a, b, c, $d \in \mathbb{R}$.

1.56. Satz. *In $\mathbb{C}$ gelten Assoziativ- und Kommutativgesetz sowohl für die Addition als auch für die Multiplikation. Neutrale Elemente sind $0_{\mathbb{C}} = (0,0)$ für die Addition und $1_{\mathbb{C}} = (1,0)$ für die Multiplikation. Es gilt das Distributivgesetz. Jedes Element (a,b) besitzt ein additives Inverses*

$$-(a,b) = (-a,-b)$$

und, falls $(a,b) \neq 0_{\mathbb{C}}$, ein multiplikatives Inverses

$$(a,b)^{-1} = \left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right) .$$

Beweis. Alle Behauptungen lassen sich direkt mit den Formeln aus Definition 1.55 nachrechnen. Beispielsweise gilt

$$(a,b) \cdot \left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right) = \left(\frac{a^2+b^2}{a^2+b^2}, \frac{-ab+ba}{a^2+b^2}\right) = (1,0) = 1_{\mathbb{C}} . \qquad \square$$

Wir sehen, dass die Abbildung $\mathbb{R} \to \mathbb{C}$ mit $a \mapsto (a,0)$ verträglich mit $+$ und $\cdot$ ist, und 0 und $1 \in \mathbb{R}$ auf $0_{\mathbb{R}}$ und $1_{\mathbb{R}}$ abbildet. Wir dürfen also $\mathbb{R}$ mit den komplexen Zahlen der Form $(\,\cdot\,,0)$ identifizieren. Wenn wir außerdem noch $i = (0,1)$ definieren, können wir uns überzeugen, dass

$$(a,b) = a \cdot (1,0) + b \cdot (0,1) = a + bi$$

für alle a, $b \in \mathbb{R}$ gilt. Damit haben wir unsere Idee vom Anfang des Abschnitts verwirklicht. Außerdem dürfen wir jetzt auch 0 und 1 für $0_{\mathbb{C}}$ und $1_{\mathbb{C}}$ schreiben.

1.57. Bemerkung. Auf $\mathbb{C}$ gibt es keine Ordnung „$\le$“, die zu Satz 1.50 (7) und (8) analoge Eigenschaften hat. Denn gäbe es solch eine Ordnung, dann gälte entweder $0 < x$ oder $0 > x$ für alle $x \neq 0$ wegen Totalität, aber wegen (7) gälte $0 > x$ genau dann, wenn $-x > 0$. Also gälte $x^2 = (-x)^2 > 0$ für alle $x \neq 0$ wegen (8), aber dann erhielten wir wegen (7) und Transitivität einen Widerspruch:

$$0 = 1^2 + i^2 \geq 1^2 > 0 .$$

1.58. Definition. Sei $z = a + bi \in \mathbb{C}$ mit a, $b \in \mathbb{R}$, dann heißt a der *Realteil* $\operatorname{Re}(z)$ von z und b der *Imaginärteil* $\operatorname{Im}(z)$ von z.

Der Imaginärteil ist also immer eine reelle Zahl, und es gilt

$$z = \operatorname{Re}(z) + \operatorname{Im}(z) \cdot i .$$

1.59. Definition. Die Abbildung $\mathbb{C} \to \mathbb{C}$ mit $z \mapsto \bar{z} = \operatorname{Re}(z) - \operatorname{Im}(z) \cdot i$ heißt *komplexe Konjugation,* $\bar{z}$ heißt das (*komplex*) *Konjugierte* von z.

1.60. Bemerkung. Die komplexe Konjugation ist verträglich mit allen Rechenoperationen, das heißt, es gilt

$$\begin{aligned} \bar{z} + \bar{w} &= \overline{z+w} , & \bar{z} \cdot \bar{w} &= \overline{z \cdot w} , \\ \overline{-z} &= -\bar{z} , & \overline{z^{-1}} &= \bar{z}^{-1} , \\ \bar{0} &= 0 , & \bar{1} &= 1 , \end{aligned}$$

auch das rechnet man leicht nach.

Es gilt $\bar{\bar{z}} = z$ für alle z, also ist die komplexe Konjugation ihre eigene Umkehrabbildung. Für eine komplexe Zahl z gilt $z = \bar{z}$ genau dann, wenn $z \in \mathbb{R} \subset \mathbb{C}$.

Man kann die komplexen Zahlen dadurch charakterisieren, dass sie die kleinste Erweiterung der reellen Zahlen $\mathbb{R}$ ist, so dass alle Rechenregeln aus Satz 1.56

gelten und eine Zahl i mit $i^2 = -1$ existiert. Aber i ist dadurch nicht eindeutig bestimmt, denn offensicht sind i und $\bar{i} = -i$ gleichberechtigt.

Die Zahl $z = i$ löst die Gleichung $z^2 + 1 = 0$. In den Übungen werden Sie sehen, dass man $z^2 = w$ für alle komplexen Zahlen w lösen kann. All das sind Spezialfälle des folgenden Satzes.

1.61. Satz (Fundamentalsatz der Algebra)**.** *Es seien $n \geq 1$ und $a_1, \dots, a_n \in \mathbb{C}$, dann existiert $z \in \mathbb{C}$, so dass*

$$z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n = 0 \;.$$

Mit rein algebraischen Methoden lässt sich dieser Satz nicht beweisen. Das liegt daran, dass die reellen Zahlen, die den komplexen ja zugrundeliegen, mit analytischen Mitteln konstruiert wurden. Einen Beweis für diesen Satz lernen Sie daher erst später, zum Beispiel in einer Vorlesung über Funktionentheorie oder Topologie.

Für $z = a + bi$ mit $a, b \in \mathbb{R}$ ist

$$z \cdot \bar{z} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 \geq 0$$

reell. Das ermöglicht folgende Definition.

1.62. Definition. Wir definieren den *Absolutbetrag* (die *Norm* oder die *Länge*) einer komplexen Zahl $z \in \mathbb{C}$ als die reelle Zahl

$$|z| = \sqrt{z \cdot \bar{z}} \geq 0 \;.$$

1.63. Bemerkung. Wir sammeln ein paar Eigenschaften des Absolutbetrages.

(1) Da $|a + bi|^2 = a^2 + b^2$, entspricht $|z| = \|z\|$ der euklidischen Norm auf $\mathbb{C} = \mathbb{R}^2$ aus Definition 1.51 (1).

(2) Unsere Konstruktion von $z^{-1} = \frac{\bar{z}}{|z|^2}$ wird jetzt etwas klarer, denn

$$z \cdot \frac{\bar{z}}{|z|^2} = \frac{|z|^2}{|z|^2} = 1 \;.$$

(3) Der Absolutbetrag ist *multiplikativ,* das heißt, für alle z und w gilt

$$|zw| = \sqrt{zw\,\overline{zw}} = \sqrt{(z\bar{z})(w\bar{w})} = \sqrt{z\bar{z}} \cdot \sqrt{w\bar{w}} = |z|\,|w| \;.$$

(4) Der Absolutbetrag ist *subadditiv* wegen (1) und der Cauchy-Schwarz-Ungleichung 1.53, das heißt, für alle $z, w \in \mathbb{C}$ gilt

$$|z + w| \leq |z| + |w| \;,$$

denn

$$\begin{aligned} |z + w|^2 &= \|z + w\|^2 = \|z\|^2 + \|w\|^2 + 2\langle z, w\rangle \\ &\leq \|z\|^2 + \|w\|^2 + 2\,\|z\|\,\|w\| = (\|z\| + \|w\|)^2 = (|z| + |w|)^2 \;. \end{aligned}$$

(5) Komplexe Konjugation ist mit dem Absolutbetrag verträglich, denn

$$|\bar{z}| = \sqrt{\bar{z}z} = \sqrt{z\bar{z}} = |z| \;.$$

Wir wollen uns Addition und Multiplikation in $\mathbb{C}$ jetzt mit Hilfer der zweidimensionalen Euklidischen Geometrie veranschaulichen. Dazu machen wir einige Anleihen aus der Schulmathematik und identifizieren $\mathbb{C}$ mit dem Vektorraum $\mathbb{R}^2$.

Die Addition in $\mathbb{C}$ entspricht der Vektoraddition in $\mathbb{R}^2$. Die komplexe Konjugation ist eine Spiegelung an der reellen Achse (also an der x-Achse).

Wir schreiben einen Vektor $z \in \mathbb{C} \setminus \{0\}$ als

$$z = |z| \cdot \frac{z}{|z|} .$$

Dann misst $|z| = \|z\|$ die Länge von z. Multiplikation mit $|z| \in \mathbb{R} \subset \mathbb{C}$ entspricht offenbar der Streckung im $\mathbb{R}^2$ mit dem Faktor $|z|$, denn

$$|z| \cdot (a + bi) = (|z| + 0i)(a + bi) = |z|\, a + |z|\, bi .$$

Der Vektor $\frac{z}{|z|}$ hat Länge 1 und beschreibt die Richtung von z. Wir nehmen jetzt an, dass bereits $|z| = 1$ gilt. Es sei φ der Winkel zwischen der positiven reellen Achse $\mathbb{R}_> \subset \mathbb{C}$ (“x-Achse”) und z (entgegen dem Uhrzeigersinn gemessen), so dass

$$z = \cos\varphi + i\, \sin\varphi .$$

Für einen beliebigen Vektor $w = c + di$ folgt

$$z \cdot w = (c\cos\varphi - d\sin\varphi) + (c\sin\varphi + d\cos\varphi)\, i .$$

Sei auf der anderen Seite R_φ die Drehung um den Winkel φ gegen den Uhrzeigersinn mit Zentrum 0. Aus der Schulzeit wissen wir, dass diese Drehung $\mathbb{R}$-linear ist. Für c, $d \in \mathbb{R}$ gilt also

$$\begin{aligned} R_\varphi(c + di) &= c\, R_\varphi(1) + d\, R_\varphi(i) \\ &= c(\cos\varphi + i\, \sin\varphi) + d\,(-\sin\varphi + i\, \cos\varphi) = z \cdot w , \end{aligned}$$

Also beschreibt die komplexe Multiplikation mit einer komplexen Zahl $z = \cos\varphi + i\, \sin\varphi$ vom Betrag 1 genau eine Drehung um φ.

Sei jetzt $z \in \mathbb{C}$, $z \neq 0$, und sei $0 \leq \varphi < 2\pi$ der Winkel zwischen z und der positiven reellen Achse, so dass

$$z = |z| \cdot (\cos\varphi + i\, \sin\varphi) .$$

Der Winkel φ heißt auch das *Argument* von z, geschrieben $\varphi = \arg(z)$, und die obige Schreibweise heißt auch die *Polardarstellung* von z. Dann entspricht Multiplikation mit z einer Drehung um den Winkel φ mit Zentrum 0 und einer anschließenden Streckung um den Faktor $|z|$.

1.64. Bemerkung (Geometrische Interpretation der komplexen Multiplikation)**.** Es seien z, $w \in \mathbb{C} \setminus \{0\}$ Zahlen mit Beträgen $r = |z|$, $s = |w|$ und Argumenten $\varphi = \arg z$ und $\psi = \arg w$. Nach unser Vorüberlegung wird der

Vektor w durch Multiplikation mit z um r gestreckt und um φ gedreht, so dass schließlich

$$|zw| = rs = |z|\,|w| \;, \qquad \arg(zw) = \varphi + \psi = \arg(z) + \arg(w)$$
$$\text{und} \qquad zw = rs\bigl(\cos(\varphi+\psi) + i\sin(\varphi+\psi)\bigr) \;.$$

Für $r = s = 1$ folgen aus der Rechnung

$$\begin{aligned}\cos(\varphi+\psi) + i\sin(\varphi+\psi) &= (\cos\varphi + i\sin\varphi)\cdot(\cos\psi + i\sin\psi)\\ &= (\cos\varphi\cos\psi - \sin\varphi\sin\psi) + i(\cos\varphi\sin\psi + \sin\varphi\cos\psi)\end{aligned}$$

die *Additionstheoreme* für Sinus und Cosinus:

$$\begin{aligned}\cos(\varphi+\psi) &= \cos\varphi\cos\psi - \sin\varphi\sin\psi\\ \text{und} \qquad \sin(\varphi+\psi) &= \cos\varphi\sin\psi + \sin\varphi\cos\psi \;.\end{aligned}$$

Man beachte aber, dass wir uns in dieser ganzen Bemerkung voll und ganz auf unsere Anschauung und unsere Schulkenntnisse in ebener Geometrie verlassen haben. Das reicht nicht als Grundlage für einen strikten Beweis, daher werden wir nach diesem Abschnitt nicht mehr auf diese Überlegungen zurückgreifen. Nichtsdestotrotz wollen wir aber aus den obigen Formeln in den Übungen noch einige interessante Folgerungen ziehen.

1.65. Bemerkung. Die Isometrien der Ebene werden erzeugt von

(1) Verschiebungen $w \mapsto a + w$ mit $a \in \mathbb{C}$,
(2) Drehungen um den Ursprung, $w \mapsto zw$, wobei $z \in \mathbb{C}$ mit $|z| = 1$, und
(3) der Spiegelung an der x-Achse, $w \mapsto \bar{w}$.

Insgesamt können wir also jede Isometrie $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ mit Hilfe komplexer Zahlen schreiben als

$$F(w) = a + zw \qquad \text{oder} \qquad F(w) = a + z\bar{w} \;,$$

wobei a und $z \in \mathbb{C}$ mit $|z| = 1$ durch F eindeutig festgelegt sind.

Es fällt auf, dass der oben benutzte Winkelbegriff nicht ganz mit dem aus dem letzten Abschnitt übereinstimmt. Hier betrachten wir Drehungen gegen den Uhrzeigersinn um beliebige Winkel, wobei der Winkel φ und der Winkel $\varphi + 2\pi n$ für alle $n \in \mathbb{Z}$ die gleiche Drehung beschreiben. Alle Winkel im Intervall

$$(-\pi, \pi] = \bigl\{\, x \in \mathbb{R} \bigm| -\pi < x \leq \pi \,\bigr\}$$

stehen für verschiedene Drehungen, insbesondere entsprechen Winkel $\varphi \in (-\pi, 0)$ Drehungen im Uhrzeigersinn um $|\varphi|$.

In Definition 1.51 (3) hingegen haben wir nur „ungerichtete“ Winkel im Intervall $[0, \pi]$ betrachtet. Besser ging es nicht, da die Winkel φ und $-\varphi$ den gleichen Cosinus haben, und die Funktion Arcus Cosinus sich nach unser Definition für Winkel in $[0, \pi]$ entscheidet.

1.6. Geometrie des Raumes und Quaternionen

Wir geben einen kurzen Abriss der Euklidischen Geometrie des Raumes, inbesondere führen wir das Kreuzprodukt ein. In Analogie zu den komplexen Zahlen definieren wir die Quaternionen, bei denen sowohl Kreuz- als auch Skalarprodukt auf dem $\mathbb{R}^3$ eine wichtige Rolle spielen. Die wichtigsten Eigenschaften der Quaternionen lernen wir später kennen.

1.66. Definition. Das *Kreuzprodukt* (*Vektorprodukt*) auf dem $\mathbb{R}^3$ ist eine Abbildung $\times\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$(u_1, u_2, u_3) \times (v_1, v_2, v_3) = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1) .$$

Beachten Sie, dass das Symbol “$\times$” sowohl das kartesiche Produkt von Mengen ($\mathbb{R}^3 \times \mathbb{R}^3$) als auch das Kreuzprodukt von Vektoren bezeichnet. Missverständnisse wird es deswegen voraussichtlich nicht geben.

1.67. Bemerkung. Für alle u, v, $w \in \mathbb{R}^3$ und alle a, $b \in \mathbb{R}$ gilt

$$(au + bv) \times w = a(u \times w) + b(v \times w) , \tag{1}$$

$$u \times v = -v \times u . \tag{2}$$

All dies folgt unmittelbar aus Definition 1.66. Man sagt, das Kreuzprodukt ist linear im ersten Argument (1) und *antisymmetrisch* (2) .

Wegen (1) und (2) ist das Kreuzprodukt auch im zweiten Argument linear, denn

$$\begin{aligned} \text{(1')} \quad u \times (av + bw) &= -(av + bw) \times u \\ &= -a(v \times u) - b(w \times u) = a(u \times v) + b(u \times w) . \end{aligned}$$

1.68. Satz. *Für alle u, v, w, $t \in \mathbb{R}^3$ gilt*

$$\langle u \times v, w\rangle = \langle v \times w, u\rangle = \langle w \times u, v\rangle , \tag{1}$$

$$(u \times v) \times w = \langle u, w\rangle \cdot v - \langle v, w\rangle \cdot u = w \times (v \times u) , \tag{2}$$

$$0 = (u \times v) \times w + (v \times w) \times u + (w \times u) \times v , \tag{3}$$

$$\langle u \times v, w \times t\rangle = \langle u, w\rangle\langle v, t\rangle - \langle u, t\rangle\langle v, w\rangle \tag{4}$$

Die Gleichung (2) heißt auch *Graßmann-Identität,* und (3) heißt *Jacobi-Identität.* Den Ausdruck $\langle u \times v, w\rangle$ in (1) nennt man auch das *Spatprodukt* der Vektoren u, v, w.

Beweis. Zu (1) berechnen wir

$$\langle u \times v, w\rangle = u_2v_3w_1 - u_3v_2w_1 + u_3v_1w_2 - u_1v_3w_2 + u_1v_2w_3 - u_2v_1w_3 ,$$

und dieser Ausdruck ist invariant unter zyklischer Vertauschung von u, v und w.

Die Graßmann-Identität (2) überprüfen wir nur in der ersten Komponente der ersten Gleichung:

$$\begin{aligned}\big((u \times v) \times w\big)_1 &= (u \times v)_2 \cdot w_3 - (u \times v)_3 \cdot w_2 \\ &= u_3 \cdot v_1 \cdot w_3 - u_1 \cdot v_3 \cdot w_3 - u_1 \cdot v_2 \cdot w_2 + u_2 \cdot v_1 \cdot w_2 \\ &= (u_1 \cdot w_1 + u_2 \cdot w_2 + u_3 \cdot w_3) \cdot v_1 \\ &\qquad - (v_1 \cdot w_1 + v_2 \cdot w_2 + v_3 \cdot w_3) \cdot u_1 \\ &= \langle u, w\rangle \cdot v_1 - \langle v, w\rangle \cdot u_1 \;;\end{aligned}$$

die zweite und dritte Komponente ergeben sich, indem man oben die Indizes 1, 2 und 3 zyklisch vertauscht. Die zweite Gleichung folgt aus der ersten mit Antisymmetrie.

Die Jacobi-Identität (3) folgt, indem man u, v und w in (2) zyklisch permutiert und dann alle drei Gleichungen addiert.

Behauptung (4) folgt aus (1) und (2) durch folgende Rechnung:

$$\begin{aligned}\langle u \times v, w \times t\rangle &= \langle (w \times t) \times u, v\rangle \\ &= \big\langle \langle w, u\rangle \cdot t - \langle t, u\rangle \cdot w, v\big\rangle = \langle u, w\rangle\langle v, t\rangle - \langle u, t\rangle\langle v, w\rangle \;. \qquad \square\end{aligned}$$

1.69. Bemerkung. Wir geben eine geometrische Interpretation.

(1) Satz 1.68 (4) und Bemerkung 1.54 (1) implizieren, dass

$$\begin{aligned}\|u \times v\| &= \sqrt{\|u\|^2 \,\|v\|^2 - \langle u, v\rangle^2} \\ &= \sqrt{\|u\|^2 \,\|v\|^2 \,(1 - \cos^2 \measuredangle(u, v))} = \|u\| \,\|v\| \sin \measuredangle(u, v) \;,\end{aligned}$$

da $\sin^2 + \cos^2 = 1$ und $\sin\varphi \geq 0$ für alle $\varphi \in [0, \pi]$. Also ist $\|u \times v\|$ gerade der Flächeninhalt des von u und v aufgespannten Parallelogramms. Aus Bemerkung 1.67 (2) und Satz 1.68 (1) folgt

$$\langle u \times v, u\rangle = \langle u \times u, v\rangle = 0 \qquad \text{und} \qquad \langle u \times v, v\rangle = \langle v \times v, u\rangle = 0 \;.$$

Also steht $u \times v$ senkrecht auf der Fläche dieses Parallelogramms. Damit haben wir eine geometrische Beschreibung des Kreuzproduktes *bis auf das Vorzeichen.* Das Vorzeichen ergibt sich durch die Wahl einer Orientierung, wie wir später in Beispiel 4.29 lernen werden.

(2) Das Spatprodukt können wir nun als Volumen des Parallelotops mit den Kanten u, v und w interpretieren. Da $u \times v$ senkrecht auf der Grundfläche steht, wird die Höhe dieses Parallelotops gerade gegeben durch

$$\|w\| \,|\cos \measuredangle(u \times v, w)| = \|w\| \,\frac{|\langle u \times v, w\rangle|}{\|u \times v\| \,\|w\|} = \frac{|\langle u \times v, w\rangle|}{\|u \times v\|} \;.$$

Als Produkt aus Grundfläche $\|u \times v\|$ und Höhe erhalten wir das Volumen also als Absolutbetrag $|\langle u \times v, w\rangle|$ des Spatproduktes. Das Vorzeichen des Spatproduktes ist wiederum eine Frage der Orientierung.

Wir erinnern uns an unsere Definition 1.55 der komplexen Zahlen. Dort wurde eine Multiplikation auf $\mathbb{R} \times \mathbb{R}$ erklärt durch

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc) \ .$$

Wir führen jetzt die etwas kompliziertere Quaternionen-Multiplikation ein. Die Quaternionen wurden von Hamilton entdeckt, daher der Buchstabe $\mathbb{H}$.

1.70. Definition. Die *Quaternionen* sind definiert als $\mathbb{H} = \mathbb{R} \times \mathbb{R}^3$, mit

$$\begin{aligned} (a, u) + (b, v) &= (a + b, u + v) \ , \\ (a, u) \cdot (b, v) &= \big(a \cdot b - \langle u, v\rangle, a \cdot v + b \cdot u + u \times v\big) \\ \text{und} \qquad \overline{(a, u)} &= (a, -u) \end{aligned}$$

für alle a, $b \in \mathbb{R}$ und alle u, $v \in \mathbb{R}^3$. Wir identifizieren $a \in \mathbb{R}$ mit $(a, 0) \in \mathbb{H}$ und $u \in \mathbb{R}^3$ mit $(0, u) \in \mathbb{H}$, und definieren Real- und Imaginärteil von (a, u) durch

$$\begin{aligned} \operatorname{Re}(a, u) &= \frac{1}{2}\big((a, u) + \overline{(a, u)}\big) = a \in \mathbb{R} \\ \text{und} \qquad \operatorname{Im}(a, u) &= \frac{1}{2}\big((a, u) - \overline{(a, u)}\big) = u \in \mathbb{R}^3 \ . \end{aligned}$$

1.71. Satz. *In $\mathbb{H}$ gelten Assoziativ- und Kommutativgesetz für die Addition. Die Multiplikation ist assoziativ aber nicht kommutativ. Es gelten Distributivgesetze*

$$(1) \qquad p \cdot (q + r) = p \cdot q + p \cdot r \qquad \textit{und} \qquad (p + q) \cdot r = p \cdot r + q \cdot r$$

für alle p, q, $r \in \mathbb{H}$. Neutrale Elemente sind $0_{\mathbb{H}} = (0, 0)$ für die Addition und $1_{\mathbb{H}} = (1, 0)$ für die Multiplikation. Jedes Element (a, u) besitzt ein additives Inverses

$$(2) \qquad -(a, u) = (-a, -u)$$

und, falls $(a, u) \neq 0_{\mathbb{H}}$, ein multipliktives Inverses

$$(3) \qquad (a, u)^{-1} = \left(\frac{a}{a^2 + \|u\|^2}, -\frac{u}{a^2 + \|u\|^2}\right) .$$

Für ein Quaternion (a, u) gilt

$$(4) \qquad (a, u) \cdot (b, v) = (b, v) \cdot (a, u)$$

für alle $(b, v) \in \mathbb{H}$ genau dann, wenn $(a, u) \in \mathbb{R}$, das heißt, wenn $u = 0$.

Für die Quaternionen-Konjugation gilt

$$(5) \qquad \overline{(a, u) + (b, v)} = \overline{(a, u)} + \overline{(b, v)} \ , \qquad \overline{-(a, u)} = -\overline{(a, u)} \ ,$$

$$(6) \qquad \overline{(a, u) \cdot (b, v)} = \overline{(b, v)} \cdot \overline{(a, u)} \ , \qquad \overline{(a, u)^{-1}} = \overline{(a, u)}^{-1}$$

und es gilt

$$(7) \qquad \overline{(a, u)} \cdot (a, u) = a^2 + \|u\|^2 = (a, u) \cdot \overline{(a, u)} \ .$$

Man beachte, dass wir in (1) zwei Distributivgesetze brauchen, da die Multiplikation in $\mathbb{H}$ nicht kommutativ ist.

Beweis. Die Rechenregeln für die Addition sind leicht zu überprüfen. Die Distributivgesetze (1) folgen aus den Bemerkungen 1.52 (1) und 1.67 (1), zum Beispiel gilt

$$\begin{aligned}
(a, u) \cdot \big((b, v) + (c, w)\big)& \\
&= (a, u) \cdot (b + c, v + w) \\
&= \big(a(b + c) - \langle u, v + w\rangle, a(v + w) + (b + c)u + u \times (v + w)\big) \\
&= (ab - \langle u, v\rangle, av + bu + u \times v) + (ac - \langle u, w\rangle, aw + cu + u \times w) \\
&= (a, u) \times (b, v) + (a, u) \times (c, w) \;.
\end{aligned}$$

Das Assoziativgesetz für die Multiplikation folgt aus Satz 1.68 (1) und (2) und bleibt Übung. Außerdem überprüft man leicht, dass

$$(a, u) + (0, 0) = (a, u) = (a, u) \cdot (1, 0) = (1, 0) \cdot (a, u) \;.$$

Auch die Formel (2) für das additive Inverse ist klar.

Es gelte (4). Aus der Symmetrie des Skalarproduktes und der Antisymmetrie des Kreuzproduktes folgt

$$\begin{aligned}
0 &= (a, u) \cdot (b, v) - (b, v) \cdot (a, u) \\
&= (0, u \times v - v \times u) = (0, 2u \times v) \;.
\end{aligned}$$

Nach Bemerkung 1.69 (1) folgt: wenn (4) für alle (b, v) gilt, dann hat für jeden Vektor $v \in \mathbb{R}^3$ das von u, v aufgespannte Parallelogramm den Flächeninhalt 0. Aber dann muss bereits $u = 0$, also $(a, u) \in \mathbb{R}$ gelten.

Es gilt

$$\begin{aligned}
\overline{(a, u)} \cdot (a, u) &= (a, -u) \cdot (a, u) \\
&= \big(a^2 + \langle u, u\rangle, au - au - u \times u\big) = a^2 + \|u\|^2 \in \mathbb{R} \;,
\end{aligned}$$

und es folgt die erste Gleichung in (7). Die zweite erhalten wir, indem wir u durch $-u$ ersetzen. Aus (7) folgt (3), denn

$$\left(\frac{a}{a^2 + \|u\|^2}, -\frac{u}{a^2 + \|u\|^2}\right) \cdot (a, u) = \frac{1}{\overline{(a, u)} \cdot (a, u)}\, \overline{(a, u)} \cdot (a, u) = 1 \;.$$

Gleichung (5) ist wiederum klar, und (6) folgt aus der Antisymmetrie des Kreuzproduktes, denn

$$\begin{aligned}
\overline{(a, u) \cdot (b, v)} &= (ab - \langle u, v\rangle, -av - bu - u \times v) \\
&= \big(ab - \langle -v, -u\rangle, b(-u) + a(-v) + (-v) \times (-u)\big) \\
&= \overline{(b, v)} \cdot \overline{(a, u)} \;.
\end{aligned}$$

□

1.72. Definition. Wir definieren den *Absolutbetrag* eines Quaternions $q \in \mathbb{H}$ als die reelle Zahl

$$|q| = \sqrt{\bar{q}q} \;.$$

Wegen Satz 1.71 (7) ist das möglich, und für $q = (a, u_1, u_2, u_3) \in \mathbb{H}$ gilt

$$|q|^2 = a^2 + u_1^2 + u_2^2 + u_3^2 ,$$

also stimmt $|q|$ wiederum mit der Euklidischen Norm $\|q\|$ auf $\mathbb{R}^4$ überein.

1.73. Bemerkung. So, wie wir den komplexen Zahlen $(1, 0)$ und $(0, 1)$ die Namen 1 und i gegeben haben, wollen wir hier die folgenden Bezeichnungen einführen:

$$1 = (1, 0) , \qquad i = (0, e_1) , \qquad j = (0, e_2) \qquad \text{und} \qquad k = (0, e_3) .$$

Wir erhalten die Multiplikationstabelle

$$\begin{array}{c|rrr} \cdot & i & j & k \\ \hline i & -1 & k & -j \\ j & -k & -1 & i \\ k & j & -i & -1 . \end{array}$$

Zusammen mit den Distributivgesetzen und Satz 1.71 (4) können wir jetzt alle Quaternionen miteinander multiplizieren.

So wie die komplexen Zahlen die Geometrie der Ebene beschreiben, beschreiben die imaginären Quaternionen die Geometrie des dreidimensionalen Raumes. Wir sehen, dass sowohl das Standard-Skalarprodukt als auch das Kreuzprodukt in der Definition auftauchen, und in der Tat erhalten wir diese zurück als

$$\langle u, v \rangle = \mathrm{Re}(\overline{(0, u)} \cdot (0, v)) \qquad \text{und} \qquad u \times v = \mathrm{Im}((0, u) \cdot (0, v)) .$$

Jetzt wollen wir Isometrien des $\mathbb{R}^3$ mit Hilfe von Quaternionen beschreiben.

1.74. Satz. *Es sei $q = (\cos\varphi, v \sin\varphi) \in \mathbb{H}$, wobei $v \in \mathbb{R}^3$ mit $\|v\| = 1$ und $\varphi \in \mathbb{R}$. Für ein imaginäres $w \in \mathbb{R}^3$ ist $qw\bar{q}$ wieder imaginär. Die Abbildung $F_q \colon \mathbb{R}^3 \to \mathbb{R}^3$ mit $w \mapsto qw\bar{q}$ beschreibt eine Drehung um die Achse durch 0 in Richtung v um den Winkel 2φ.*

Beweis. Ein Quaternion w ist imaginär genau dann, wenn $\bar{w} = -w$ gilt. Wenn w imaginär ist, ist auch $qw\bar{q}$ imaginär, denn

$$\overline{qw\bar{q}} = \bar{\bar{q}}\bar{w}\bar{q} = -qw\bar{q} .$$

Die Abbildung F_q ist $\mathbb{R}$-linear wegen Satz 1.71 (1) und (4). Das gleiche gilt für die Drehung $R_{v,2\varphi}$ um die Achse durch 0 in Richtung v um den Winkel 2φ. Wir zerlegen $w \in \mathbb{R}^3$ wie im Beweis der Cauchy-Schwarz-Ungleichung 1.53 als

$$w = \langle v, w \rangle v + (w - \langle v, w \rangle v) ,$$

so dass der zweite Vektor wegen $\|v\| = 1$ senkrecht auf v steht. Wegen Linearität reicht es, $F_q v = R_{v,2\varphi} v$ und $F_q w = R_{v,2\varphi} w$ für alle Vektoren w mit $|w| = 1$ und $\langle v, w \rangle = 0$ zu zeigen.

Betrachte zunächst v. Wegen $\langle v,v\rangle = 1$ und $v\times v = 0$ gilt in diesem Fall

$$\begin{aligned} qw\bar q &= (\cos\varphi, v\sin\varphi)\cdot(0,v)\cdot(\cos\varphi,-v\sin\varphi)\\ &= \bigl(-\sin\varphi, v\cos\varphi\bigr)\cdot(\cos\varphi,-v\sin\varphi)\\ &= \bigl(-\cos\varphi\sin\varphi+\cos\varphi\sin\varphi, v\sin^2\varphi+v\cos^2\varphi\bigr) = (0,v)\ , \end{aligned}$$

da $\cos^2\varphi+\sin^2\varphi = 1$. Auch die Drehung $R_{v,2\varphi}$ hält v fest, es gilt also $F_q v = v = R_{v,2\varphi}v$.

Es gelte jetzt $\langle v,w\rangle = 0$ und $\|w\| = 1$. Wegen $\langle v\times w, v\rangle = 0$ und der Graßmann-Identität gilt in diesem Fall

$$\begin{aligned} qw\bar q &= (\cos\varphi, v\sin\varphi)\cdot(0,w)\cdot(\cos\varphi,-v\sin\varphi)\\ &= \bigl(0, w\cos\varphi + v\times w\sin\varphi\bigr)\cdot(\cos\varphi,-v\sin\varphi)\\ &= \bigl(0, w\cos^2\varphi + v\times w\cos\varphi\sin\varphi - w\times v\cos\varphi\sin\varphi - (v\times w)\times v\sin^2\varphi\bigr)\\ &= \bigl(0, w(\cos^2\varphi-\sin^2\varphi) + v\times w\cdot 2\cos\varphi\sin\varphi\bigr)\ . \end{aligned}$$

Cosinus und Sinus des doppelten Winkels berechnen sich als

$$\cos(2\varphi) = \cos^2\varphi - \sin^2\varphi \qquad\text{und}\qquad \sin(2\varphi) = 2\cos\varphi\sin\varphi\ .$$

Wenn wir $\|w\|$ annehmen, dann folgt aus Bemerkung 1.69, dass die Vektoren v, w und $v\times w$ aufeinander sekrecht stehen, und dass auch

$$\|v\times w\| = \|v\|\times\|w\|\times\sin\measuredangle(v,w) = 1\ .$$

Insbesondere bilden w und $v\times w$ eine Orthonormalbasis der zu v senkrechten Ebene. Die Drehung $R_{v,2\varphi}$ bildet den Vektor w also ab auf

$$R_{v,2\varphi}w = \cos(2\varphi)\,w + \sin(2\varphi)\,v\times w = F_q w\ . \qquad\square$$

1.75. Bemerkung. Die Isometrien des Raumes werden erzeugt von

(1) Verschiebungen $w\mapsto u+w$ mit $u\in\mathbb{R}^3$,
(2) Drehungen um die Achse durch den Ursprung in Richung v mit Winkel φ, also $w\mapsto F_q w$, wobei jetzt

$$q = \cos\frac{\varphi}{2} + v\,\sin\frac{\varphi}{2}\ ,$$

(3) Die Punktspiegelung $w\mapsto -w$.

In Analogie zu Bemerkung 1.65 können wir also jede Isometrie schreiben als

$$F(w) = u + qw\bar q \qquad\text{oder}\qquad F(w) = u + q\bar w\bar q\ .$$

Dabei sind $u\in\operatorname{Im}\mathbb{H}$ und $q\in\mathbb{H}$ mit $|q| = 1$ durch F fast eindeutig festgelegt — man kann nur noch q durch $-q$ ersetzen. Dieses Phänomen nennt man „Spin“. Es hat sowohl in der Mathematik als auch in der Physik eine Bedeutung.

Die obige Darstellung hat zwei interessante Eigenschaften.

- Sei $G(w) = v + rw\bar{r}$ eine weitere Isometrie, dann hat auch die Verkettung $F \circ G$ die gleiche Form:
$$(F \circ G)(w) = u + q(v + rw\bar{r})\bar{q} = (u + qv\bar{q}) + qr\,w\,\overline{qr}\ .$$
- Anhand der obigen Formel kann man q leicht bestimmen, wenn man Drehachse und -winkel kennt. Umgekehrt man Drehachse und -winkel ablesen, wenn q bekannt ist.

Aufgrund dieser beiden Vorteile werden Quaternionen in der Praxis eingesetzt, zum Beispiel in der Robotersteuerung und in der dreidimensionalen Bildverarbeitung.

1.76. Bemerkung. Analog zu den Bemerkungen 1.65 und 1.75 können wir auch alle Isometrien des $\mathbb{R}^4$ beschreiben durch
$$F(w) = v + pw\bar{q} \qquad \text{oder} \qquad F(w) = v + p\bar{w}\bar{q}\ .$$
Hierbei ist $w \in \mathbb{R}^4 = \mathbb{H}$, und die Quaternionen v, p, $q \in \mathbb{H}$ mit $|p| = |q| = 1$ sind durch F fast eindeutig festgelegt — man darf nur das Paar (p, q) durch das Paar $(-p, -q)$ ersetzen. Es gibt also auch hier einen „Spin". Der Zusammenhang zwischen dem Paar (z, w) und der Gestalt der Isometrie ist nicht so einfach zu erklären wie in Bemerkung 1.75.

Für $\mathbb{R}^n$ mit $n \geq 5$ gibt es leider keine so schönen Beschreibungen der Isometrien mehr. Wir werden im zweiten Semester sehen, wie man Isometrien generell durch Matrizen darstellen kann.

KAPITEL 2

Vektorräume und Moduln

In diesem Kapitel lernen wir mit Vektoren zu rechnen, indem wir Koordinaten angeben und lineare Abbildungen als Matrizen schreiben. Einem Vektor in Koordinaten entspricht ein Element in einem freien Modul, und einer Matrix entspricht eine lineare Abbildung zwischen freien Moduln. Anschließend überlegen wir uns, warum und wie Matrixrechnung funktioniert.

Für das Rechnen mit Matrizen reicht uns zunächst einmal ein Ring, obwohl wir später meistens einen Körper, zum Beispiel $\mathbb{R}$, zugrunde legen werden. Die etwas größere Allgemeinheit verursacht keinen zusätzlichen Aufwand; außerdem müssen wir später gelegentlich mit Matrizen über Ringen arbeiten. Die zahlreichen Vorteile, die die Arbeit über Körpern (auch Schiefkörpern) mit sich bringt, lernen wir dann im nächsten Kapitel kennen.

Als erstes führen wir ein paar algebraische Grundbegriffe ein: Vektoren sind Elemente von Vektorräumen über Körpern oder Schiefkörpern. Etwas allgemeiner ist der Begriff eines Moduls über einem Ring. Und sowohl Ringen als auch Moduln liegen abelsche Gruppen zugrunde, mit denen wir daher beginnen werden. Nachdem wir Moduln eingeführt haben, betrachten wir spezielle „strukturerhaltende“ Abbildungen. Zum Schluss konstruieren wir neue Moduln aus gegebenen und überlegen uns ihre Eigenschaften.

2.1. Gruppen, Ringe, Körper

Wir definieren eine Reihe wichtiger algebraischer Strukturen. Unser Hauptziel sind Körper. Aber auch Gruppen und Ringe werden uns noch häufiger begegnen.

2.1. Definition. Eine *Gruppe* $(G,*)$ ist eine Menge G mit einer Verknüpfung $*\colon G\times G\to G$, für die ein neutrales Element $e\in G$ und für alle $g\in G$ ein inverses Element $g^{-1}\in G$ existiert, so dass für alle g, h und k die folgenden Gruppenaxiome gelten:

(G1) $\quad g*(h*k)=(g*h)*k \quad$ (*Assoziativgesetz*),

(G2) $\quad e*g=g \quad$ (*linksneutrales Element*),

(G3) $\quad g^{-1}*g=e \quad$ (*linksinverse Elemente*).

Eine Gruppe heißt *kommutativ* oder *abelsch,* wenn außerdem für alle g, $h\in G$ gilt

(G4) $\quad g*h=h*g \quad$ (*Kommutativgesetz*).

2.2. Beispiel. Wir kennen schon Beispiele von abelschen Gruppen.

(1) Die ganzen Zahlen $\mathbb{Z}$ bilden eine abelsche Gruppe $(\mathbb{Z}, +)$, genannt die *unendliche zyklische Gruppe,* siehe auch Satz 1.47.
(2) Sei $\Bbbk = \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$. Dann ist $(\Bbbk, +)$ eine abelsche Gruppe, die sogenannte *additive Gruppe* von $\Bbbk$, siehe dazu die Sätze 1.50, 1.56 und 1.71, sowie Punkt (1) am Anfang von Abschnitt 1.4.
(3) Die natürlichen Zahlen $\mathbb{N}$ bilden keine Gruppe, denn es fehlen die inversen Elemente, siehe Übung 3(b) von Blatt 2.

Die Gruppenaxiome sind bewusst sparsam formuliert. Dadurch hat man relativ wenig zu tun, um nachzuweisen, dass eine bestimmte Verknüpfung auf einer Menge eine Gruppe definiert. Beim Rechnen in Gruppen hilft die folgende Definition.

2.3. Proposition. *Sei $(G, *)$ eine Gruppe, dann sind das neutrale Element e und das Inverse g^{-1} zu jedem $g \in G$ eindeutig bestimmt. Außerdem gilt für alle $g \in G$, dass*

$$g * e = g \,, \tag{G2'}$$

$$g * g^{-1} = e \,. \tag{G3'}$$

Insbesondere muss man das neutrale Element und die Abbildung, die einem Gruppenelement sein Inverses zuordnet, in der Notation „$(G, *)$" nicht mit angeben, da beide eindeutig festgelegt sind. Das spart etwas Schreibarbeit. Und wir dürfen tatsächlich von neutralen und inversen Elementen reden, nicht von linksneutralen und linksinversen Elementen.

Beweis. Wir leiten aus den Gruppenaxiomen der Reihe nach einige interessante Rechenregeln ab. Für alle g, h, $k \in G$ gilt

(1) Linkskürzungsregel: aus $g * h = g * k$ folgt $h = k$, denn
$$\begin{aligned} h = e * h = (g^{-1} * g) * h = g^{-1} * (g * h)& \\ = g^{-1} * (g * k) = (g^{-1} * g) * k = e * k = k \,.& \end{aligned}$$
(2) Die Aussage (G2') folgt aus der Linkskürzungsregel (1) und
$$g^{-1} * (g * e) = (g^{-1} * g) * e = e * e = e = g^{-1} * g \,.$$
(3) Eindeutigkeit des neutralen Elements: Es gelte $f * g = g$ für alle $g \in G$, dann folgt insbesondere
$$f = f * e = e \,.$$
Umgekehrt gelte $g * f = g$ für alle $g \in G$, dann folgt ebenfalls
$$f = e * f = e \,.$$
(4) Aussage (G3') folgt aus der Linkskürzungsregel (1) und
$$g^{-1} * (g * g^{-1}) = (g^{-1} * g) * g^{-1} = e * g^{-1} = g^{-1} = g^{-1} * e \,.$$

(5) Rechtskürzungsregel: aus $h * g = k * g$ folgt $h = k$, denn

$$\begin{aligned} h = h * e = h * (g * g^{-1}) &= (h * g) * g^{-1} \\ &= (k * g) * g^{-1} = k * (g * g^{-1}) = k * e = k \;. \end{aligned}$$

(6) Eindeutigkeit des Inversen: aus $g * h = e$ folgt $h = g^{-1}$ wegen der Linkskürzungsregel (1) und

$$g * h = e = g * g^{-1} \;,$$

umgekehrt folgt $k = g^{-1}$ aus $k * g = e$ wegen der Rechtskürzungsregel (2) und

$$k * g = e = g^{-1} * g \;. \qquad \square$$

2.4. Bemerkung. Wir erinnern uns an die Verkettung „$\circ$" von Abbildungen aus Definition 1.19, an die Identität id_M aus Beispiel 1.18 (1) und an die Umkehrabbildungen aus Satz 1.23.

(1) Es seien K, L, M, N Mengen und $F\colon M \to N$, $G\colon L \to M$ und $H\colon K \to L$ Abbildungen,

$$K \xrightarrow{\;H\;} L \xrightarrow{\;G\;} M \xrightarrow{\;F\;} N \;.$$

Dann gilt $F \circ (G \circ H) = (F \circ G) \circ H$, denn für alle $k \in K$ ist

$$\begin{aligned} (F \circ (G \circ H))(k) &= F((G \circ H)(k)) = F(G(H(k))) \\ &= (F \circ G)(H(k)) = ((F \circ G) \circ H)(k) \;. \end{aligned}$$

(2) Für $F\colon M \to N$ wie oben gilt $\mathrm{id}_N \circ F = F$, denn für alle $m \in M$ gilt

$$(\mathrm{id}_N \circ F)(m) = \mathrm{id}_N(F(m)) = F(m) \;.$$

(3) Es sei F bijektiv. Dann existiert eine Umkehrabbildung S nach Satz 1.23, und es gilt

$$S \circ F = \mathrm{id}_M \;.$$

Diese Beziehungen sehen fast so aus wie die Gruppenaxiome (G1)–(G3). Man sollte aber beachten, dass die Abbildungen F, G, H, id_M, id_N und S im Allgemeinen von verschiedenen Typen sind. Das heißt, wenn die Mengen K, L, M, N paarweise verschieden sind, gehören keine zwei dieser Abbildungen zur gleichen Grundmenge, etwa $F \in \mathrm{Abb}(M, N)$, $\mathrm{id}_M \in \mathrm{Abb}(M, M)$, und so weiter.

2.5. Beispiel. Es sei M eine Menge. Wir definieren die Menge der *Automorphismen* von M als

$$\mathrm{Aut}(M) = \bigl\{\, F\colon M \to M \bigm| F \text{ ist bijektiv} \,\bigr\} \;.$$

Dann bildet $(\mathrm{Aut}(M), \circ)$ eine Gruppe. Dazu überlegen wir uns

(1) Seien F und G bijektiv, dann ist $F \circ G$ bijektiv nach Satz 1.22 (3). Also ist die Verknüpfung „$\circ$" auf $\mathrm{Aut}(M)$ wohldefiniert.
(2) Es gilt das Assoziativgesetz (G1) nach Bemerkung 2.4 (1).
(3) Die Identität id_M aus Beispiel 1.18 (1) ist bijektiv. Nach Bemerkung 2.4 (2) ist id_M das neutrale Element in $(\mathrm{Aut}(M), \circ)$.

(4) Das Inverse zu $F \in \operatorname{Aut}(M)$ ist die Umkehrabbildung G aus Satz 1.23. Aus Satz 1.22 (4) und (5) folgt, dass G wieder bijektiv ist, und das Axiom (G3) ist gerade Punkt (3) in Bemerkung 2.4.

Später werden wir häufiger Gruppen begegnen, die aus speziellen bijektiven Abbildungen F einer Menge M in sich bestehen.

2.6. Definition. Ein *Ring* $(R, +, \cdot\,)$ ist eine Menge R mit zwei Verknüpfungen $+,\ \cdot : R \times R \to R$, so dass $(R, +)$ eine abelsche Gruppe bildet, und so dass für alle r, s, $t \in R$ die folgenden Ringaxiome gelten:

(R1) $\qquad (r \cdot s) \cdot t = r \cdot (s \cdot t) \qquad$ (*Assoziativgesetz*),

(R2) $\qquad \begin{cases} r \cdot (s + t) = r \cdot s + r \cdot t \\ (r + s) \cdot t = r \cdot t + s \cdot t \end{cases} \qquad$ (*Distributivgesetze*).

Ein Ring heißt *unitär* oder *Ring mit Eins,* wenn es ein neutrales Element 1_R gibt, so dass für alle $r \in R$ gilt:

(R3) $\qquad 1_R \cdot r = r \cdot 1_R = r \qquad$ (*multiplikatives neutrales Element*).

Ein Ring heißt *kommutativ,* wenn für alle r, $s \in R$ gilt:

(R4) $\qquad r \cdot s = s \cdot r \qquad$ (*Kommutativgesetz*).

Man beachte, dass die Axiome (R3) und (R4) unabhängig voneinander erfüllt sein können. Wir werden in dieser Vorlesung fast nur Ringe mit Eins betrachten.

In allgemeinen Ringen haben wir kein Kommutativgesetz und auch keine Links- oder Rechtskürzungsregeln für die Multiplikation, da uns die multiplikativen Inversen fehlen. Aus diesem Grund brauchen wir beide Gleichungen in (R2) und (R3).

Die Gruppe $(R, +)$ heißt die additive Gruppe des Rings $(R, +, \cdot\,)$. Ihr neutrales Element wird mit 0 oder 0_R bezeichnet, und das additive Inverse von $r \in R$ wird $-r$ geschrieben. Die Bezeichnung r^{-1} ist für multiplikative Inverse reserviert (wenn sie existieren). Das Symbol für die Multiplikation wird häufig weggelassen, somit steht rs kurz für $r \cdot s$.

2.7. Beispiel. Wir kennen bereits einige Ringe.

(1) Die ganzen Zahlen $(\mathbb{Z}, +, \cdot\,)$ bilden einen kommutative Ring mit Eins, siehe Satz 1.47.
(2) Sei $\Bbbk = \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$. Dann ist $(\Bbbk, +, \cdot\,)$ ein Ring mit Eins; siehe dazu die Sätze 1.50, 1.56 und 1.71, sowie Punkt (1) am Anfang von Abschnitt 1.4. Bis auf $\mathbb{H}$ sind diese Ringe auch kommutativ.
(3) Auf den natürlichen Zahlen $\mathbb{N}$ sind zwar Addition und Multiplikation erklärt, und (R1)–(R4) gelten. Aber da $(\mathbb{N}, +)$ keine Gruppe ist, ist $(\mathbb{N}, +, \cdot\,)$ kein Ring, siehe Beispiel 2.2 (3).

Auch aus den Ringaxiomen lassen sich Folgerungen ziehen.

2.8. Proposition. *Es sei $(R, +, \cdot)$ ein Ring. Dann gilt für alle r, $s \in R$, dass*

$$0_R \cdot r = r \cdot 0_R = 0_R \ , \tag{1}$$

$$r \cdot (-s) = (-r) \cdot s = -r \cdot s \ . \tag{2}$$

In einem Ring mit Eins ist die Eins eindeutig, und es gilt entweder $0_R \neq 1_R$, oder aber $R = \{0_R\}$.

Aufgrund der letzten Aussage wird bei einem Ring mit Eins manchmal zusätzlich $0_R \neq 1_R$ gefordert.

Beweis. Aus dem Distributivgesetz (R2) folgt

$$0_R \cdot r = (0_R + 0_R) \cdot r = 0_R \cdot r + 0_R \cdot r \ ,$$

also $0_R = 0_R \cdot r$ nach der Kürzungsregel für die Addition. Genauso folgt $r \cdot 0_R = 0_R$.

Aussage (2) folgt aus

$$0_R = r \cdot 0_R = r \cdot (s + (-s)) = r \cdot s + r \cdot (-s) \ ,$$

genauso erhält man die zweite Gleichung.

Die Eindeutigkeit der Eins folgt wie in Proposition 2.3.

Wenn in einem Ring mit Eins $0_R = 1_R$ gilt, folgt aus (R3) und (1) für alle $r \in R$, dass

$$r = 1_R \cdot r = 0_R \cdot r = 0_R \ . \qquad \square$$

Der Ring $R = \{0\}$ heißt auch *Nullring* oder „trivialer Ring“.

2.9. Beispiel. Sei $n \in \mathbb{N}$, $n \geq 1$. Wir definieren eine Relation „$\equiv \bmod n$“ auf $\mathbb{Z}$ durch

$$a \equiv b \mod n \qquad \Longleftrightarrow \qquad \text{es gibt } k \in \mathbb{Z} \text{ mit } a - b = kn \ ,$$

lies: „a ist kongruent zu b modulo n“.

Wir wollen zeigen, dass es sich um eine Äquivalenzrelation handelt. Die Relation is reflexiv (Ä1), denn $a - a = 0 \cdot n$ für alle $a \in \mathbb{Z}$. Für a, $b \in \mathbb{Z}$ gelte $a - b = kn$ mit $k \in \mathbb{Z}$, dann folgt $b - a = (-k) \cdot n$, also ist die Relation symmetrisch (Ä2). Schließlich ist sie auch transitiv (Ä3), denn gelte $a - b = kn$ und $b - c = \ell n$ für a, b, c, k, $\ell \in \mathbb{Z}$, dann folgt $a - c = (\ell + k) \cdot n$.

Die Äquivalenzklasse von $a \in \mathbb{Z}$ heißt *Restklasse von a* und hat die Form

$$[a] = \{\, a + k \cdot n \mid k \in \mathbb{Z} \,\} = \{\dots, a - n, a, a + n, \dots\} \ .$$

Der Quotient heißt *Menge der Restklassen modulo n* und wird mit $\mathbb{Z}/n$ oder $\mathbb{Z}/n\mathbb{Z}$ bezeichnet. Indem wir $a \in \mathbb{Z}$ mit Rest durch n dividieren, erhalten wir b, $k \in \mathbb{Z}$ mit $0 \leq b < n$, so dass $a = kn + b$. Es folgt

$$\mathbb{Z}/n = \big\{[0], \dots, [n-1]\big\} \ ,$$

insbesondere hat $\mathbb{Z}/n$ die Mächtigkeit n.

Analog zu Abschnitt 1.3 wollen wir zeigen, dass Addition und Multiplikation in $\mathbb{Z}$ auf dem Quotienten $\mathbb{Z}/n$ wohldefinierte Rechenoperationen definieren. Es sei etwa $a-b=kn$ und $c-d=\ell n$, dann folgt

$$\begin{aligned} (a+c)-(b+d) &= (k+\ell)\cdot n\;, \\ (a\cdot c)-(b\cdot d) &= (a-b)\cdot c + b\cdot(c-d) = (kc+b\ell)\cdot n \\ \text{und}\qquad (-a)-(-b) &= (-k)\cdot n\;. \end{aligned}$$

Somit erhalten wir Verknüpfungen $+,\ \cdot : (\mathbb{Z}/n\mathbb{Z})\times(\mathbb{Z}/n\mathbb{Z})\to\mathbb{Z}/n\mathbb{Z}$ sowie $-\,\cdot : \mathbb{Z}/n\mathbb{Z}\to\mathbb{Z}/n\mathbb{Z}$ mit

$$[a]+[c]=[a+c]\;,\qquad [a]\cdot[c]=[a\cdot c]\qquad\text{und}\qquad -[a]=[-a]\;.$$

Schließlich wollen wir die Axiome (G1)–(G4) und (R1)–(R4) überprüfen, um zu zeigen, dass $(\mathbb{Z}/n\mathbb{Z},+,\,\cdot\,)$ ein kommutativer Ring mit Eins ist. Dazu setzen mit $0_{\mathbb{Z}/n\mathbb{Z}}=[0]$ und $1_{\mathbb{Z}/n\mathbb{Z}}=[1]$. Jetzt folgt jedes einzelne der obigen Axiome aus der entsprechenden Rechenregel für $(\mathbb{Z},+,\cdot)$, zum Beispiel

$$\begin{aligned} ([a]+[b])+[c] &= [a+b]+[c] = [(a+b)+c] \\ &= [a+(b+c)] = [a]+[b+c] = [a]+([b]+[c])\;, \\ [a]\cdot([b]+[c]) &= [a]\cdot[b+c] = [a\cdot(b+c)] \\ &= [ab+ac] = [ab]+[ac] = [a]\cdot[b]+[a]\cdot[c] \\ \text{und}\qquad [1]\cdot[a] &= [1\cdot a] = [a] = [a\cdot 1] = [a]\cdot[1]\;. \end{aligned}$$

Somit ist $(\mathbb{Z}/n\mathbb{Z},+,\,\cdot\,)$ ein kommutativer Ring mit Eins. Seine additive Gruppe $(\mathbb{Z}/n\mathbb{Z},+)$ heißt auch die *zyklische Gruppe der Ordnung n.*

2.10. Definition. Ein *Schiefkörper* $(K,+,\,\cdot\,)$ ist ein Ring mit Eins 1_K und additivem neutralem Element 0_K, in dem für alle $k\in K\setminus\{0_K\}$ ein multiplikatives Inverses k^{-1} existiert, so dass für alle $k\in K\setminus\{0_K\}$ die folgenden Körperaxiome gelten:

(K1) $\qquad k^{-1}\cdot k = 1_K \qquad$ (*multiplikatives linksinverses Element*),

(K2) $\qquad 1_K \neq 0_K \qquad$ (*Nichttrivialität*).

Ein Schiefkörper heißt *Körper,* wenn der zugrundeliegende Ring kommutativ ist.

2.11. Beispiel. Wir kennen bereits einige Körper und Schiefkörper.

(1) Es sei $\Bbbk=\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$, dann ist $(\Bbbk,+,\,\cdot\,)$ ein Körper, siehe dazu die Sätze 1.50, 1.56 sowie Punkt (1) am Anfang von Abschnitt 1.4. Insbesondere sind $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ auch Schiefkörper.
(2) Die Quaternionen bilden einen “echten”, also nichtkommutativen Schiefkörper, siehe Satz 1.71.
(3) Die natürlichen Zahlen $(\mathbb{N},+,\,\cdot\,)$ sind kein (Schief-) Körper, da sie noch nicht einmal einen Ring bilden. Die ganzen Zahlen $(\mathbb{Z},+,\,\cdot\,)$ sind zwar ein kommutativer Ring mit Eins, aber kein (Schief-) Körper, da multiplikative Inverse fehlen.

Die Körperaxiome werden in der Literatur oft unterschiedlich formuliert. Manchmal fasst man (G1)–(G4), (R1)–(R4), (K1) und (K2) (oder kleine Variationen davon) zu Axiomen (K1)–(K10) zusammen. Es folgt eine weitere Möglichkeit.

2.12. Proposition. *Eine Menge K mit Verknüpfungen $+$, $\cdot : K \times K \to K$ und Elementen 0_K, $1_K \in K$ bildet genau dann einen Schiefkörper $(K, +, \cdot\,)$, wenn*

(1) $(K, +)$ *eine Gruppe bildet,*
(2) $(K \setminus \{0_K\}, \cdot\,)$ *eine Gruppe bildet, und*
(3) *die Distributivgesetze* (R2) *gelten.*

Falls die Gruppe $(K \setminus \{0_K\}, \cdot\,)$ abelsch ist, ist $(K, +, \cdot\,)$ ein Körper.

Beweis. $\Longrightarrow$: Sei $(K, +, \cdot\,)$ ein Körper, dann ist $(K, +)$ nach den Definitionen 2.6 und 2.10 eine abelsche Gruppe. Auch die Distributivgesetze (R2) haben wir vorausgesetzt, somit gelten (1) und (3).

Zu (2) betrachte a, $b \neq 0_K$. Es gilt $a^{-1} \neq 0$, denn ansonsten wäre

$$1_K = a^{-1} \cdot a = 0_K$$

nach Proposition 2.8 (1), im Widerspruch zu (K2). Es gilt auch $a \cdot b \neq 0_K$, denn sonst wäre

$$b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = 0_K$$

nach Proposition 2.8 (1). Somit definiert die Multiplikation eine Verknüpfung auf der Menge $K \setminus \{0_K\}$, und auch 1_K und die Inversen a^{-1} liegen in $K \setminus \{0_K\}$. Die Gruppenaxiome für $(K \setminus \{0_K\}, \cdot\,)$ folgen jetzt aus (R1), (R3) und (K1).

$\Longleftarrow$: Wenn (1)–(3) erfüllt sind, gelten zunächst einmal (G1)–(G3) und (R2) wegen (1) und (3).

Es gilt $a + b = b + a$ sicher, falls $a = 0$ oder $b = 0$ (wegen (G2) und (G2')), oder falls $a + b = 0$ (wegen (G3) und (G3')). Ansonsten folgt aus dem Axiom (G2) für $(K \setminus \{0_K\}, \cdot\,)$ und den Distributivgesetzen, dass

$$\begin{aligned} a + a + b + b &= (1 + 1) \cdot a + (1 + 1) \cdot b = (1 + 1) \cdot (a + b) \\ &= 1 \cdot (a + b) + 1 \cdot (a + b) = a + b + a + b\ . \end{aligned}$$

Die Kürzungsregeln (1) und (5) aus dem Beweis von Proposition 2.3 liefern (G4).

Das Assoziativgesetz (R1) folgt aus (G1) für die Gruppe $(K \setminus \{0_K\}, \cdot\,)$, falls r, s, $t \in K \setminus \{0_K\}$. Falls mindestens eines der drei Elemente 0_K ist, sind rechte und linke Seite von (R1) auch 0_k wegen Proposition 2.8 (1) — dabei benutzen wir, dass wir im Beweis von Proposition 2.8 das Assoziativgesetz noch nicht benutzt haben. Genauso folgt (R3) aus (G2) und aus (G2') in Proposition 2.3 falls $r \neq 0_K$, und aus Proposition 2.8 (1), falls $r = 0_K$.

Das Axiom (K1) ist gerade (G1) für $(K \setminus \{0_K\}, \cdot\,)$, und (K2) folgt, da $1_K \in K \setminus \{0_K\}$. Also ist $(K, +, \cdot\,)$ ein Körper. $\square$

Wir schreiben $K^\times = K \setminus \{0_K\}$ und nennen $(K^\times, \cdot)$ die *multiplikative Gruppe* von K. Manche Autoren schreiben auch K^*; wir wollen uns das Sternchen aber für andere Zwecke aufsparen.

2.13. Bemerkung. In jedem Körper oder Schiefkörper $(K, +, \cdot)$ gilt Proposition 2.3 für die additive Gruppe $(K, +)$ sowie für die multiplikative Gruppe $(K^\times, \cdot)$. Im Fall $(K^\times, \cdot)$ gelten manche der Aussagen in Proposition 2.3 und ihrem Beweis immer noch, wenn einzelne Elemente 0_K sind. Zur Begründung benutzen wir wieder Proposition 2.8 (1).

(1) *Kürzungsregeln:* Aus $a \cdot b = a \cdot c$ oder $b \cdot a = c \cdot a$ folgt $b = c$ oder $a = 0_K$, genau wie in Satz 1.40 (5).
(2) *Nullteilerfreiheit:* Aus $a \cdot b = 0_K$ folgt $a = 0_K$ oder $b = 0_K$. Das ist äquivalent zu (1).
(3) *neutrales Element:* Es gilt $1 \cdot a = a \cdot 1 = a$ für alle $a \in K$;
(4) *Eindeutigkeit der Eins:* aus $a \cdot b = a$ oder $b \cdot a = a$ für *ein* $a \in K^\times$ und ein $b \in K$ folgt $b = 1$;
(5) *Eindeutigkeit des Inversen:* aus $a \cdot b = 1$ oder $b \cdot a = 1$ für $a, b \in K$ folgt $a, b \in K^\times$ und $b = a^{-1}$.

Unter *Nullteilern* in einem Ring $(R, +, \cdot)$ versteht man Elemente $r, s \in R \setminus \{0\}$ mit $r \cdot s = 0$. Körper sind also *nullteilerfrei* nach (2). In Ringen kann es Nullteiler geben, zum Beispiel gilt

$$[2] \cdot [3] = [6] = [0] \qquad \in \mathbb{Z}/6\mathbb{Z} .$$

2.14. Definition. Sei R ein Ring mit Eins. Falls es eine Zahl $n \in \mathbb{N} \setminus \{0\}$ gibt mit

$$(*) \qquad \underbrace{1_R + \cdots + 1_R}_{n \text{ Summanden}} = 0_R ,$$

dann heißt die kleinste solche Zahl die *Charakteristik* $\chi(R)$ von R. Andernfalls ist $\chi(R) = 0$.

Man beachte, dass aus $\chi(R) = n$ bereits für alle $r \in R$ folgt:

$$\underbrace{r + \cdots + r}_{n \text{ Summanden}} = \bigl(\underbrace{1_R + \cdots + 1_R}_{n \text{ Summanden}}\bigr) \cdot r = 0 .$$

2.15. Beispiel. Für einige Ringe kennen wir die Charakteristik.

(1) Aus dem ersten Peano-Axiom 1.28 (1) folgt für alle $n \in \mathbb{N} \setminus \{0\}$, dass
$$\underbrace{1 + \cdots + 1}_{n \text{ Summanden}} = n \neq 0 .$$
Da $\mathbb{N}$ eine Teilmenge von $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ und $\mathbb{H}$ ist, folgt
$$\chi(\mathbb{Z}) = \chi(\mathbb{Q}) = \chi(\mathbb{R}) = \chi(\mathbb{C}) = \chi(\mathbb{H}) = 0 .$$
(2) Der Ring $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ aus Beispiel 2.9 hat Charakteristik $\chi(\mathbb{Z}/n\mathbb{Z}) = n$.

Aus der Schule kenne wir den Begriff der *Primzahl.* Es sei $1 \leq n \in \mathbb{N}$. Wir nennen $a \in \mathbb{N}$ einen *Teiler* von n, kurz $a \mid n$, wenn es $b \in \mathbb{N}$ mit $ab = n$ gibt. Eine Primzahl ist eine Zahl $p \in \mathbb{Z}$ mit $p > 1$, deren einzige Teiler 1 und p sind. Die Zahl 1 selbst ist keine Primzahl.

2.16. Proposition. *Die Charakteristik eines Körpers ist entweder* 0 *oder eine Primzahl.*

Beweis. Wir wollen annehmen, dass $\chi(K) \neq 0$. Aus (K2) folgt $1_K \neq 0_K$, also ist $\chi(K) \neq 1$. Falls jetzt $\chi(K) = a \cdot b$ mit a, $b > 1$ gilt, betrachte die Gleichung

$$0 = \underbrace{1_K + \dots + 1_K}_{a\cdot b \text{ Summanden}} = \big(\underbrace{1_K + \dots + 1_K}_{a \text{ Summanden}}\big) \cdot \big(\underbrace{1_K + \dots + 1_K}_{b \text{ Summanden}}\big) .$$

Da K als Körper nullteilerfrei ist, muss bereits einer der beiden Faktoren oben 0_K sein. Ohne Einschränkung dürfen wir annehmen, dass es sich um den ersten handelt (ansonsten vertausche a und b). Nun ist aber $a < a \cdot b$ da $1 < b$, und gleichzeitig ist $a \cdot b$ nach Definition 2.14 die kleinste Zahl mit der Eigenschaft (*). Aufgrund dieses Widerspruchs kann $\chi(K)$ kein echtes Produkt sein. □

2.17. Beispiel. Der Ring $\mathbb{Z}/n\mathbb{Z}$ aus Beispiel 2.9 kann also nur ein Körper sein wenn n also eine Primzahl ist.

Sei also p eine Primzahl und $K = \mathbb{Z}/p\mathbb{Z}$. Wir wissen schon, dass $\mathbb{Z}/p\mathbb{Z}$ ein kommutativer Ring mit Eins $[1] \neq [0]$ ist. Wir wollen noch die Existenz multiplikativer Inverser beweisen (K1). Jedes Element $[a] \in \mathbb{Z}/p \setminus \{[0]\}$ hat genau p verschiedene Vielfache in $\mathbb{Z}/p\mathbb{Z}$, denn sonst gäbe es $[b]$, $[c] \in \mathbb{Z}/p$ mit $[b] \neq [c]$ aber $[a] \cdot [b] = [a] \cdot [c]$, also $a \cdot (b - c) = k \cdot p$ für ein $k \in \mathbb{Z}$, aber weder a noch $b - c$ enthalten den Primteiler p, Widerspruch. Also ist die Abbildung $F\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ mit $F([b]) = [a][b]$ injektiv, und daher auch surjektiv (Übung), somit existiert $[b] \in \mathbb{Z}/p\mathbb{Z}$ mit $[a][b] = [1]$, das heißt, $[a]$ hat ein multiplikatives Inverses. Es gibt also *endliche Körper,* das heißt, Körper mit endlich vielen Elementen.

Man kann (K1) auch expliziter beweisen, indem man ein Inverses angibt. Sei dazu $1 \leq a < p$, dann gibt es keine Zahl $c > 1$, die a und p teilt. Nach Satz 2.18 (2) unten für $a_0 = p > a_1 = a$ existieren Zahlen d_0 und $d_1 \in \mathbb{Z}$ mit

$$1 = d_1 a_0 + d_0 a_1 = d_1 p + d_0 a .$$

Dann ist $d_0 a \equiv 1$ modulo p, also ist $[d_0] = [a]^{-1} \in \mathbb{Z}/p\mathbb{Z}$ das multiplikative Inverse von $[a]$.

Für den folgenden Satz brauchen wir *Division mit Rest:* Zu je zwei Zahlen m, $n \in \mathbb{N}$ mit $n \neq 0$ gibt es eindeutige Zahlen q, $r \in \mathbb{N}$ mit $0 \leq r < n$, so dass

$$m = qn + r .$$

2.18. Satz (Euklidischer Algorithmus)**.** *Es seien* a_0, $a_1 \in \mathbb{N} \setminus \{0\}$ *mit* $a_1 \leq a_0$, *Dann existieren eindeutige Zahlen* $i_0 \in \mathbb{N}$, $a_1 > a_2 > \dots > a_{i_0} > a_{i_0+1} = 0$ *und* $b_2, \dots, b_{i_0+1} \in \mathbb{N}$, *so dass*

$$a_{i-1} = b_{i+1} a_i + a_{i+1} \qquad \text{für alle } 1 \leq i \leq i_0 \ . \tag{1}$$

Die Zahl a_{i_0} *ist die größte Zahl in* $\mathbb{N}$, *die* a_0 *und* a_1 *teilt.*

Setze $d_{i_0+1} = 1$, $d_{i_0} = 0$ *und bestimme* $d_{i_0-1}, \dots, d_1, d_0 \in \mathbb{Z}$ *so, dass*

$$d_{i-1} = d_{i+1} - d_i b_{i+1} \qquad \text{für } i_0 \geq i \geq 1 \ . \tag{2}$$

Dann gilt $a_{i_0} = d_1 a_0 + d_0 a_1$.

Die Zahl a_{i_0} heißt der *größte gemeinsame Teiler* von a_0 und a_1, kurz $a_{i_0} = \mathrm{ggT}(a_0, a_1)$.

Beweis. Nach Definition der Division mit Rest existieren die Zahlen a_i und b_i, sind eindeutig bestimmt durch (1) und werden immer kleiner. Also erreichen wir $a_{i_0+1} = 0$ nach $i_0 \leq a_1$ vielen Schritten.

Es sei $0 < c \in \mathbb{N}$ eine Zahl, die a_0 und a_1 teilt, dann teilt c auch alle Zahlen $a_2, \dots, a_{i_0}$ wegen (1). Also kann es keine Zahl größer als a_{i_0} geben, die a_0 und a_1 teilt. Aus (1) für i_0 folgt, dass a_{i_0} auch a_{i_0-1} teilt. Indem wir (1) für immer kleinere i benutzen, folgt, dass a_{i_0} auch $a_{i_0-2}, \dots, a_1$ und a_0 teilt. Also ist $a_{i_0} = \mathrm{ggT}(a_0, a_1)$.

Seien jetzt d_i wie in (2) gegeben. Betrachte die Gleichung

$$a_{i_0} = d_{i+1} a_i + d_i a_{i+1} \ . \tag{3}$$

Wegen $a_{i_0+1} = 0$ und $d_{i_0+1} = 1$ gilt (3) für $i = i_0$. Aus den Gleichungen (1)–(3) für i erhalten wir

$$\begin{aligned} a_{i_0} &= d_{i+1} a_i + d_i (a_{i-1} - b_{i+1} a_i) \\ &= d_i a_{i-1} + (d_{i+1} - d_i b_{i+1}) a_i = d_i a_{i-1} + d_{i-1} a_i \ . \end{aligned}$$

Also gilt (3) auch für $i - 1$. Für $i = 0$ erhalten wir die Behauptung. □

2.19. Bemerkung. Es gibt einen Körper mit n Elementen genau dann, wenn sich $n = p^a$ schreiben lässt, wobei p eine Primzahl ist und $a \geq 1$. Dieser Körper wird F_{p^a} genannt und hat die Charakteristik p. Sie lernen ihn in der Algebra-Vorlesung kennen. Es gibt auch Körper mit Charakteristik p und unendlich vielen Elementen.

Wir sollten in der linearen Algebra immer von Augen haben, dass es diese endlichen Körper gibt; insbesondere Körper der Charakteristik 2 erfordern ein wenig zusätzliche Aufmerksamkeit.

2.2. Moduln und Vektorräume

Gruppen, Ringe und Körper begegnen uns oft dadurch, dass sie auf anderen Strukturen "operieren". Uns interessiert hier zunächst der Fall von Ring- und Körperoperationen; Gruppenoperationen lernen wir später auch noch kennen.

2.20. Definition. Sei $(R, +, \cdot)$ ein Ring. Ein (*Rechts-*) *R-Modul* $(M, +, \,.\,)$ besteht aus einer abelschen Gruppe $(M, +)$ und einer *skalaren Multiplikation* $\,.\,\colon M \times R \to M$, so dass für alle m, $n \in M$ und alle r, $s \in R$ die folgenden Modulaxiome gelten

(M1) $\quad m\,.\,(r \cdot s) = (m\,.\,r)\,.\,s \quad$ (*Assoziativgesetz*),

(M2) $\quad m\,.\,(r + s) = m\,.\,r + m\,.\,s \quad$ (*Erstes Distributivgesetz*),

(M3) $\quad (m + n)\,.\,r = m\,.\,r + n\,.\,r \quad$ (*Zweites Distributivgesetz*).

Sei $(R, +, \cdot)$ ein Ring mit Eins 1. Ein *unitärer* (*Rechts-*) *R-Modul* $(M, +, \,.\,)$ ist ein Rechtsmodul $(M, +, \,.\,)$, so dass zusätzlich gilt:

(M4) $\quad m\,.\,1 = m \quad$ (*Wirkung der Eins*).

Ist der Ring $R = K$ ein Schiefkörper oder Körper, so heißen unitäre Rechts-K-Moduln auch (*Rechts-*) *K-Vektorräume* oder (*Rechts-*) *Vektorräume über K*.

Man beachte, dass das Symbol „+" in (M2) zwei verschiedene Bedeutungen hat. Die Punkte für die Multiplikation kann man oft weglassen. Wir sprechen von Rechts-R-Moduln, weil R durch skalare Multiplikation „von rechts" auf M wirkt. Analog definiert man Links-R-Moduln mit einer skalaren Multiplikation $\cdot\colon R \times M \to M$. In diesem Fall dreht sich in (M1)–(M4) jeweils die Reihenfolge der Faktoren um, beispielsweise würde (M1) zu

$$(r \cdot s) \cdot m = r \cdot (s \cdot m)\ .$$

2.21. Beispiel. Wir können einige Moduln und Vektorräume angeben.

(1) $(R, +, \,.\,)$ ist ein Rechts-R-Modul, wobei "+" und " . " die gleichen Verknüpfungen sind wie in R, jedoch aufgefasst als $+\colon M \times M \to M$ und $\,.\,\colon M \times R \to M$. Nach Definition 2.6 ist nämlich $(R, +)$ eine abelsche Gruppe, (R1) liefert (M1), und die Distributivgesetze (R2) liefern (M2) und (M3). Falls R eine Eins 1 besitzt, ist M auch unitär, denn (M4) folgt dann aus (R3). Völlig analog kann man R zu einem Linksmodul machen.

(2) Der „kleinste" Rechts-R-Modul ist $(\{0\}, +, \,.\,)$ mit $0\,.\,r = 0$ für alle $r \in R$. Er heißt der *Nullmodul.*

(3) Jede abelsche Gruppe A wird zu einem Rechts R-Modul mit $a\,.\,r = 0_A$ für alle $a \in A$ und alle $r \in R$. Damit reduzieren sich (M1)–(M3) zur trivialen Aussage $0_A = 0_A$. Dieser Modul ist allerdings nicht unitär, es sei denn, er wäre bereits der Nullmodul aus (2).

(4) Die Vektorräume $\mathbb{R}^n$, speziell $\mathbb{R}^2$ und $\mathbb{R}^3$ aus den Abschnitten 1.4–1.6 sind Vektorräume über $\mathbb{R}$.

(5) In der Analysis lernen Sie viele $\mathbb{R}$-Vektorräume kennen. So sind die Räume der Folgen und der Nullfolgen mit Werten in $\mathbb{R}$ Vektorräume über $\mathbb{R}$. Auch die Räume der stetigen oder der differenzierbaren Funktionen auf einem Intervall $I \subset \mathbb{R}$ sind $\mathbb{R}$-Vektorräume.

2.22. Proposition. *Es sei $(M, +, .)$ ein $(R, +, \cdot)$-Rechtsmodul. Dann gilt für alle $m \in M$ und $r \in R$, dass*

$$0_M \,.\, r = m \,.\, 0_R = 0_M \;, \tag{1}$$

$$m \,.\, (-s) = (-m) \,.\, s = -m \cdot s \;. \tag{2}$$

Analoge Aussagen gelten für Linksmoduln.

Beweis. All das folgt aus den Distributivgesetzen (M2), (M3) wie im Beweis von Proposition 2.8. □

2.23. Bemerkung. Sei $(R, +, \cdot)$ ein kommutativer Ring, zum Beispiel ein Körper. Dann kann man aus jedem Rechts-R-Modul $(M, +, .)$ einen Links-R-Modul $(M, +, \cdot)$ machen und umgekehrt, indem man $r \cdot m = m \,.\, r$ für alle $r \in R$ und $m \in M$ setzt. Das einzige fragliche Axiom ist (M1), und wir rechnen nach, dass

$$s \cdot (r \cdot m) = (m \,.\, r) \,.\, s = m \,.\, (r \cdot s) = (r \cdot s) \cdot m = (s \cdot r) \cdot m$$

für alle r, $s \in R$ und $m \in M$. Wir dürfen in diesem Fall also einfach von *Moduln* reden.

Da wir im letzten Schritt das Kommutativgesetz (R4) benutzt haben, zeigt diese Rechnung aber auch, dass wir über einem nicht kommutativen Ring genau zwischen Links- und Rechtsmoduln unterscheiden müssen.

Abbildung 1 gibt einen Überblick über die bis jetzt definierten algebraischen Strukturen. Der Übersicht halber haben wir nicht-unitäre Moduln von (Schief-) Körpern und Ringen mit Eins weggelassen.

Es sei $(A, +)$ eine abelsche Gruppe, $n \in \mathbb{N}$, und $a_1, \dots, a_n \in A$. Wir setzen $s_0 = 0$ und definieren induktiv

$$s_i = s_{i-1} + a_i \in A \qquad \text{für } i = 1, \dots, n \;.$$

Dann ist die *Summe der a_n für i von 1 bis n* definiert als

$$\sum_{i=1}^{n} a_i = s_n = a_1 + \dots + a_n \in A \;.$$

Allgemeiner sei I eine Menge. Unter einer *Familie in A mit Indexmenge I* verstehen wir eine Abbildung $a \colon I \to A$, geschrieben $(a_i)_{i \in I}$, mit $i \mapsto a_i$. Wir schreiben $A^I = \operatorname{Abb}(I, A)$ für die Menge aller Familien. Beispielsweise ist eine *Folge* in A gerade eine Familie mit Indexmenge $\mathbb{N}$, und $\mathbb{R}^{\mathbb{N}}$ ist die Menge der reellwertigen Folgen. Wir sagen $a_i = 0$ *für fast alle* $i \in I$, wenn nur endlich viele $i \in I$ nicht auf 0_A abgebildet werden, das heißt, wenn die Menge

$$J = \{\, i \in I \mid a_i \neq 0 \,\}$$

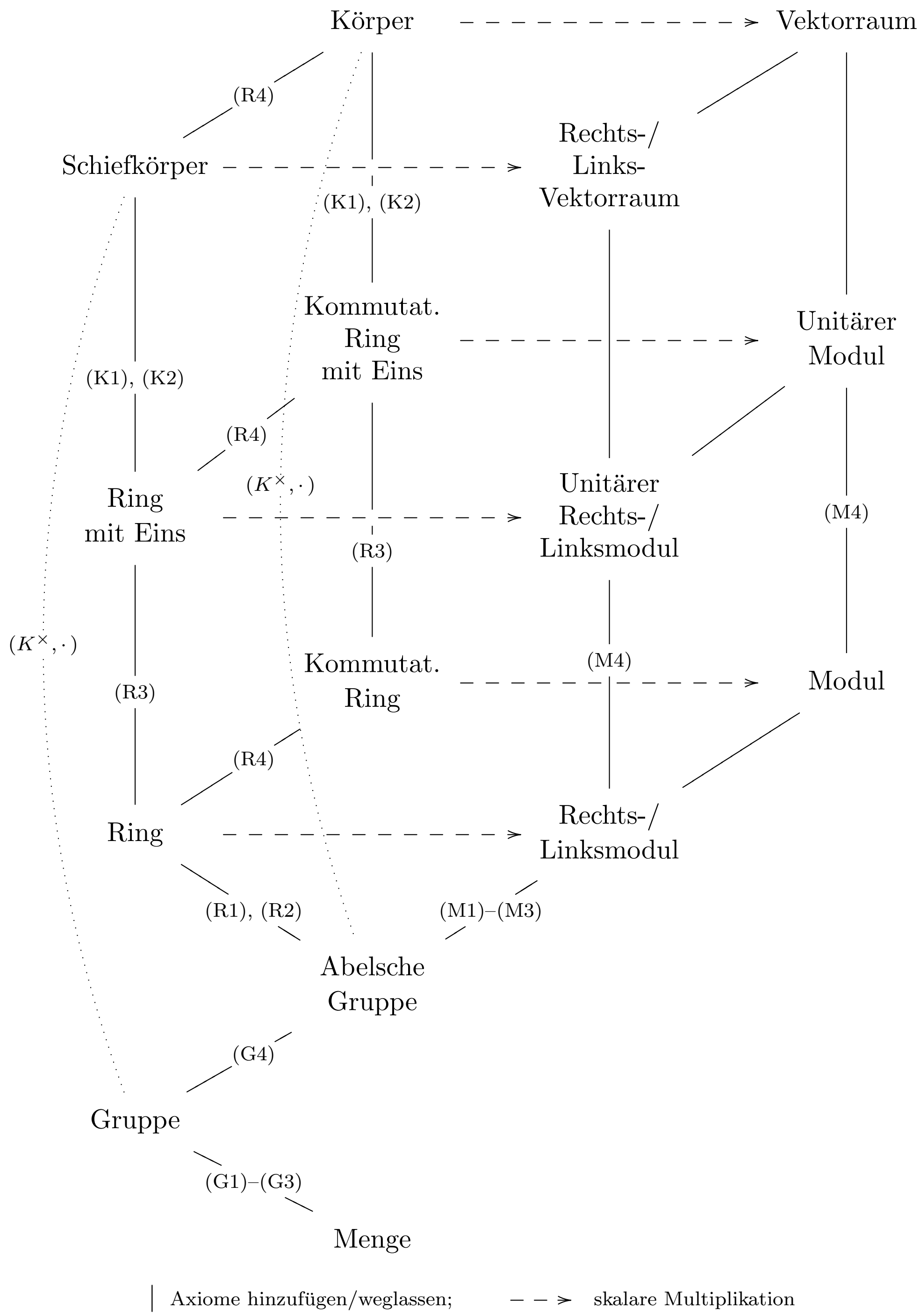

Abbildung 1. Strukturen aus den Abschnitten 2.1 und 2.2

endlich ist. Dann sei $i\colon \{1,\dots,\#J\} \to J$ eine bijektive Abbildung und

$$\sum_{i\in I} a_i = s_n = \sum_{j=1}^{\#J} a_{i(j)} \in A\;.$$

Somit ist $\sum_{i\in I} a_i$ die (endliche) *Summe* derjenigen a_i mit $i \in I$, die von 0_A verschieden sind. Wegen des Kommutativgesetzes (G4) für die Addition in R kommt es dabei nicht auf Reihenfolge der Summation an. Das Ergebnis hängt also nicht von der Wahl der Abbildung i ab. Wir lesen „Summe der a_i für $i \in I$."

Obwohl wir unendliche Indexmengen erlauben, betrachten wir in Wirklichkeit nur endliche Summen, da wir verlangen, dass fast alle Summanden 0_A sind. Man beachte den Unterschied zur Analysis, wo auch gewisse unendliche Summen erlaubt sind.

2.24. Definition. Sei M ein Rechts-R-Modul, und sei $E \subset M$ eine Teilmenge. Sei $(r_e)_{e\in E} \in R^E$, so dass $r_e = 0_R$ für fast alle $e \in E$, dann heißt

$$\sum_{e\in E} e\,.\,r_e \qquad \in M$$

eine *Linearkombination* der $e \in E$. Ein Element $m \in M$ heißt *als Linearkombination der $e \in E$ darstellbar,* wenn es $(r_e)_{e\in E} \in R^E$ mit 0_R für fast alle $e \in E$ gibt, so dass $m = \sum_{e\in E} e\,.\,r_e$. Das *Erzeugnis* von E (*über* R) ist die Menge

$$\langle\, E\,\rangle = \left\{ \sum_{e\in E} e\,.\,r_e \;\middle|\; r_e \in R \text{ mit } r_e = 0_R \text{ für fast alle } e \in E \right\}.$$

Falls $M = \langle\, E\,\rangle$, heißt E eine *Erzeugermenge* von M, und E *erzeugt* M (*über* R). Falls es eine endliche Menge E gibt, die M erzeugt, heißt M *endlich erzeugt* (*über* R).

Hier haben wir immer die Summe der Familie $(e\,.\,r_e)_{e\in E}$ gebildet. Beachte also: falls E endlich ist, dürfen wir jede beliebige Familie $(r_e)_{e\in E}$ in R zur Bildung einer Linearkombination heranziehen. Falls E unendlich ist, müssen wir $r_e = 0$ für fast alle $e \in E$ fordern, damit die Summe endlich bleibt.

Das Erzeugnis einer Menge E wird manchmal auch mit $\operatorname{span}(E)$ bezeichnet.

2.25. Bemerkung. Linearkombinationen werden uns regelmäßig begegnen. Zum Beispiel haben wir im Beweis der Cauchy-Schwarz-Ungleichung 1.53 eine Linearkombination $y - (\langle x,y\rangle / \|x\|^2)\,x$ der Vektoren x und y betrachtet, die senkrecht auf x steht. Im Beweis von Satz 1.74 haben wir einen Vektor $w \in \mathbb{R}^3$ mit $\langle v,w\rangle = 0$ um die Achse durch v gedreht und das Ergebnis als Linearkombination der Vektoren w und $v \times w$ geschrieben.

2.26. Beispiel. Es sei R ein Ring mit Eins, und es sei M ein (Rechts-) R-Modul. Dann ist die zugrundeliegende Menge M selbst immer eine Erzeugermenge, denn

$$m = m\,.\,1 = \sum_{n\in M} n\,.\,\delta_{mn}\;.$$

Hierbei ist das *Kronecker-Symbol* δ definiert durch

$$\delta_{ij} = \begin{cases} 1_R & \text{falls } i = j, \text{ und} \\ 0_R & \text{sonst,} \end{cases}$$

insbesondere ist $\delta_{mn} = 0_R$ für fast alle $n \in M$.

2.27. Beispiel. Als Vektorraum über $\mathbb{C}$ wird $\mathbb{C}$ selbst erzeugt von der Menge $\{1\}$. Wir können $\mathbb{C} \cong \mathbb{R}^2$ aber auch als Vektorraum über $\mathbb{R}$ auffassen. Dann erzeugt $\{1\}$ über R nur die Teilmenge $\mathbb{R} \subset \mathbb{C}$, während $\{1, i\}$ eine Erzeugermenge über $\mathbb{R}$ ist. Aus diesem Grund ist es manchmal sinnvoll, den zugrundeliegenden Ring oder Körper mit anzugeben.

Noch schlimmer wird es, wenn wir $\mathbb{C}$ als Vektorraum über $\mathbb{Q}$ auffassen. Da $\mathbb{Q}$ abzählbar ist und $\mathbb{R}$ und $\mathbb{C}$ überabzählbar sind, ist $\mathbb{C}$ über $\mathbb{R}$ endlich erzeugt, aber nicht über $\mathbb{Q}$.

2.28. Definition. Es sei M ein Rechts-R-Modul und $E \subset M$. Falls

$$0_M = \sum_{e \in E} e \,.\, r_e \qquad \Longrightarrow \qquad r_e = 0_R \text{ für alle } e \in E$$

für alle Familien $(r_e)_{e \in E} \in R^E$ gilt, bei denen $r_e = 0$ für fast alle $e \in E$, dann heißt E *linear unabhängig.* Andernfalls heißt E *linear abhängig.*

Sei M ein Rechts-R-Modul. Eine (*ungeordnete*) *Basis* von M ist eine linear unabhängige Erzeugermenge $E \subset M$ von M. Ein Rechts-R-Modul M heißt *frei* (*über* R), wenn er eine Basis besitzt.

2.29. Beispiel. Es sei $n \geq 1$. Wir können $M = \mathbb{Z}/n\mathbb{Z}$ als unitären $\mathbb{Z}$-Modul auffassen. Dazu definieren wir eine skalare Multiplikation durch $[a] \,.\, r = [ar]$ für alle a, $r \in \mathbb{Z}$. Mit analogen Überlegungen wie in Beispiel 2.9 folgt, dass das wohldefiniert ist, und dass die Modulaxiome gelten.

Für alle $[a] \in \mathbb{Z}/n\mathbb{Z}$ gilt $[a] \,.\, n = [an] = [0]$, also ist jede nichtleere Teilmenge $E \subset \mathbb{Z}/n\mathbb{Z}$ linear abhängig. Genauer: sei $f = [a] \in E$, dann wähle $(r_e)_{e \in E} = (\delta_{ef} \cdot n)_{e \in E}$; es folgt

$$\sum_{e \in E} e \,.\, (\delta_{ef} \cdot n) = [a] \,.\, n = [0] \;,$$

da der Faktor δ_{ef} in einer Summe über e nach Definition des Kronecker-Symbols nur den Summanden mit $e = f$ übriglässt.

Auf der anderen Seite erzeugt die leere Menge den Modul $\mathbb{Z}/n\mathbb{Z}$ nur dann, wenn $n = 1$. Somit ist $\mathbb{Z}/n\mathbb{Z}$ nicht frei über $\mathbb{Z}$, wenn $n > 1$.

Allerdings ist $M = \mathbb{Z}/n\mathbb{Z}$ ein freier Modul über dem Ring $R = \mathbb{Z}/n\mathbb{Z}$ mit Basis $E = \{[1]\}$, denn E erzeugt M. Aus $[0] = [1] \,.\, r$ folgt $r = [0]$, da $[1]$ gleichzeitig das Einselement von R ist. Also ist E aus linear unabhängig über R. Aus diesem Grund empfiehlt es sich auch bei linearer Abhängigkeit, im Zweifelsfall den Grundring mit anzugeben.

2.30. Beispiel. Es sei I eine Menge und R ein Ring mit Eins. Wir definieren einen Rechts-R-Modul $R^{(I)}$ durch

$$\begin{aligned} R^{(I)} &= \big\{\, (r_i)_{i\in I} \in R^I \;\big|\; r_i = 0 \text{ für fast alle } i \in I \,\big\} \;, \\ (r_i)_{i\in I} + (s_i)_{i\in I} &= (r_i + s_i)_{i\in I} && \text{für alle } (r_i)_{i\in I},\, (s_i)_{i\in I} \in R^{(I)} \;, \\ (r_i)_{i\in I} \,.\, s &= (r_i \cdot s)_{i\in I} && \text{für alle } (r_i)_{i\in I} \in R^{(I)},\, s \in R \;. \end{aligned}$$

Addition und skalare Multiplikation nehmen wieder Werte in $R^{(I)}$ an: seien etwa $(r_i)_{i\in I}$, $(s_i)_{i\in I}$ wie oben, dann gibt es nur endlich viele Indizes $i \in I$, an denen $r_i \neq 0$ oder $s_i \neq 0$ gilt; an allen anderen Stellen gilt $r_i + s_i = 0_R$. Das neutrale Element ist die Familie $0_{R^{(I)}} = (0_R)_{i\in I}$, die an allen $i \in I$ den Wert 0_R hat. Jetzt lassen sich die Modulaxiome (M1)–(M3) leicht überprüfen. Wenn R ein Ring mit Eins ist, ist $R^{(I)}$ sogar unitär, das heißt, es gilt auch (M4).

Ab jetzt nehmen wir an, dass R ein Ring mit Eins ist. Für alle $j \in I$ sei

$$e_j = (\delta_{ij})_{i\in I} \in R^{(I)} \tag{1}$$

die Familie, die genau an der Stelle $j \in I$ den Wert 1_R hat, und sonst überall 0_R. Dann ist die Teilmenge

$$E = \big\{\, e_j \in R^{(I)} \;\big|\; j \in I \,\big\} \subset R^{(I)} \;.$$

eine Erzeugermenge, denn für alle $(r_i)_{i\in I} \in R^{(I)}$ gilt

$$\sum_{i\in I} e_i \,.\, r_i = \sum_{j\in I} e_j \,.\, r_j = \sum_{j\in I} (\delta_{ij})_{i\in I} \,.\, r_i = \Bigg(\sum_{j\in I} \delta_{ij} \cdot r_j\Bigg)_{i\in I} = (r_i)_{i\in I} \;. \tag{2}$$

Außerdem ist $r_i = 0$ für fast alle $i \in I$ nach Definition von $R^{(I)}$, so dass wir die obigen Summen bilden dürfen.

Die Teilmenge E ist auch *linear unabhängig,* denn sei

$$\sum_{j\in I} e_j \,.\, r_j = 0_{R^{(I)}} = (0_R)_{i\in I} \;,$$

dann folgt

$$\Bigg(\sum_{j\in I} \delta_{ij} \cdot r_j\Bigg)_{i\in I} = (0_R)_{i\in I} \;,$$

also ergibt jede einzelne Summe den Wert 0_R. Nach Definition von δ_{ij} folgt für den i-ten Eintrag, dass

$$0_R = \sum_{j\in I} \delta_{ij} \cdot r_j = r_i \;.$$

Da das für alle $i \in I$ gelten muss, gilt $r_i = 0$ für alle i, und somit ist E linear unabhängig.

Der Modul $R^{(I)}$ heißt auch der *von I erzeugte freie Rechts-R-Modul.* Er ist frei mit der *Standardbasis* E, und man beachte, dass jedem $i \in I$ genau ein Basiselement e_i entspricht. Mitunter identifiziert man i und e_i, schreibt also

$$(r_{i\in I})_{i\in I} = \sum_{i\in I} i \,.\, r_i \;;$$

das geht aber nur, wenn dadurch keine Misverständnisse entstehen.

Eine analoge Konstruktion liefert den von I erzeugten freien Links-R-Modul ${}^{(I)}R$; nach Bemerkung 2.23 dürfen wir beide identifizieren, falls R kommutativ ist.

Man beachte den Unterschied zwischen R^I und $R^{(I)}$. Beispielsweise ist $\mathbb{R}^{\mathbb{N}}$ der Vektorraum aller reellwertigen Folgen, während $\mathbb{R}^{(\mathbb{N})}$ nur diejenigen Folgen enthält, bei denen ab einer bestimmten Stelle alle Einträge 0 sind. Insbesondere ist die Menge $\{\,(\delta_{mn})_{n\in\mathbb{N}} \mid m \in \mathbb{N}\,\}$ der Folgen, bei denen genau ein Eintrag 1 und alle anderen 0 sind, nur eine Basis von $\mathbb{R}^{(\mathbb{N})}$, nicht vom Raum aller Folgen $\mathbb{R}^{\mathbb{N}}$. Man kann sogar zeigen, dass eine Basis von $\mathbb{R}^{\mathbb{N}}$ überabzählbar viele Elemente haben muss.

2.31. Beispiel. Wir betrachten den Spezialfall $I = \{1, \dots, n\}$ für $n \in \mathbb{N}$. In diesem Fall schreiben wir R^n für $R^{(I)}$, und stellen die Elemente als Spalten dar:

$$\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = (r_i)_{i\in I} = (r_i)_{i=1,\dots,n} \quad \in \quad R^n = R^{(I)} .$$

Die Rechenoperationen sind dann gegeben durch

$$\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} + \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} r_1 + s_1 \\ \vdots \\ r_n + s_n \end{pmatrix} \qquad \text{und} \qquad \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} . s = \begin{pmatrix} r_1 \cdot s \\ \vdots \\ r_n \cdot s \end{pmatrix} .$$

Die Basis $\{e_1, \dots, e_n\}$ heißt *Standardbasis* des R^n und besteht aus den *Standardbasisvektoren*

$$e_1 = \begin{pmatrix} 1_R \\ 0_R \\ \vdots \\ 0_R \end{pmatrix}, \dots, e_n = \begin{pmatrix} 0_R \\ \vdots \\ 0_R \\ 1_R \end{pmatrix} .$$

Der Vektor e_j hat also als j-ten Eintrag die 1_R, und sonst überall 0_R. Natürlich gilt

$$\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} 1_R \\ 0_R \\ \vdots \\ 0_R \end{pmatrix} . r_1 + \dots + \begin{pmatrix} 0_R \\ \vdots \\ 0_R \\ 1_R \end{pmatrix} . r_n .$$

Analog schreiben wir ${}^nR$ für den freien Links-R-Modul ${}^{(I)}R$ und stellen die Elemente als Zeilen dar:

$$(r_1, \dots, r_n) = (r_i)_{i\in I} = (r_i)_{i=1,\dots,n} \quad \in \quad {}^nR = {}^{(I)}R .$$

Als Standardbasisvektoren erhalten wir entsprechend

$$\varepsilon_1 = (1, 0, \dots, 0)\,, \dots, \varepsilon_n = (0, \dots, 0, 1) .$$

In diesem Fall ist

$$(r_1, \dots, r_n) = r_1 \cdot (1, 0, \dots, 0) + \dots + r_n \cdot (0, \dots, 0, 1) .$$

2.32. Proposition. *Es sei M ein freier Rechts-R-Modul mit Basis B. Dann existiert zu jedem $m \in M$ genau eine Familie $(m_b)_{b\in B} \in R^B$ mit $m_b = 0_R$ für fast alle $b \in B$, so dass*

$$\sum_{b\in B} b \,.\, m_b = m \,. \tag{1}$$

Ein analoges Resultat gilt für freie Links-R-Moduln.

Beweis. Da B eine Basis ist, erzeugt B den Modul M. Also existiert eine Familie $(m_b)_{b\in B} \in R^B$ mit der Eigenschaft (1).

Sei jetzt $(n_b)_{b\in B} \in R^B$ eine weitere Familie mit $n_b = 0_R$ für fast alle $b \in B$, so dass
$$\sum_{b\in B} b \,.\, n_b = m \,.$$
Dann folgt
$$0_M = m - m = \sum_{b\in B} b \,.\, n_b - \sum_{b\in B} b \,.\, m_b = \sum_{b\in B} b \,.\, (n_b - m_b) \,,$$
und es gilt $n_b - m_b = 0_R$ für fast alle $b \in B$. Da B linear unabhängig ist, folgt $n_b - m_b = 0_R$ für alle $b \in B$. Also gilt $(m_b)_{b\in B} = (n_b)_{b\in B}$, das heißt, die Familie $(m_b)_{b\in B}$ ist auch eindeutig. □

Das bedeutet, dass wir mit Hilfe einer Basis ein beliebiges Element in einem freien Modul ersetzen können durch eine Ansammlung von Ringelementen. Das ist insbesondere zum Rechnen sehr hilfreich.

2.33. Definition. Es sei M ein freier Rechts-R-Modul mit Basis B, und es sei $m \in M$. Dann heißen die $(m_b)_{b\in B}$ in R aus Proposition 2.32 die *Koordinaten* von m bezüglich der Basis B. Die Abbildung $M \to R^{(B)}$ mit $m \mapsto (m_b)_{b\in B}$ heißt die *Koordinatenabbildung* zur Basis B. Umgekehrt ist die *Basisabbildung* von M zur Basis B die Abbildung $R^{(B)} \to M$ mit
$$(r_b)_{b\in B} \longmapsto \sum_{b\in B} b \,.\, r_b \,.$$

2.34. Bemerkung. Nach Proposition 2.32 ist die Basisabbildung bijektiv. Ihre Umkehrabbildung ist die Koordinatenabbildung.

2.3. Lineare Abbildungen

2.35. Definition. Sei $(R, +, \cdot)$ ein Ring und seien $(M, +, \cdot)$ und $(N, +, \cdot)$ Rechts-R-Moduln, dann heißt eine Abbildung $F\colon M \to N$ ein (*Rechts-R-*) *Modulhomomorphismus* oder (*rechts-*) *R-linear* (kurz: *linear*), falls für alle ℓ, $m \in M$ und alle $r \in R$ gilt

(L1) $\quad F(\ell + m) = F(\ell) + F(m) \qquad$ (*Additivität*),

(L2) $\quad F(m \,.\, r) = F(m) \,.\, r \qquad$ (*Homogenität*).

Falls R ein (Schief-) Körper ist, nennt man lineare Abbildungen zwischen (Rechts-R-) Vektorräumen auch *Vektorraumhomomorphismen.* Die Menge aller (rechts-) R-linearer Abbildungen von M nach N wird mit $\mathrm{Hom}_R(M,N)$ bezeichnet. Analog definieren wir *Links-R-Modulhomomorphismen.* Die Menge aller Links-R-Modulhomomorphismen von A nach B wird mit ${}_R\mathrm{Hom}(M,N)$ bezeichnet.

Für lineare Abbildungen gilt wegen Proposition 2.22 (1) insbesondere immer

$$F(0_M) = F(0_M \,.\, 0_R) = F(0_M) \,.\, 0_R = 0_N \;.$$

Wir bemerken, dass die Addition in (L1) einmal in M und einmal in N stattfindet. Genauso wird in (L2) einmal in M und einmal in N skalar multipliziert. Aus diesem Grund ist es wichtig, dass beide Moduln über demselben Ring R definiert sind. Wenn R kommutativ ist, gibt es nach Bemerkung 2.23 keinen Unterschied zwischen Links- und Rechts-R-Moduln. Wir sprechen dann nur noch von Modulhomomorphismen, und schreiben $\mathrm{Hom}(M,N)$ oder $\mathrm{Hom}_R(M,N)$ für die Menge aller linearer Abbildungen.

2.36. Beispiel. Wir kennen bereits Beispiele linearer Abbildungen.

(1) Wir haben bereits in den Abschnitten 1.5 und 1.6 benutzt (aber noch nicht bewiesen), dass Isometrien des $\mathbb{R}^2$ und des $\mathbb{R}^3$, die den Nullpunkt festhalten, $\mathbb{R}$-linear sind. Dazu gehören Drehungen um den Nullpunkt und Spiegelungen an Achsen durch den Nullpunkt im $\mathbb{R}^2$, siehe Bemerkung 1.65, sowie Drehungen um Achsen durch den Nullpunkt, die Punktspiegelung am Ursprung, sowie Spiegelungen an Ebenen durch den Nullpunkte im $\mathbb{R}^3$, siehe Bemerkung 1.75.

(2) Wir betrachten $M = N = \mathbb{C}$ zunächst als Modul über $\mathbb{C}$. Die komplexe Konjugation entspricht der Spiegelung an der reellen Achse. Wir überprüfen die Axiome (L1), (L2). Nach Bemerkung 1.60 gilt

$$\overline{z+w} = \overline{z} + \overline{w} \qquad \text{und} \qquad \overline{z \,.\, w} = \overline{z} \,.\, \overline{w} \;.$$

Also ist komplexe Konjugation additiv, aber nicht homogen, da im allgemeinen $w \neq \overline{w}$. Wenn wir aber $\mathbb{C}$ als $\mathbb{R}$-Modul auffassen, dann gilt auch (L2), da $w = \overline{w}$ genau dann, wenn $w \in \mathbb{R}$. Also kommt es auch bei Linearität auf den zugrundeliegenden Ring oder (Schief-) körper an.

(3) Sei $M = \mathbb{H}$ ein Rechts-$\mathbb{H}$-Vektorraum wie in Beispiel 2.21 (1), so dass $m \,.\, q = mq$ für $m \in M$, $q \in \mathbb{H}$. Es sei $f\colon M \to M$ rechts $\mathbb{H}$-linear und $p = f(1)$. dann folgt

$$f(m) = f(1 \,.\, m) = f(1) \,.\, m = pm \;,$$

also wird f durch Linksmultiplikation mit $p = f(1)$ gegeben. Umgekehrt ist Linksmultiplikation mit einem beliebigen Quaternion eine rechts-$\mathbb{H}$-lineare Abbildung.

(4) Sei $(M, +, \cdot\,)$ ein Links-$\mathbb{H}$-Vektorraum. Wir konstruieren daraus einen Rechts-$\mathbb{H}$-Vektorraum $(M, +, \,.\,)$ (und umgekehrt), so dass $m \,.\, q = \overline{q} \cdot m$

für alle $m \in M$ und $q \in \mathbb{H}$. Das Axiom (M1) ist erfüllt, denn wegen Satz 1.71 (6) gilt

$$m \,.\, (q \cdot r) = \overline{q \cdot r} \cdot m = \overline{r} \cdot \overline{q} \cdot m = (\overline{q} \cdot m) \,.\, r = (m \,.\, q) \,.\, r$$

für alle $m \in M$ und q, $r \in H$. Die anderen Modulaxiome lassen sich ebenso leicht nachprüfen.

Wir fassen jetzt $N = \mathbb{H}$ als Links-$\mathbb{H}$-Vektorraum auf und betrachten außerdem $M = \mathbb{H}$ wie in (3). Dann ist die Quaternionen-Konjugation $F = \overline{\cdot}\colon \overline{N} \to M$ rechts-$\mathbb{H}$-linear, denn

$$F(n \,.\, q) = F(\overline{q} \cdot n) = \overline{\overline{q} \cdot n} = \overline{n} \cdot \overline{\overline{q}} = \overline{n} \cdot q = F(n) \,.\, q \,.$$

2.37. Bemerkung. Auch in der Analysis spielen lineare Abbildungen eine wichtige Rolle. Beispielsweise dient die Ableitung einer Funktion $f\colon I \to \mathbb{R}$ auf einem offenen Intervall $I \subset \mathbb{R}$ dazu, die Funktion an einer Stelle $x_0 \in I$ zu beschreiben als

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + o(x - x_0) \,, \tag{1}$$

dabei ist der zweite Term linear in $x-x_0$, und der Rest $o(x-x_0)$ geht für $x \to x_0$ schneller gegen 0 als jede lineare Funktion in $x - x_0$ außer der konstanten Funktion 0. Viele wichtige Eigenschaften von f lassen sich bereits von der „Linearisierung“ $f(x_0) + f'(x_0) \cdot (x - x_0)$ ablesen: wenn $f'(x_0) \neq 0$ ist, ist x_0 keine lokale Extremstelle von f, und f besitzt sogar lokal eine differenzierbare Umkehrfunktion.

Eine Funktion $f\colon U \to \mathbb{R}^m$ auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ nähert man wieder wie in (1) an, dabei ist diesmal $f'(x_0)\colon \mathbb{R}^n \to \mathbb{R}^m$ selbst eine lineare Abbildung. Im Fall $m = 1$ folgt aus $f'(x_0) \neq 0$ wieder, dass x_0 keine lokale Extremstelle von f ist. Im Fall $m = n$ hat f genau dann eine differenzierbare lokale Umkehrfunktion, wenn $f'(x_0)$ als lineare Abbildung invertierbar ist. Ist $f'(x_0)$ injektiv, so ist das Bild der Einschränkung von f auf eine kleine Umgebung von x_0 eine „glatte“ Teilmenge des $\mathbb{R}^m$. Ist $f'(x_0)$ surjektiv, so ist das Urbild $f^{-1}\bigl(\{f(x_0)\}\bigr)$ nahe x_0 eine „glatte“ Teilmenge des $\mathbb{R}^n$.

Als Beispiel betrachten wir zwei Funktionen $f\colon \mathbb{R} \to \mathbb{R}^2$ und $F\colon \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(t) = \begin{pmatrix} t^3 \\ t^2 \end{pmatrix} \qquad \text{und} \qquad F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2 - y^3 \,.$$

Dann ist $f'(t) = \binom{3y^2}{2t}\colon \mathbb{R} \to \mathbb{R}^2$ injektiv außer an der Stelle $t = 0$, und $F'\bigl(\binom{x}{y}\bigr) = (2x, -3y^2)\colon \mathbb{R}^2 \to \mathbb{R}$ ist surjektiv außer an der Stelle $\binom{x}{y} = 0$. Die Teilmenge

$$\operatorname{im} f = F^{-1}(\{0\}) \subset \mathbb{R}^2$$

ist glatt außer an der Stelle $\binom{0}{0} = f(0)$, wo die Ableitungen verschwinden.

Auch aufgrund dieser späteren Anwendungen lohnt es sich, lineare Abbildungen und ihre Eigenschaften genauer zu studieren.

2.38. Beispiel. Es sei M, N Rechts-R-Moduln. Dann sind die folgenden Abbildungen immer R-linear.

(1) Die Identität aus Beispiel 1.18 (1) ist immer linear, denn

$$\begin{aligned} \mathrm{id}_M(\ell+m) &= \ell+m = \mathrm{id}_M(\ell)+\mathrm{id}_M(m) \;, \\ \text{und} \quad \mathrm{id}_M(m\,.\,r) &= m\,.\,r = \mathrm{id}_M(m)\,.\,r \;. \end{aligned}$$

(2) Die *Nullabbildung* $0\colon M \to N$ mit $0(m) = 0_N$ für alle $m \in M$ ist ebenfalls linear, denn

$$\begin{aligned} 0(\ell+m) &= 0_N = 0_N + 0_N = 0(\ell)+0(m) \;, \\ \text{und} \quad 0(m\,.\,r) &= 0_N = 0_N\,.\,r = 0(m)\,.\,r \;. \end{aligned}$$

2.39. Bemerkung. Es sei I eine Menge und N ein Rechts-R-Modul, dann wird auch $N^I = \mathrm{Abb}(I,N)$ zu einem Rechts-R-Modul mit den Rechenoperationen

$$(F+G)(i) = F(i)+G(i) \qquad \text{und} \qquad (F\,.\,r)(i) = F(i)\,.\,r \qquad \in N \;.$$

Auf diese Weise erhält man beispielsweise auch den Vektorraum $\mathbb{R}^{\mathbb{N}}$ der reellwertigen Folgen. Für die folgende Konstruktion ist wichtig, dass man Abbildungen mit Werten in einem Modul addieren kann, indem man die Bilder addiert.

2.40. Proposition. *Die Hintereinanderausführung von linearen Abbildungen ist linear. Die Umkehrabbildung einer bijektiven linearen Abbildung ist linear. Die Summe linearer Abbildungen ist linear.*

Beweis. Seien L, M und N Rechts-R-Moduln, und seien $F\colon M \to N$ und $G\colon L \to M$ R-linear. Dann folgt aus der Linearität von F und G für alle ℓ, $m \in L$ und alle $r \in R$, dass

$$\begin{aligned} (F\circ G)(\ell+m) &= F\big(G(\ell+m)\big) = F\big(G(\ell)+G(m)\big) \\ &= F\big(G(\ell)\big)+F\big(G(m)\big) = (F\circ G)(\ell)+(F\circ G)(m) \;, \\ \text{und} \quad (F\circ G)(\ell\,.\,r) &= F\big(G(\ell\,.\,r)\big) = F\big(G(\ell)\,.\,r\big) \\ &= F\big(G(\ell)\big)\,.\,r = (F\circ G)(\ell)\,.\,r \;. \end{aligned}$$

Also ist auch $F\circ G$ linear.

Sei jetzt $F\colon M \to N$ eine bijektive lineare Abbildung und $G\colon N \to M$ ihre Umkehrabbildung, siehe Satz 1.23. Es seien p, $q \in N$ beliebig und $\ell = G(p)$, $m = G(q) \in M$, so dass $F(\ell) = p$ und $G(m) = q$. Außerdem sei $r \in R$. Aus der Linearität von F folgt

$$\begin{aligned} G(p+q) &= G\big(F(\ell)+F(m)\big) = G\big(F(\ell+m)\big) = \ell+m = G(p)+G(q) \;, \\ \text{und} \quad G(q\,.\,r) &= G\big(F(m)\,.\,r\big) = G\big(F(m\,.\,r)\big) = m\,.\,r = G(q)\,.\,r \;. \end{aligned}$$

Also ist die Umkehrabbildung G linear.

Seien jetzt $F, G\colon M \to N$ linear, dann ist auch $F+G$ linear, denn für alle ℓ, $m \in M$ und alle $r \in R$ gilt

$$\begin{aligned} (F+G)(\ell+m) &= F(\ell+m)+G(\ell+m) = F(\ell)+F(m)+G(\ell)+G(m) \\ &= (F+G)(\ell)+(F+G)(m) \;, \\ (F+G)(m\,.\,r) &= F(m\,.\,r)+G(m\,.\,r) = F(m)\,.\,r+G(m)\,.\,r \\ &= (F+G)(m)\,.\,r \;. \end{aligned}$$

□

Achtung: im allgemeinen ist das Vielfache einer linearen Abbildung nicht linear. Betrachte dazu $F\colon M \to N$, $m \in M$ und r, s$\in R$.

$$(F\,.\,r)(m\,.\,s) = F(m\,.\,s)\,.\,r = F(m)\,.\,s\,.\,r\;,(F\,.\,r)(m)\,.\,s \quad = F(m)\,.\,r\,.\,s\;.$$

Diese beiden Ausdrücke sind im allgemeinen verschieden.

2.41. Definition. Es seien M, N Rechts-R-Moduln. Bijektive lineare Abbildungen $F\colon M \to N$ heißen (*Rechts-R-*) *Modulisomorphismen.* Lineare Abbildungen $F\colon M \to M$ heißen (*Rechts-R-*) *Modulendomorphismen,* und wenn sie bijektiv sind, (*Rechts-R-*) *Modulautomorphismen.* Falls R ein Körper ist, sprechen wir von *Vektorraumiso-, -endo- und -automorphismen.* Die Menge aller Modul- oder Vektorraumisomorphismen von M nach N wird mit $\mathrm{Iso}_R(M,N) \subset \mathrm{Hom}_R(M,N)$ bezeichnet, die Menge aller Modul- oder Vektorraumendo- oder -automorphismen von M mit $\mathrm{End}_R(M)$ beziehungsweise $\mathrm{Aut}_R M \subset \mathrm{End}_R M$. Analoge Bezeichnungen ${}_R\mathrm{Iso}(M,N)$, ${}_R\mathrm{End}\,M$ und ${}_R\mathrm{Aut}\,M$ führen wir für Links-R-Moduln oder -Vektorräume ein.

Bei Aut_R und End_R lässt man gelegentlich die Klammern weg, es ist also $\mathrm{End}_R M = \mathrm{End}_R(M)$. Analoge Bezeichnungen (Hom, End, Iso und Aut) werden in der Mathematik häufig für Abbildungen benutzt, die eine bestimmte „Struktur“ (hier die eines Moduls beziehungsweise Vektorraums) erhalten.

2.42. Folgerung (aus Prop2.40)**.** *Es sei R ein Ring, und M und N seien Rechts-R-Moduln.*

(1) *Die Automorphismen von M bilden eine Gruppe* $(\mathrm{Aut}_R M, \circ)$*, die* Automorphismengruppe *von M.*
(2) *Die Endomorphismen von M bilden einen Ring* $(\mathrm{End}_R M, +, \circ)$ *mit Eins* id_M*, den* Endomorphismenring *von M.*
(3) *Der Modul M ist ein Links-*$\mathrm{End}_R M$*-Modul, die skalare Multiplikation wirkt für alle $F \in \mathrm{End}_R M$ und alle $m \in M$ durch $F \cdot m = F(m) \in M$.*
(4) *Die Homomorphismen* $\mathrm{Hom}_R(M,N)$ *bilden einen unitären Recht-*$\mathrm{End}_R M$*-Modul, und einen unitären Links-*$\mathrm{End}_R N$*-Modul.*

Analoge Aussagen gelten, wenn M und N Links-R-Moduln sind.

Beweis. Der Beweis von (1) orientiert sich am Beispiel 2.5 der Automorphismengruppe einer Menge. Zunächst einmal ist die Verknüpfung zweier Automorphismen ein Automorphismus nach Proposition 2.40, genauso wie die Umkehrabbildung eines Automorphismus. Nach Beispiel 2.38 (1) ist auch die Identität ein Automorphismus. Die Gruppenaxiome ergeben sich wieder aus Bemerkung 2.4 (1)–(3).

Die Addition auf $\mathrm{End}_R(M)$ in (2) ist die gleiche wie in Bemerkung (2.39), Man üperprüft leicht die Axiome (G1)-(G4). Aus Bemerkung 2.4 (1) und (2) folgen (R1), (R3). Als nächstes seien F, G, $H \in \mathrm{End}_R(M)$, dann gilt

$$(*) \qquad (F+G)\circ H = F\circ H + G\circ H \quad \text{und} \quad F\circ(H+K) = F\circ H + F\circ K\;,$$

wie man durch Einsetzen von $m \in M$ leicht überprüft. Es folgt (R2) in (2). Also bildet $(\operatorname{End}_R M, +, \circ)$ einen Ring mit Eins $1_{\operatorname{End}_R M} = \mathrm{id}_M$.

Da M ein Rechts-R-Modul ist, ist $(M, +)$ eine abelsche Gruppe. Das Axiom (M1) für Linksmoduln folgt aus der Definition 1.19 der Verkettung, denn $(F \circ G)(m) = F(G(m))$ für alle F, $G \in \operatorname{End}_R(M)$ und alle $m \in M$. Axiom (M2) ist die Definition der Addition auf $\operatorname{End}_R(M)$ in Bemerkung (2.39), und (M3) ist gerade die Additivität (L1) der Endomorphismen. Schließlich ist M unitär, da die Eins in $\operatorname{End}_R M$ gerade id_M ist.

Es sei $F \in \operatorname{Hom}_R(M, N)$, $G \in \operatorname{End}_R M$ und $H \in \operatorname{End}_R M$. Dann folgt

$$F \circ G \in \operatorname{Hom}_R(M, N) \qquad \text{und} \qquad H \circ F \in \operatorname{Hom}_R(M, N)\ .$$

Also wirkt $\operatorname{End}_R M$ von rechts und $\operatorname{End}_R N$ von links auf $\operatorname{Hom}_R(M, N)$. Der Beweis von (4) funktioniert danach im wesentlichen genauso wie der von (2). Beispielsweise zeigt man die Distributivgesetze (M2), (M3) genau wie in (*), und (M1), (M4) folgen aus Bemerkung 2.4 (1) und (2). □

Wir betrachten $\operatorname{Hom}_R(M, R)$ als Spezialfall von (4), so dass $N = R$ als Rechts-R-Modul wie in Beispiel 2.21 (1). Dann ist $\operatorname{Hom}_R(M, R)$ ein Links-$\operatorname{End}_R R$-Modul. Für alle $r \in R$ ist Linksmultiplikation mit r rechts-R-linear, denn

$$r \cdot (s + t) = r \cdot s + r \cdot t \qquad \text{und} \qquad r \cdot (s \,.\, t) = (r \cdot s) \,.\, t$$

für alle s, $t \in R$. Also gilt $R \subset \operatorname{End}_R R$. Wenn R unitär ist, dann folgt wie in Beispiel (2.36) (3) sogar $\operatorname{End}_R R = R$. In jedem Fall können wir $\operatorname{Hom}_R(M, R)$ als Links-R-Modul auffassen.

2.43. Definition. Sei M ein Rechts-R-Modul, dann ist $M^* = \operatorname{Hom}_R(M, R)$ der zu M *duale* Links-R-Modul, beziehungsweise der zu M duale Links-R-Vektorraum, falls R ein (Schief-) Körper ist. Analog definieren wir den dualen Rechts R=Modul ${}^*N$ zu einem Links-R-Modul N.

Sie lernen einige duale Moduln in den Übungen kennen.

2.4. Unterräume und Quotienten

In diesem Abschnitt lernen wir, wie man aus gegebenen Moduln neue konstruieren kann.

2.44. Definition. Es sei $(M, +, .\,)$ ein Rechts-R-Modul und $U \subset M$ eine Teilmenge. Dann heißt U ein (*Rechts-R-*) *Untermodul,* falls für alle u, $v \in U$ und alle $r \in R$ die folgenden Untermodulaxiome gelten:

(U1) $\quad 0_M \in U$ $\qquad$ (*Neutrales Element*) ,

(U2) $\quad u + v \in U\ , \quad -u \in U$ $\qquad$ (*abgeschlossen unter Addition*) ,

(U3) $\quad u \,.\, r \in U$ $\qquad$ (*abgeschlossen unter skalarer Multiplikation*) .

Analog definieren wir Links-R-Untermoduln von Links-R-Moduln. Falls R ein (Schief-) Körper ist, sprechen wir stattdessen von (*Rechts-/Links-*) *Untervektorräumen*, kurz *Unterräumen.*

Anstelle von (U1) hätte es gereicht zu fordern, dass $U \neq \emptyset$. Denn sei $u \in U$, dann folgt $0_M = u \,.\, 0_R \in U$ aus (U3) und Proposition 2.22.

2.45. Beispiel. Wir kennen bereits Beispiele von Untervektorräumen.

(1) Wir fassen die Quaternionen $\mathbb{H}$ als $\mathbb{R}$-Vektorraum auf. In Abschnitt 1.6 haben wir die Unterräume $\mathbb{R} \subset \mathbb{H}$ der reellen und $\mathbb{R}^3 \subset \mathbb{H}$ der imaginären Quaternionen betrachtet.
(2) In der Analysis trifft man häufig auf Untervektorräume. Beispielsweise bilden die Nullfolgen einen Unterraum des Vektorraums aller Folgen. Für ein offenes Intervall I bilden die stetigen Funktionen auf I einen Unterraum des Raumes aller Funktionen auf I, und die differenzierbaren Funktionen einen Unterraum des Raumes der stetigen Funktionen auf I.

2.46. Bemerkung. Jeder Untermodul U eines Rechts-R-Moduls $(M, +, \,.\,)$ ist selbst ein Rechts-R-Modul. Zunächst einmal existiert ein Nullelement 0_M und die Verknüpfungen $+\colon U \times U \to U$ und $.\colon U \times R \to U$ sind wohldefiniert dank (U1)–(U3). Da die Axiome (G1)–(G4) und (M1)–(M3) gelten, wenn man für die Variablen Elemente aus M einsetzt, gelten sie erst recht, wenn man nur Elemente aus U zulässt. Beispielsweise gilt $0_M + u = u$ in M für alle $u \in U$, also auch in U.

Die Inklusion $U \to M$ aus Bemerkung 1.21 ist linear, da (L1) und (L2) offensichtlich gelten.

Wenn R ein Ring mit Eins und M ein unitärer Modul ist, dann ist auch jeder Untermodul U unitär mit der gleichen Begründung wie oben. In diesem Fall darf man auf die Forderung $-u \in U$ in (U2) verzichten, da $-u = u \,.\, (-1)$.

Auf völlig analoge Weise kann man *Untergruppen* und *Unterringe* definieren. Beispielsweise sollte ein Unterring $U \subset R$ das Element 0_R enthalten, und die Summe und das Produkt von Elementen von U sollte wieder in U liegen. Bei Körpern bevorzugt man aus naheliegenden Gründen den Begriff *Teilkörper.*

Wir wollen nun Quotientenmoduln in Analogie zu Beispiel 2.9 konstruieren. Dazu sei $(M, +, \,.\,)$ ein Rechts-R-Modul und $U \subset M$ ein Untermodul. Dann definieren wir eine Relation „$\sim$“ auf M für alle m, $n \in M$ durch

$$m \sim n \qquad \Longrightarrow \qquad n - m \in U\ .$$

Das ist eine Äquivalenzrelation, denn (Ä1)–(Ä3) folgen für ℓ, m, $n \in M$ aus

$$m - m \in U\ , \qquad n - m \in U \quad \Longrightarrow \quad m - n = -(n - m) \in U\ ,$$

$$\text{sowie} \quad m - \ell \in U \text{ und } n - m \in U \quad \Longrightarrow \quad n - \ell = (n - m) + (m - \ell) \in U\ .$$

2.47. Definition. Der Quotient $M/U = M/\!\sim$ heißt der *Quotientenmodul* von M nach U (lies „M modulo U“). Falls R ein Körper ist heißt M/U der *Quotientenvektorraum,* kurz *Quotientenraum.*

Man beachte hier, dass wir zur Definition der Äquivalenzrelation „$\sim$“ und der Menge M/U nur die additive Struktur des Moduls M benutzt haben. Die skalare Multiplikation können wir nachträglich definieren. Es sei $p\colon M \to M/{\sim}$ die Quotientenabbildung, siehe Definition 1.42.

2.48. Proposition. *Es sei $(M, +, .\,)$ ein Rechts-R-Modul und $U \subset M$ ein Untermodul. Dann induzieren „$+$“ und „$.$“ Verknüpfungen*

$$+\colon M/U \times M/U \to M/U \qquad \textit{und} \qquad .\,\colon M/U \times R \to M/U\ ,$$

und $(M/U, +, .\,)$ ist ein Rechts-R-Modul. Die Quotientenabbildung $p\colon M \to M/U$ ist rechts-R-linear. Wenn R ein Ring mit Eins und M ein unitärer Modul ist, ist auch M/U ein unitärer Modul.

Beweis. Wir gehen vor wie in Beispiel 2.9. Seien $m, n, p, q \in M$ mit $n-m \in U$ und $q - p \in U$, und sei $r \in R$, dann folgt

$$\begin{aligned} (n+q) - (m+p) &= (n-m) + (q-p) && \in U\ ,\\ (n\,.\,r) - (m\,.\,r) &= (n-m)\,.\,r && \in U\\ \text{und} \qquad (-n) - (-m) &= -(n-m) && \in U\ , \end{aligned}$$

also sind Addition und skalare Multiplikation auf M/U wohldefiniert durch

$$[m] + [p] = [m+p]\ , \qquad -[m] = [-m] \qquad \text{und} \qquad [m]\,.\,r = [m\,.\,r]\ .$$

Wir setzen $0_{M/U} = [0_M]$. Jetzt können wir die Axiome (G1)–(G4), (M1)–(M3) und gegebenenfalls (M4) auf die entsprechenden Axiome in M zurückführen. Beispielsweise gilt (M1), denn

$$([m]\,.\,r)\,.\,s = [m\,.\,r]\,.\,s = [(m\,.\,r)\,.\,s] = [m\,.\,(r \cdot s)] = [m]\,.\,(r \cdot s)\ .$$

Schließlich zur Linearität der Quotientenabbildung: für alle $m, n \in M$ und r, $s \in R$ gilt

$$p(m\,.\,r + n\,.\,s) = [m\,.\,r + n\,.\,s] = [m]\,.\,r + [n]\,.\,s = p(m)\,.\,r + p(n)\,.\,s\ . \qquad \square$$

2.49. Beispiel. Wir betrachten $M = \mathbb{Z}$ als $\mathbb{Z}$-Modul und

$$U = n\mathbb{Z} = \langle\, \{n\}\, \rangle = \{\, an \mid a \in \mathbb{Z}\,\} = \{\, \dots, -n, 0, n, \dots\,\}\ .$$

Dann ist U ein Untermodul, und der Quotient $M/U = \mathbb{Z}/n\mathbb{Z}$ ist gerade der Modul aus Beispiel 2.29.

2.50. Bemerkung. In Bemerkung 2.46 haben wir gesehen, dass geeignete Teilmengen von Gruppen, Ringen oder (Schief-) Körpern selbst wieder Gruppen, Ringe beziehungsweise Körper sind. Die Quotientenkonstruktion ist leider nicht so allgemein: Der Quotient einer Gruppe nach einer Untergruppe U beziehungsweise eines Ringes nach einem Unterring ist nur dann wieder Gruppe beziehungsweise Ring, wenn U gewisse zusätzliche Bedingungen erfüllt (siehe Übungen). Körper und Schiefkörper haben keine Quotienten.

2.51. Definition. Es seien M und N Rechts-R-Moduln, und es sei $F\colon M \to N$ rechts-R-linear. Dann definieren wir den *Kern* $\ker F$ durch

$$\ker F = F^{-1}(\{0_N\}) = \{\, m \in M \mid F(m) = 0\,\}\ .$$

Wir erinnern uns auch an das Bild $\operatorname{im} F$, siehe Definition 1.15.

2.52. Proposition. *Es seien M und N Rechts-R-Moduln, und $F\colon M \to N$ sei rechts-R-linear.*

(1) *Der Kern* $\ker F$ *ist ein Untermodul von M, und F ist genau dann injektiv, wenn* $\ker F = \{0_M\}$.
(2) *Das Bild* $\operatorname{im} F$ *ist ein Untermodul von N, und F ist genau dann surjektiv, wenn* $\operatorname{im} F = N$.

Die letzte Aussage in (2) ist klar nach Definition 1.17. Wir haben sie nur angefügt, um die Analogie zu (1) herzustellen.

Beweis. Die Untermodulaxiome folgen aus der Linearität von F, denn für alle m, $n \in M$ und alle $r \in R$ gilt

$$\begin{aligned} &F(0_M) = 0_N\ , \\ F(m) = F(n) = 0_N \quad &\Longrightarrow \quad F(m+n) = F(m) + F(n) = 0\ , \\ F(m) = 0_N \quad &\Longrightarrow \quad F(m\,.\,r) = F(m)\,.\,r = 0 \end{aligned}$$

Wenn F injektiv ist, hat insbesondere $\ker F = F^{-1}(\{0\})$ höchstens ein Element. Aus $F(0_M) = 0_N$ folgt dann $\ker F = \{0_M\}$.

Sei umgekehrt $\ker F = \{0_M\}$ und $F(m) = F(n) \in N$, dann folgt

$$F(m-n) = F(m) - F(n) = 0_N$$

aus der Additivität (L1) von F, somit ist $m - n \in \ker F$, also nach Voraussetzung $m - n = 0$, das heißt $m = n$. Also ist F injektiv, und (1) ist gezeigt.

Die Untermodulaxiome für $\operatorname{im} F \subset N$ folgen wieder aus der Linearität von F: für alle m, $n \in N$, p, $q \in N$ und $r \in R$ gilt

$$\begin{aligned} &0_N = F(0_M)\ , \\ p = F(m)\ , \quad q = F(n) \quad &\Longrightarrow \quad p + q = F(m+n)\ , \\ p = F(m) \quad &\Longrightarrow \quad p\,.\,r = F(p\,.\,r)\ . \end{aligned} \qquad \square$$

Der folgende Satz entspricht Proposition 1.43 (3).

2.53. Proposition (Universelle Eigenschaft des Quotienten)**.** *Es seien M und N Rechts-R-Moduln, es sei $U \subset M$ ein Untermodul mit Quotientenabbildung $p\colon M \to M/U$, und es sei $F\colon M \to N$ eine rechts-R-lineare Abbildung. Dann existiert genau dann eine Abbildung $\overline{F}\colon M/U \to N$ mit $F = \overline{F} \circ p$, wenn $U \subset \ker F$. In diesem Fall ist $\overline{F}$ eindeutig bestimmt und rechts-R-linear. Es gilt*

$$\operatorname{im} \overline{F} = \operatorname{im} F \qquad \text{und} \qquad \ker \overline{F} = \ker F / U\ .$$

Wenn $\overline{F}$ existiert, erhalten wir folgendes Diagramm:

$$\begin{tikzcd} M \arrow[r, "F"] \arrow[d, "p"'] & N \\ M/U\ . \arrow[ur, dashed, "\overline{F}"'] & \end{tikzcd}$$

Beweis. Zu „$\Longrightarrow$“ nehmen wir an, dass $\overline{F}$ existiert. Für alle $u \in U$ gilt $[u] = 0_{M/U}$, somit

$$F(u) = \overline{F}([u]) = \overline{F}(0_{M/U}) = F(0_M) = 0_N \;,$$

es folgt $U \subset \ker F$.

Zu „$\Longleftarrow$“ nehmen wir an, dass $U \subset \ker F$. Seien $m, n \in M$ mit $[m] = [n] \in M/U$, dann folgt

$$m - n \in U \subset \ker F \quad\Longrightarrow\quad F(m) - F(n) = F(m-n) = 0_N \;,$$

also gilt $F(m) = F(n)$, und $\overline{F}([m]) = F(m)$ ist wohldefiniert.

Die Eindeutigkeit von $\overline{F}$ folgt aus Proposition 1.43 (3). Außerdem ist $\overline{F}$ linear, denn

$$\begin{aligned} \overline{F}([m] + [n]) = F(m+n) = F(m) + F(n) = \overline{F}([m]) + \overline{F}([n]) \;, \\ \overline{F}([m] \,.\, r) = F(m \,.\, r) = F(m) \,.\, r = \overline{F}([m]) \,.\, r \end{aligned}$$

für alle $m, n \in M$ und alle $r \in R$.

Wir sehen leicht, dass $\operatorname{im} \overline{F} = \operatorname{im} F$. Es gilt $[m] \in \ker \overline{F} \subset M/U$ genau dann, wenn $m \in \ker F$, somit folgt

$$\ker \overline{F} = \ker F / U \;. \qquad \square$$

2.54. Folgerung (Homomorphiesatz). *Es seien M und N Rechts-R-Moduln und $F\colon M \to N$ linear. Dann induziert F einen Isomorphismus*

$$\overline{F}\colon M/\ker F \to \operatorname{im} F \;.$$

Beweis. Wir wenden Proposition 2.52 an mit $U = \ker F$. Da $\operatorname{im} \overline{F} = \operatorname{im} F$ gilt, dürfen wir $\overline{F}$ als Abbildung mit Bildbereich $\operatorname{im} F$ auffassen. Dann ist $\overline{F}$ linear. Da $\ker \overline{F} = \ker F / \ker F = \{[0_M]\}$, ist $\overline{F}$ injektiv nach Proposition 2.52 (1). Außerdem ist $\overline{F}$ surjektiv, da $\operatorname{im} \overline{F} = \operatorname{im} F$. Also ist $\overline{F}$ ein Isomorphismus. $\square$

Wir können also jede lineare Abbildung $F\colon M \to N$ wie folgt zerlegen:

$$\begin{tikzcd} M \arrow[rrr, "F"] \arrow[dr, "p"'] & & & N \\ & M/\ker F \arrow[r, "\overline{F}", "\cong"'] & \operatorname{im} F \,. \arrow[ur, "\iota"'] & \end{tikzcd}$$

Dabei ist p die Quotientenabbildung und ι die Inklusion. Um F zu verstehen, bieten sich die folgenden Schritte an.

(1) Bestimme $\ker F$ als Untermodul von M.
(2) Bestimme $\operatorname{im} F$ als Untermodul von N.
(3) Bestimme den Isomorphismus $\overline{F}\colon M/\ker F \to \operatorname{im} F$.

2.55. Beispiel. Wir betrachten eine Ebene $V \subset \mathbb{R}^3$ und eine Gerade $U \subset \mathbb{R}^3$, so dass sich U und V nur in einem Punkt schneiden. Wir wollen annehmen, dass das der Nullpunkt ist; dann sind U und V Unterräume. Unsere Anschauung sagt uns, dass es durch jeden Punkt $x \in \mathbb{R}^3$ genau eine zu V parallele Gerade gibt, und dass diese Gerade die Ebene U genau in einem Punkt schneidet. Wir definieren $F\colon \mathbb{R}^3 \to U$ so, dass $F(x)$ gerade dieser Schnittpunkt ist. Diese Abbildung ist $\mathbb{R}$-linear — all das wird im nächsten Kapitel klarer werden.

Nach Konstruktion werden genau die Punkte auf der Geraden V auf den Schnittpunkt 0 von U und V abgebildet, also ist $\ker F = V$. Jeder Punkt in der Ebene U wird auf sich abgebildet, also ist F insbesondere surjektiv. Aus dem Homomorphiesatz 2.54 folgt

$$\mathbb{R}^3/V = \mathbb{R}^3/\ker F \cong \operatorname{im} F = U\ .$$

Das Besondere hier ist, dass U selbst ein Unterraum von $\mathbb{R}^3$ ist mit $F|_U = \mathrm{id}_U$.

Sei jetzt wieder $x \in \mathbb{R}^3$ beliebig. Nach Konstruktion ist $x - F(x) \in V$, da eine zu V parallele Gerade durch x und $F(x)$ geht. Es folgt

$$x = u + v \qquad \text{mit} \qquad u = F(x) \in U \qquad \text{und} \qquad v = x - F(x) \in V\ .$$

Diese Zerlegung ist eindeutig, denn wäre $x = u' + v'$ eine weitere Zerlegung, dann würde folgen

$$u' + v' = u + v \qquad \longrightarrow \qquad u' - u = v - v' \in U \cap V = \{0\}\ ,$$

also $u = u'$ und $v = v'$. Somit liefern die Unterräume U und V ein Beispiel für die folgende Definition.

2.56. Definition. Es sei M ein Rechts-R-Modul und $U, V \subset M$ Untermoduln. Die *Summe* von U und V ist gegeben durch

$$U + V = \bigl\{\, u + v \bigm| u \in U, v \in V \,\bigr\} \subset M\ .$$

Falls $U \cap V = \{0\}$ heißt die Summe *direkt,* und wir schreiben statt $U + V$ auch $U \oplus V$. Falls M die direkte Summe $U \oplus V$ ist, sagen wir, dass V ein *Komplement* von U in M ist (und umgekehrt), oder, dass U und V *komplementäre Untermoduln* sind. Wenn R ein (Schief-) Körper ist, sprechen wir analog von *komplementären Unterräumen.*

Man beachte, dass wegen (U1) stets $0_M \in U \cap V$ gilt. Einen kleineren Durchschnitt als $\{0_M\}$ können zwei Untermoduln also nicht haben.

2.57. Beispiel. Wir geben Beispiele von direkten Summen und komplentären Untermoduln an.

(1) In den Übungen zeigen Sie, dass $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$. Also sind die Untermoduln

$$\begin{aligned} U &= 2\mathbb{Z}/6\mathbb{Z} = \{[0],[2],[4]\} \cong \mathbb{Z}/3\mathbb{Z} \\ \text{und} \qquad V &= 3\mathbb{Z}/6\mathbb{Z} = \{[0],[3]\} \cong \mathbb{Z}/2\mathbb{Z} \end{aligned}$$

von $M = \mathbb{Z}/6\mathbb{Z}$ zueinander komplementär.

(2) Ähnlich wie in (1) betrachte
$$V = 2\mathbb{Z}/4\mathbb{Z} = \{[0],[2]\} \subset M = \mathbb{Z}/4\mathbb{Z}\ .$$
Dann ist $V \cong \mathbb{Z}/2\mathbb{Z}$. Es gibt keinen komplementären Untermodul U, denn dieser müsste mindestens ein Element aus $M \setminus V$ enthalten, also entweder $[1]$ oder $[3]$. In beiden Fällen wäre $[2] \in U$, denn $[2] = [1] + [1] = [3] + [3]$, und somit $U \cap V \neq \{[0]\}$. Also existiert nicht immer ein komplementärer Untermodul.

Es sei $V \subset M$ ein Untermodul. Wir erinnern uns an die Quotientenabbildung $p\colon M \to M/V$ aus Proposition 2.48.

2.58. Proposition. *Es seien U, V Untermoduln eines Rechts-R-Moduls M.*

(1) *Die Summe $U + V \subset M$ ist ein Untermodul.*
(2) *Wenn die Summe direkt ist, existiert eine bijektive Abbildung*
$$U \times V \to U \oplus V \qquad \text{mit} \qquad (u,v) \mapsto u + v\ .$$
(3) *Es sei $p\colon M \to M/V$ die Quotientenabbildung. Wenn U und V komplementäre Untermoduln sind, dann ist $p|_U\colon U \to M/V$ ein Modulisomorphismus.*

Beweis. Die Unterraumaxiome für $U + V$ gelten, da
$$\begin{aligned} 0_M &= 0_M + 0_M &&\in U + V\ ,\\ (t+v) + (u+w) &= (t+u) + (v+w) &&\in U + V\\ \text{und}\qquad (u+v)\,.\,r &= u\,.\,r + v\,.\,r &&\in U + V \end{aligned}$$
für alle t, $u \in U$, v, $w \in V$ und $r \in R$.

Die Abbildung in (2) ist immer surjektiv nach Definition der Summe. Wenn die Summe direkt ist, ist für jedes Element $s \in U \oplus V$ die Zerlegung $s = u + v$ mit $u \in U$ und $v \in V$ eindeutig, denn aus $s = u' + v'$ mit $u' \in U$, $v' \in V$ folgt
$$u' - u = v - v' \in U \cap V \quad \Longrightarrow \quad u' - u = v - v' = 0_M\ .$$
Also ist die Abbildung in (2) auch injektiv.

Die Quotientenabbildung $p\colon M \to M/V$ ist linear nach Proposition 2.48. Die Inklusion $\iota\colon U \to M$ ist linear nach Bemerkung 2.46. Also ist auch die Abbildung $p|_U = p \circ \iota$ in (3) linear nach Proposition 2.40.

Aus $p(u) = p(u') \in M/V$ folgt, dass ein $v \in V$ existiert mit $u' = u + v$. Wie in (2) folgt aus $v = u - u' \in U \cap V$, dass $u = u'$. Also ist $p|_U$ injektiv.

Sei schließlich $[m] \in M/V$ mit $m \in M$, dann existieren $u \in U$, $v \in V$ mit $m = u + v$, da $M = U \oplus V$. Da $p(u) = [u] = [m]$, ist $p|_U$ auch surjektiv. □

2.59. Bemerkung. Wir können also den Quotientenmodul M/V mit Hilfe von $p|_U$ mit einem komplementären Untermodul U identifizieren, falls ein solcher existiert. Wenn U ein zu V komplementärer Untermodul ist, gibt es meistens noch andere komplementäre Untermoduln, siehe etwa Beispiel 2.55, wo in

Richtung von V auf verschiedene Ebenen in $\mathbb{R}^3$ projizieren kann. Das bedeutet, dass die Identifikation $M/V \cong U$ von der Wahl des Komplements abhängt. Obwohl man oft leichter mit dem komplementären Untermodul U als mit dem Quotienten M/V arbeiten kann, ist es daher manchmal sinnvoll, den Quotienten M/V zu betrachten.

Die direkte Summe erfüllt gleich zwei „universelle Eigenschaften". Sei dazu $M = U \oplus V$, dann betrachten wir die Inklusionsabbildungen $\iota_U \colon U \to M$ und $\iota_V \colon V \to M$. Wenn wir wie oben $M/U \cong V$ und $M/V \cong U$ identifizieren, erhalten wir auch Projektionen $p_U \colon M \to U$ und $p_V \colon M \to V$, so dass insbesondere

$$m = p_U(m) + p_V(m)$$

für alle $m \in M$.

2.60. Proposition (Universelle Eigenschaften der direkten Summe)**.** *Es sei M ein Rechts-R-Moduln und U, $V \subset M$ Untermoduln, so dass $M = U \oplus V$.*

(1) *Die Inklusions- und Projektionsabbildungen erfüllen*

$$p_U \circ \iota_U = \mathrm{id}_U \;, \qquad p_U \circ \iota_V = 0 \colon V \to U \;,$$
$$p_V \circ \iota_U = 0 \colon U \to V \qquad \textit{und} \qquad p_V \circ \iota_V = \mathrm{id}_V \;.$$

(2) Universelle Eigenschaft des Koproduktes: *Sei N ein weiterer Rechts-R-Modul und seien $F \colon U \to N$ und $G \colon V \to N$ linear, dann existiert genau eine lineare Abbildung $H \colon U \oplus V \to N$, so dass $F = H \circ \iota_U$ und $G = H \circ \iota_V$.*

(3) Universelle Eigenschaft des Produktes: *Sei N ein weiterer Rechts-R-Modul und seien $P \colon N \to U$ und $Q \colon N \to V$ linear, dann existiert genau eine lineare Abbildung $L \colon N \to U \oplus V$, so dass $P = p_U \circ L$ und $Q = p_V \circ L$.*

Wie eng diese beiden Eigenschaften miteinander verwandt sind, zeigen die folgenden Diagramme, die sich nur in der Richtung der Pfeile unterscheiden. Man sagt auch, die Diagramme sind zueinander *dual*.

$$\begin{tikzcd}
& N & \\
& U \oplus V \arrow[u, dashed, "H"'] & \\
U \arrow[uur, "F"] \arrow[ur, "\iota_U"'] & & V \arrow[uul, "G"'] \arrow[ul, "\iota_V"]
\end{tikzcd} \;, \qquad
\begin{tikzcd}
& N \arrow[d, dashed, "L"] \arrow[ddl, "P"'] \arrow[ddr, "Q"] & \\
& U \oplus V \arrow[dl, "p_U"] \arrow[dr, "p_V"'] & \\
U & & V
\end{tikzcd} \;.$$

Beweis. Zu (1) sei $u \in U$, dann folgt

$$(p_U \circ \iota_U)(u) = p_U(u + 0_M) = u = \mathrm{id}_U(u) \;,$$
$$(p_V \circ \iota_U)(u) = p_V(u + 0_M) = 0 = 0(u) \in V \;.$$

Also gilt $p_U \circ \iota_U = \mathrm{id}_U$ und $p_V \circ \iota_U = 0$. Die beiden anderen Gleichungen folgen genauso.

Zu (2) zeigen wir zunächst die Eindeutigkeit. Sei also eine lineare Abbildung H gegeben mit $H \circ \iota_U = F$ und $H \circ \iota_V = G$. Für $m = u + v$ folgt

$$\begin{aligned} H(m) = H(u) + H(v) &= H(\iota_U(u)) + H(\iota_V(v)) \\ &= F(u) + G(v) = F(p_U(m)) + G(p_V(m)) \;, \end{aligned}$$

also ist H eindeutig bestimmt.

Auf der anderen Seite ist die Abbildung $F \circ p_U + G \circ p_V$ linear nach Proposition 2.40. Sie leistet das Gewünschte, denn wegen (1) gilt

$$\begin{aligned} \bigl(F \circ p_U + G \circ p_V\bigr) \circ \iota_U &= F \circ \underbrace{p_U \circ \iota_U}_{=\mathrm{id}_U} + G \circ \underbrace{p_V \circ \iota_U}_{=0} = F \;, \\ \bigl(F \circ p_U + G \circ p_V\bigr) \circ \iota_V &= F \circ p_U \circ \iota_V + G \circ p_V \circ \iota_V = G \;. \end{aligned}$$

Der Beweis zu (3) verläuft analog. Es sei $n \in N$ und $m = L(n) = u + v$ mit $u \in U$ und $v \in V$, dann folgt $u = p_U(L(n)) = P(n)$ und $v = p_V(L(n)) = Q(n)$, also ist L eindeutig bestimmt.

Umgekehrt ist die Abbildung

$$\iota_U \circ P + \iota_V \circ Q \colon N \to M = U \oplus V$$

linear nach Proposition 2.40. Mithilfe von (1) überprüft man wieder, dass

$$p_U \circ \bigl(\iota_U \circ P + \iota_V \circ Q\bigr) = P \qquad \text{und} \qquad p_V \circ \bigl(\iota_U \circ P + \iota_V \circ Q\bigr) = Q \;. \qquad \square$$

Wir können Summen auch für mehr als zwei Unterräume definieren. Sei etwa M ein Rechts-R-Modul, sei I eine Indexmenge, und sei $(U_i)_{i \in I}$ eine Familie von Untermoduln, aufgefasst als Familie in der Potenzmenge von M, siehe Definition 1.11. Dann definieren wir ihre Summe als

$$\sum_{i \in I} U_i = \left\{ \sum_{i \in I} u_i \;\middle|\; u_i \in U_i \text{ für alle } i \in I,\ u_i = 0_M \text{ für fast alle } i \in I \right\} \subset M \;.$$

Wenn $U_i \cap \sum_{j \in I, j \neq i} U_j = \{0_M\}$ für alle $i \in I$, nennen wir diese Summe wieder direkt und schreiben

$$\bigoplus_{i \in I} U_i = \sum_{i \in I} U_i \;.$$

Manchmal definiert man auch eine direkte Summe von beliebigen Rechts-R-Moduln, die nicht Untermoduln eines festen Moduls M sind.

2.61. Definition. Es seien M_i Rechts-R-Modul. Dann definieren wir ihre *direkte Summe* und ihr *direktes Produkt* als

$$\begin{aligned} \coprod_{i \in I} M_i &= \left\{ (m_i)_{i \in I} \;\middle|\; m_i \in M_i \text{ für alle } i \in I,\ u_i = 0_{M_i} \text{ für fast alle } i \in I \right\} , \\ \prod_{i \in I} M_i &= \bigl\{ (m_i)_{i \in I} \bigm| m_i \in M_i \text{ für alle } i \in I \bigr\} \;. \end{aligned}$$

Wir erhalten für alle $j \in I$ Inklusionen $\iota_j\colon M_j \to \coprod_{i\in I} M_i$ beziehungsweise $\iota_j\colon M_j \to \prod_{i\in I} M_i$ mit

$$\iota_j(m) = (m_i)_{i\in I}\ , \qquad \text{mit} \qquad m_i = \begin{cases} m & \text{falls } i = j, \text{ und} \\ 0 & \text{falls } i \neq j, \end{cases}$$

und Projektionen $p_j\colon \coprod_{i\in I} M_i \to M_j$ beziehungsweise $p_j\colon \prod_{i\in I} M_i \to M_j$ mit

$$p_j\bigl((m_i)_{i\in I}\bigr) = m_j \in M_j\ .$$

Man überzeugt sich leicht, dass beides wieder Moduln sind. Dabei geht man ähnlich vor wie in Beispiel 2.30. In der Tat gilt

$$R^{(I)} = \coprod_{i\in I} R \qquad \text{und} \qquad R^I = \prod_{i\in I} R\ .$$

Die folgenden universellen Eigenschaften werden analog zu Proposition 2.60 bewiesen:

2.62. Proposition. *Es sei M_i ein Rechts-R-Modul für alle $i \in I$.*

(1) *Für die Inklusions- und Projektionsabbildungen gilt*

$$p_i \circ \iota_i = \mathrm{id}_{M_i} \qquad \textit{und} \qquad p_i \circ \iota_j = 0\colon M_j \to M_i$$

für alle i, $j \in I$ mit $i \neq j$.

(2) Universelle Eigenschaft des Koproduktes: *Sei N ein weiterer Rechts-R-Modul und sei $F_j\colon M_j \to N$ linear für alle $j \in I$, dann existiert genau eine lineare Abbildung $H\colon \coprod_{i\in I} M_i \to N$, so dass $F_j = H \circ \iota_j$ für alle $j \in I$.*

(3) Universelle Eigenschaft des Produktes: *Sei N ein weiterer Rechts-R-Modul und seien $P_j\colon N \to M_j$ linear für alle $j \in I$, dann existiert genau eine lineare Abbildung $L\colon N \to \prod_{i\in I} M_i$, so dass $P_j = p_j \circ L$ für alle $j \in I$.*

2.5. Matrizen

Wir wollen jetzt verstehen, wie man lineare Abbildungen durch Matrizen beschreiben kann. Das ist zum Beispiel dann wichtig, wenn man numerische Berechnungen durchführen will (also Berechnungen mit „echten“ Zahlen, nicht abstrakte Überlegungen mit Variablen). Auf der anderen Seite sollten wir Matrizen nur als nützliche Rechenhilfen verstehen. Im Vordergrund des Interesses werden weiterhin lineare Abbildungen stehen.

Vorab überlegen wir uns, dass $(R^{(I)})^J$ gerade die Menge derjenigen Familien $(a_{ij})_{i\in I, j\in J}$ in R bezeichnet, so dass für jedes $j \in J$ nur endlich viele $i \in I$ mit $a_{ij} \neq 0_R$ existieren.

2.63. Proposition. *Es sei R ein Ring mit Eins, I, J seien Mengen und $R^{(I)}$, $R^{(J)}$ die von ihnen erzeugten freien Rechts-R-Moduln. Für alle $j \in J$ bezeichne e_j den Basisvektor aus Beispiel 2.30* (1)*. Dann existiert zu jeder linearen Abbildung $A\colon R^{(J)} \to R^{(I)}$ eine Familie $(a_{ij})_{i\in I, j\in J} \in (R^{(I)})^J$, so dass*

$$(1) \qquad (a_{ij})_{i\in I} = A(e_j) \in R^{(I)} \qquad \text{für alle } j \in J\ .$$

Für alle $(b_{ij})_{i\in I, j\in J} \in (R^{(I)})^J$ existiert eine lineare Abbildung $B\colon R^{(J)} \to R^{(I)}$, so dass

$$(2) \qquad B\big((r_j)_{j\in J}\big) = \Bigg(\sum_{j\in J} b_{ij} \cdot r_j\Bigg)_{i\in I}\ .$$

Die in (1) *und* (2) *konstruierten Abbildungen $\Phi\colon \mathrm{Hom}_R\big(R^{(J)}, R^{(I)}\big) \to (R^{(I)})^J$ und $\Psi\colon (R^{(I)})^J \to \mathrm{Hom}_R\big(R^{(J)}, R^{(I)}\big)$ sind zueinander invers und daher bijektiv.*

Beweis. Da $A(e_j) \in R^{(I)}$, gibt es für jedes $j \in J$ nur endlich viele $i \in I$ mit $a_{ij} \neq 0_R$, und es folgt (1). Wir erhalten also eine Abbildung $\Phi\colon \mathrm{Hom}_R(R^{(J)}, R^{(I)}) \to (R^{(I)})^J$ mit $A \mapsto (a_{ij})_{i\in I, j\in J}$.

Es seien jetzt $(b_{ij})_{i\in I, j\in J} \in (R^{(I)})^J$ und $(r_j)_{j\in J} \in R^{(J)}$ gegeben. Dann ist

$$\Bigg(\sum_{j\in J} b_{ij} \cdot r_j\Bigg)_{i\in I} = \sum_{j\in J} (b_{ij} \cdot r_j)_{i\in I} = \sum_{j\in J} (b_{ij})_{i\in I} \,.\, r_j$$

eine Linearkombination aus Elementen $(b_{ij})_{i\in I} \in R^{(I)}$, also erhalten wir in (2) eine Abbildung $B\colon R^{(J)} \to R^{(I)}$.

Für alle $(r_j)_{j\in J}$, $(s_j)_{j\in J} \in R^{(J)}$ und alle $t \in R$ folgt

$$\begin{aligned}
B\big((r_j)_{j\in J} + (s_j)_{j\in J}\big) = B\big((r_j + s_j)_{j\in J}\big) &= \Bigg(\sum_{j\in J} b_{ij} \cdot (r_j + s_j)\Bigg)_{i\in I} \\
&= \Bigg(\sum_{j\in J} b_{ij} \cdot r_j + \sum_{j\in J} b_{ij} \cdot s_j\Bigg)_{i\in I} \\
&= \Bigg(\sum_{j\in J} b_{ij} \cdot r_j\Bigg)_{i\in I} + \Bigg(\sum_{j\in J} b_{ij} \cdot s_j\Bigg)_{i\in I} \\
&= B\big((r_j)_{j\in J}\big) + B\big((s_j)_{j\in J}\big)\ ,
\end{aligned}$$

und

$$\begin{aligned}
B\big((r_j)_{j\in J} \,.\, t\big) = B\big((r_j \cdot t)_{j\in J}\big) &= \Bigg(\sum_{j\in J} b_{ij} \cdot (r_j \cdot t)\Bigg)_{i\in I} \\
&= \Bigg(\sum_{j\in J} (b_{ij} \cdot r_j) \cdot t\Bigg)_{i\in I} = \Bigg(\sum_{j\in J} b_{ij} \cdot r_j\Bigg)_{i\in I} \,.\, t \\
&= B\big((r_j)_{j\in J}\big) \,.\, t\ .
\end{aligned}$$

Also ist die Abbildung B in (2) auch rechts-R-linear. Wir erhalten also eine Abbildung $\Psi\colon (R^{(I)})^J \to \mathrm{Hom}_R\big(R^{(J)}, R^{(I)}\big)$ mit $(b_{ij})_{i\in I, j\in J} \mapsto B$.

Sei wieder $(b_{ij})_{i\in I,j\in J} \in (R^{(I)})^J$ und $B = \Psi\bigl((b_{ij})_{i\in I,j\in J}\bigr)$. Für die Familie $(a_{ij})_{i\in I,j\in J} = \Phi(B)$ folgt

$$(a_{ij})_{i\in I} = B(e_j) = \Biggl(\sum_{k\in J} b_{ik}\cdot\delta_{kj}\Biggr)_{i\in I} = (b_{ij})_{i\in I}$$

für alle $j \in J$, also gilt $\Phi\circ\Psi = \mathrm{id}_{(R^{(I)})^J}$.

Sei umgekehrt $A \in \mathrm{Hom}_R\bigl(R^{(J)}, R^{(I)}\bigr)$ und $(a_{ij})_{i\in I,j\in J} = \Phi(A)$. Es sei $B = \Psi\bigl(\bigr)$, dann folgt aus der Linearität von A und aus Beispiel (2.30) (2), dass

$$\begin{aligned} B\bigl((r_j)_{j\in J}\bigr) &= \Biggl(\sum_{j\in J} a_{ij}\cdot r_j\Biggr)_{i\in I} = \sum_{j\in J}(a_{ij})_{i\in I}\,.\,r_j \\ &= \sum_{j\in J} A(e_j)\,.\,r_j = A\Biggl(\sum_{j\in J} e_j\,.\,r_j\Biggr) = A\bigl((r_j)_{j\in J}\bigr)\ . \end{aligned}$$

Also gilt auch $\Psi\circ\Phi = \mathrm{id}_{\mathrm{Hom}_R\left(R^{(J)},R^{(I)}\right)}$, somit sind Φ und Ψ zueinander invers und insbesondere bijektiv nach Satz 1.22 (4), (5). □

Diese Proposition ist die Grundlage für das Rechnen mit Matrizen, wie wir jetzt sehen werden.

2.64. Definition. Es sei R ein Ring und $m, n \in \mathbb{N}$. Eine $m\times n$-*Matrix* über R ist eine Familie $A = (a_{ij})_{i=1\dots m,j=1\dots n}$ in R, geschrieben

(1) $$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\ .$$

Die Menge aller $m\times n$-Matrizen über R wird mit $M_{m,n}(R)$ bezeichnet.

Wir definieren die *Matrixaddition* $+\colon M_{m,n}(R)\times M_{m,n}(R) \to M_{m,n}(R)$ durch

(2) $$A + B = (a_{ij} + b_{ij})_{i=1\dots m,j=1\dots n} \in M_{m,n}(R)$$

für alle $B = (b_{ij})_{i=1\dots m,j=1\dots n} \in M_{m,n}(R)$, und die *Matrizenmultiplikation* $\cdot : M_{\ell,m}(R)\times M_{m,n}(R) \to M_{\ell,n}(R)$ mit $\ell\in\mathbb{N}$ durch

(3) $$C\cdot A = \Biggl(\sum_{j=1}^{m} c_{ij}\cdot a_{jk}\Biggr)_{i=1\dots\ell,k=1\dots n} \in M_{\ell,n}(R)$$

für alle $C = (c_{ij})_{i=1\dots\ell,j=1\dots m} \in M_{\ell,m}(R)$.

Wenn die Größe einer Matrix bekannt ist, schreiben wir auch kurz $(a_{ij})_{ij} \in M_{m,n}(R)$ — daraus ergibt sich, dass $1 \le i \le m$ und $1 \le j \le n$.

Die Matrixaddition erfolgt komponentenweise, genau wie in Beispiel (2.30). Zwei Matrizen kann man nur addieren, wenn sie die gleiche Anzahl von Zeilen und die gleiche Anzahl von Spalten haben.

Zwei Matrizen lassen sich multiplizieren, wenn die erste soviele Spalten hat wie die zweite Zeilen. Die Matrixmultiplikation lässt sich am besten am folgenden Schema verdeutlichen:

$$\begin{array}{cc} & \begin{pmatrix} \cdot & \cdots & b_{1k} & \cdots & \cdot \\ \vdots & & \vdots & & \vdots \\ \cdot & \cdots & b_{nk} & \cdots & \cdot \end{pmatrix} \\ \begin{pmatrix} \cdot & \cdots & \cdot \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ \cdot & \cdots & \cdot \end{pmatrix} & \begin{pmatrix} \cdot & \cdots & \vert & \cdots & \cdot \\ \vdots & & \vert & & \\ \text{—} & & a_{i1}b_{1k}+\dots+a_{in}b_{nk} & & \vdots \\ \vdots & & & & \\ \cdot & & \cdots & & \cdot \end{pmatrix} \end{array}$$

Hierbei steht die Matrix A links, die Matrix B oben, und das Produkt $A \cdot B$ unten rechts. Der Eintrag an der Stelle (i,k) sieht also genauso aus wie das „Skalarprodukt“ aus der Zeile i der Matrix A und der Spalte k der Matrix B, vergleiche Definition 1.51 (1).

2.65. Bemerkung. Wir betrachten die folgenden Spezialfälle.

(1) Wenn $m = 0$ oder $n = 0$ ist, enthält $M_{m,n}(R)$ nur ein Element, die leere Matrix $(\,)$.
(2) Für $m = 1 = n$ identifizieren wir $M_{1,1}(R)$ mit R. Addition und Multiplikation von 1×1-Matrizen entsprechen genau der Addition und Multiplikation in R:
$$(r) + (s) = (r + s) \qquad \text{und} \qquad (r) \cdot (s) = (r \cdot s)\ .$$
(3) Es sei $n = 1$, dann ist $M_{m,1}(R) = R^m$ der „Raum der Spalten“ der Länge m, und Addition funktioniert genau wie in Beispiel 2.31. Wir können von rechts mit einer 1×1-Matrix aus (2) multiplizieren und erhalten die skalare Multiplikation
$$\begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} \cdot (s) = \begin{pmatrix} r_1 \cdot s \\ \vdots \\ r_m \cdot s \end{pmatrix} = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} . s\ .$$
Aus diesem Grund ist es sinnvoll, Spalten von rechts mit Skalaren zu multiplizieren.
(4) Für $m = 1$ ist $M_{1,n}(R) = {}^nR$ der „Raum der Zeilen“ der Länge n. Addition funktioniert wieder wie in Beispiel 2.31, Multiplikation mit einer 1×1-Matrix von links entspricht der Multiplikation mit einem Skalar.

Die nächsten zwei Spezialfälle sind so wichtig, dass wir sie separat formulieren.

2.66. Folgerung (aus Proposition 2.63)**.** *Es sei R ein Ring mit Eins und m, $n \in \mathbb{N}$. Dann existiert eine natürliche Bijektion*

$$\Phi\colon \operatorname{Hom}_R(R^n, R^m) \to M_{m,n}(R)\ . \tag{1}$$

Dabei steht das Bild des Basisvektors e_j von R^n unter $A\colon R^n \to R^m$ in der j-ten Spalte der Matrix $(a_{ij})_{i,j} = \Phi(A)$. Matrixmultiplikation $\cdot\colon M_{m,n}(R) \times R^n \to R^m$ entspricht dem Anwenden einer linearen Abbildung, genauer

$$A\left(\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}\right) = \Phi(A) \cdot \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \in R^m \;. \tag{2}$$

Die Matrixaddition entspricht der Addition linearer Abbildungen, für A, $B \in \operatorname{Hom}_R(R^n, R^m)$ gilt also

$$\Phi(A + B) = \Phi(A) + \Phi(B) \;. \tag{3}$$

Für ℓ, m, $n \in \mathbb{N}$ seien $A\colon R^m \to R^\ell$ und $B\colon R^n \to R^m$ rechts-R-linear. Dann gilt

$$\Phi(A \circ B) = \Phi(A) \cdot \Phi(B) \in M_{\ell,n}(R) \;, \tag{4}$$

das heißt, die Matrixmultiplikation $\cdot\colon M_{\ell,m}(R) \times M_{m,n}(R) \to M_{\ell,n}(R)$ entspricht der Verkettung linearer Abbildungen.

Beweis. Wir setzen $I = \{1, \dots, m\}$ und $J = \{1, \dots, n\}$ in Proposition 2.63. Dann ist

$$R^{(I)} = R^m \;, \qquad R^{(J)} = R^n \qquad \text{und} \qquad (R^{(I)})^J = M_{m,n}(R) \;.$$

Beachte, dass I und J hier endliche Mengen sind, es gibt also keine zusätzlichen Bedingungen an die Zahlen a_{ij}, wenn $(a_{ij})_{i,j} \in (R^{(I)})^J = M_{m,n}(R)$. Also liefert Proposition 2.63 (1) die Abbildung in (1). Die j-te Spalte der Matrix ist dabei wie gefordert

$$A(e_j) = (a_{ij})_{i=1,\dots,m} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \;.$$

Es sei wieder $(a_{ij})_{i,j} = \Phi(A)$. Wir schreiben $(r_j)_{j=1,\dots,n}$ für die Spalte aus (2). Aus Proposition 2.63 (2) folgt

$$A\left(\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}\right) = A\bigl((r_j)_{j=1,\dots,n}\bigr) = \left(\sum_{j=1}^n a_{ij} \cdot r_j\right)_{i=1,\dots,m} = (a_{ij})_{i,j} \cdot \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \;.$$

Die letzte Gleichung ist gerade die Definition 2.64 (3) der Matrixmultiplikation in dem Fall, dass der zweite Faktor $(r_j)_{j=1,\dots,n}$ eine Spalte ist.

Zu (3) seien $(a_{ij})_{i,j} = \Phi(A)$, $(b_{ij})_{i,j} = \Phi(B) \in M_{m,n}(R)$. Wir bestimmen $\Phi(A + B)$, indem wir die Bilder der Vektoren $e_k \in R^n$ berechnen. Nach Definition von $A + B$ in Bemerkung 2.39 und Proposition 2.63 (1) gilt

$$(A + B)(e_k) = A(e_k) + B(e_k) = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} + \begin{pmatrix} b_{1k} \\ \vdots \\ b_{mk} \end{pmatrix} = \begin{pmatrix} a_{1k} + b_{1k} \\ \vdots \\ a_{mk} + b_{mk} \end{pmatrix} \in R^m \;,$$

und das ist genau die k-te Spalte der Matrix $\Phi(A) + \Phi(B)$.

Zu (4) sei $(a_{ij})_{i,j} = \Phi(A) \in M_{\ell,m}(R)$ und $(b_{jk})_{j,k} = \Phi(B) \in M_{m,n}(R)$. Um die Matrix $\Phi(A \circ B)$ zu erhalten, müssen wir die Bilder der Vektoren $e_k \in R^n$ bestimmen. Nach Proposition 2.63 (1) ist

$$B(e_k) = (b_{jk})_{j=1,\dots,m} \ .$$

Nach Proposition 2.63 (2) gilt

$$A\bigl(B(e_k)\bigr) = \left(\sum_{j=1}^{m} a_{ij} \cdot b_{jk}\right)_{i=1,\dots,\ell} = \begin{pmatrix} a_{11} \cdot b_{1k} + \dots + a_{1m} \cdot b_{mk} \\ \vdots \\ a_{\ell 1} \cdot b_{1k} + \dots + a_{\ell m} \cdot b_{mk} \end{pmatrix} \in R^\ell \ ,$$

Also hat $\Phi(A \circ B)$ die gleiche k-te Spalte wie das Matrixprodukt $\Phi(A) \cdot \Phi(B)$. Daraus folgt unsere Behauptung. □

2.67. Bemerkung. Nach Folgerung 2.66 bietet es sich an, Matrizen $A = (a_{ij})_{i,j} \in M_{m,n}(R)$ mit den zugehörigen Abbildungen $A\colon R^n \to R^m$ zu identifizieren. Somit ist

$$\mathrm{Hom}_R(R^n, R^m) = M_{m,n}(R) \ .$$

Die Abbildungen Φ und Ψ aus Proposition 2.63 brauchen wir dann natürlich nicht mehr.

Man beachte: auf die Rechts-R-Moduln R^n wirken Matrizen von links. Auf diese Weise kommt die skalare Multiplikation der Matrix nicht „in die Quere“, das heißt, die Multiplikation mit einer Matrix von links ist rechts-R-linear.

Bei Zeilen ist es genau spiegelbildlich: hier operieren Skalare von links wegen Bemerkung 2.65 (4) und Matrizen von rechts, es folgt also

$$_R\mathrm{Hom}({}^mR, {}^nR) = M_{m,n}(R) \ .$$

Wenn man die Verkettung linearer Abbildungen als Matrixprodukt schreibt, dreht sich die Reihenfolge der Faktoren um. Aus diesem Grund ist es einfacher, mit Rechts-R-Moduln zu arbeiten.

Tatsächlich kann man einige Fehler vermeiden, wenn man Skalare konsequent von rechts wirken lässt — selbst dann, wenn man über einem kommutativen Ring oder Körper arbeitet, bei dem es nach Bemerkung 2.23 eigentlich keinen Unterschied zwischen Rechts- und Linksmoduln gibt.

2.68. Folgerung. *Die Matrixmultiplikation ist assoziativ.*

Beweis. Diese Behauptung könnte man beispielsweise mit Hilfe der Definition 2.64 (3) der Matrixmultiplikation mit etwas Aufwand nachrechnen.

Einfacher ist es, Matrizen $A \in M_{\ell,m}(R) = \mathrm{Hom}_R(R^m, R^\ell)$, $B \in M_{m,n}(R) = \mathrm{Hom}_R(R^n, R^m)$ und $C \in M_{n,p}(R) = \mathrm{Hom}_R(R^p, R^n)$ mit den entspechenden linearen Abbildungen zu identifizieren. Da die Verkettung von Abbildungen assoziativ ist nach Bemerkung 2.4 (1), folgt aus Folgerung 2.66 (4), dass

$$A \cdot (B \cdot C) = A \circ (B \circ C) = (A \circ B) \circ C = (A \cdot B) \cdot C) \ . \qquad \square$$

2.69. Definition. Es sei R ein Ring mit Eins. Eine $m \times n$-Matrix über R heißt *quadratisch,* wenn $m = n$. Der Raum der quadratischen $n \times n$-Matrizen über R wird mit $M_n(\mathbb{R})$ bezeichnet. Die quadratische Matrix $E_n = (\delta_{ij})_{i,j} \in M_n(R)$ heißt *Einheitsmatrix.* Eine quadratische Matrix $A \in M_n(R)$ heißt *invertierbar,* wenn es eine Matrix $B \in M_n(R)$ mit $A \cdot B = B \cdot A = E_n$ gibt. In diesem Fall heißt B die zu A *inverse Matrix;* sie wird auch mit A^{-1} bezeichnet.

2.70. Bemerkung. Identifiziere $M_n(R)$ mit $\operatorname{End}_R R^n = \operatorname{Hom}_R(R^n, R^n)$ wie in Bemerkung 2.67.

(1) Die Einheitsmatrix entspricht der Identität id_{R^n}, denn für alle $m = (r_j)_j \in R^n$ gilt
$$E_n \cdot m = \left(\sum_{j=1}^n \delta_{ij}\, r_j\right)_i = (r_i)_i = m\ .$$
(2) Es sei $A\colon R^n \to R^n$ eine lineare Abbildung, $A \in M_n(R)$. Wegen Folgerung 2.66 (4) ist A genau dann als lineare Abbildung umkehrbar, also ein Isomorphismus, wenn A als Matrix invertierbar ist. In diesem Fall wird die Umkehrabbildung von A genau durch die inverse Matrix A^{-1} beschrieben. Aus Folgerung 2.71 (1) unten folgt mit Proposition 2.3, dass die inverse Matrix eindeutig bestimmt ist.

2.71. Folgerung (aus Folgerungen 2.42 und 2.66)**.** *Es sei R ein Ring mit Eins.*

(1) *Die invertierbaren $n \times n$-Matrizen bilden eine Gruppe $(GL(n,R), \cdot\,)$, die* allgemeine lineare Gruppe, *und es gilt $GL(n,R) \cong \operatorname{Aut}_R R^n$.*
(2) *Die quadratischen $n \times n$-Matrizen bilden einen Ring $(M_n(R), +, \cdot\,)$ mit Eins E_n, den* Matrixring, *und es gilt $M_n(R) \cong \operatorname{End}_R R^n$.*
(3) *Der Raum der Spalten R^n wird durch Matrixmultiplikation zu einem unitären $M_n(R)$-Linksmodul.*
(4) *Der Raum $M_{m,n}(R)$ wird durch Matrixmultiplikation zu einem unitären Rechts-$M_n(R)$-Modul und zu einem unitären Links-$M_m(R)$-Modul.*

Beweis. Nach Bemerkung 2.67 gilt $\operatorname{End}_R R^n = M_n(R)$, und nach Bemerkung 2.70 (1) ist $E_n = \mathrm{id}_{R^n}$ die Eins. Es folgt (2).

Nach Bemerkung 2.70 (2) entsprechen die invertierbaren Matrizen genau den umkehrbaren linearen Abbildungen, und es folgt (1).

Die Punkte (3) und (4) folgen aus den entsprechenden Punkten in Folgerung 2.42 und Folgerung 2.66 (2) und (4). □

2.72. Bemerkung. Es sei R Ring mit Eins und M ein Rechts-R-Modul. In Definition 2.43 haben wir den dualen Links-R-Modul
$$M^* = \operatorname{Hom}_R(M; R)$$
eingeführt. Im Spezialfall $M = R^m$ folgt nach den Identifikation aus Bemerkung 2.67 und 2.65 (4), dass
$$(R^m)^* = \operatorname{Hom}_R(R^m, R) = M_{1,m}(R) = {}^mR\ .$$

Somit der Links-R-Modul der m-elementigen Zeilen dual zum Rechts-R-Modul der m-elementigen Spalten.

Es sei $(e_1, \dots, e_m)$ die Standardbasis des R^m. Als Basis der Zeilen wählen wir $\varepsilon_1, \dots, \varepsilon_m$, wobei an der i-ten Stellen von ε_i eine 1 steht und sonst nur Nullen. Diese Basis nennen wir die *Standardbasis* von ${}^mR$. Zwischen den Basen $(e_j)_j$ und $(\varepsilon_i)_i$ besteht die folgenden Beziehung:

$$\varepsilon_i(e_j) = \sum_{k=1}^m \delta_{ik}\delta_{kj} = \delta_{ij} ;$$

wir sagen dazu, dass die Basis $(\varepsilon_i)_i$ *dual* zur Basis $(e_j)_j$ ist.

Für Links-Moduln N können wir analog den Dualraum ${}^*N = {}_R\mathrm{Hom}(N, R)$ definieren. Analog zu oben folgt ${}^*({}^mR) = R^m$, und wiederum ist die Basis $(e_j)_j$ zur Basis $(\varepsilon_i)_i$ dual.

Zum Schluss dieses Abschnitts wollen wir auch in freien Moduln mit festen Basen mit Koordinaten und Matrizen rechnen. Dafür ist es praktisch, den Begriff einer Basis etwas anders zu fassen als in Definition 2.28.

2.73. Definition. Es sei R ein Ring mit Eins und M ein Rechts-R-Modul. Ein Tupel $(b_1, \dots, b_m)$ aus Elementen von M heißt

(1) *Erzeugendensystem,* wenn $\{b_1, \dots, b_m\}$ eine Erzeugermenge bildet;
(2) *linear abhängig,* wenn es $(r_1, \dots, r_m) \in R^m \setminus \{0\}$ gibt, so dass

$$b_1 . r_1 + \dots + b_m . r_m = 0 ,$$

und sonst *linear unabhängig;*
(3) (*angeordnete*) *Basis* von M, wenn es ein linear unabhängiges Erzeugendensystem bildet.

Der Hauptunterschied ist, dass wir hier mit Tupeln anstelle von Mengen arbeiten. Insbesondere hat jedes Basiselement jetzt einen Index aus $\{1, \dots, m\}$, und wegen (2) darf kein Vektor doppelt vorkommen, was in einer Menge wie in Definition 2.28 gar nicht möglich ist. Im Folgenden seien alle Basen angeordnet.

2.74. Bemerkung. Es sei M ein freier Rechts-R-Modul mit Basis $B = (b_1, \dots, b_m)$. Wie in Definition 2.33 erhalten wir eine Basisabbildung $B \colon R^m \to M$ mit

$$B\left(\begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}\right) = \sum_{i=1}^m b_i . r_i .$$

Wir benutzen hier den gleichen Buchstaben wie für die Basis B, und tatsächlich verhält sich die Basisabbildung oben formal wie die Matrixmultiplikation der „Zeile“ B aus Modulementen mit der Spalte $(r_i)_i \in R^m$.

Wie Bemerkung 2.34 ist die Basisabbildung bijektiv, und ihre Umkehrabbildung ist die Koordinatenabbildung $M \to R^m$. Die Basisabbildung ist auch linear, also ein Isomorphismus, denn für alle $(r_i)_i$, $(s_i)_i \in R^m$ und alle $t \in R$ gilt

$$\begin{aligned} B\bigl((r_i)_i + (s_i)_i\bigr) &= \sum_{i=1}^m b_i \,.\, (r_i + s_i) \\ &= \sum_{i=1}^m b_i \,.\, r_i + \sum_{i=1}^m b_i \,.\, s_i = B\bigl((r_i)_i\bigr) + B\bigl((s_i)_i\bigr) \\ \text{und}\quad B\bigl((r_i)_i \,.\, t\bigr) &= \sum_{i=1}^m b_i \,.\, (r_i \cdot t) = \Bigl(\sum_{i=1}^m b_i \,.\, r_i\Bigr) \,.\, t = B\bigl((r_i)_i\bigr) \,.\, t \,. \end{aligned}$$

Nach Proposition 2.40 ist die Koordinatenabbildung dann ebenfalls linear. Die Linearität dieser Abbildungen bedeutet, dass wir mit den Koordinaten genauso rechnen dürfen wie mit den Modulelementen selbst. Es ist also egal, ob wir erst Vektoren addieren und mit Skalaren multiplizieren und dann Koordinaten bilden, oder erst Koordinaten der einzelnen Modulelemente nehmen und dann mit ihnen weiterrechnen.

Sei jetzt umgekehrt M ein beliebiger Rechts-R-Modul, sei $B\colon R^m \to M$ eine lineare Abbildung und $b_i = B(e_i) \in M$ für $i = 1, \dots, m$. Dann bildet $(b_1, \dots, b_m)$ genau dann ein Erzeugendensystem, wenn B surjektiv ist, und genau dann linear unabhängig, wenn B injektiv ist. Also entsprechen die m-elementigen Basen von M genau den Isomorphismen $B\colon R^m \to M$.

2.75. Folgerung. *Es sei R ein Ring mit Eins, es sei M ein freier Rechts-R-Modul mit Basis $B = (b_1, \dots, b_m)$ und N ein freier Rechts-R-Modul mit Basis $C = (c_1, \dots, c_n)$. Dann entspricht jeder linearen Abbildung $F\colon N \to M$ genau eine Matrix A, die* Abbildungsmatrix von F bezüglich B und C *so dass das folgende Diagramm kommutiert.*

$$\begin{tikzcd} N \arrow[r, "F"] & M \\ R^n \arrow[u, "C"] \arrow[r, "A"] & R^m \arrow[u, "B"'] \end{tikzcd}$$

Dabei stehen in der j-ten Spalte von A die B-Koordinaten des Bildes des j-ten Basisvektors c_j. Für jedes Element $v = C((r_i)_i) \in N$ hat das Bild $F(v)$ also die Koordinaten $A \cdot (r_i)_i$.

Beweis. Wir bezeichnen die Umkehrabbildung der Basisabbildung B mit B^{-1} und setzen

$$A = B^{-1} \circ F \circ C \,,$$

dann kommutiert das Diagramm offensichtlich. Die restlichen Aussagen ergeben sich aus Folgerung 2.66. $\square$

2.76. Bemerkung. Der Spezialfall $M = N$ und $F = \mathrm{id}_M$ ist interessant. In diesem Fall erhalten wir das kommutative Diagramm

$$\begin{tikzcd} & M & \\ R^n \arrow[ur, "C"] \arrow[rr, "A"] & & R^n \arrow[ul, "B"'] \end{tikzcd} .$$

Multiplikation mit der Matrix A macht aus C-Koordinaten B-Koordinaten. Also besteht die j-te Spalte von $A = (a_{ij})_{i,j}$ aus den B-Koordinaten des Vektor c_j, das heißt

$$c_j = \sum_{i=1}^{n} b_i a_{ij} .$$

Anders formuliert erhalten wir die Vektoren der Basis C, indem wir die „Zeile“ B mit den Spalten von A multiplizieren. Aus diesem Grund nennt man die Matrix A auch *Basiswechselmatrix.* Die obigen Sachverhalte sind zwei Lesarten der „Gleichung“ $C = BA$. Man beachte, dass die „Richtung“ des Basiswechsels für die Koordinaten („von C nach B“) und für die Basisvektoren („von B nach C“) genau umgekehrt ist. Um Fehler zu vermeiden, sollte man daher immer das obige kommutative Diagramm vor Augen haben.

2.77. Proposition. *Es sei M ein freier R-Modul mit Basis $B(b_1, \dots, b_m)$. Dann besteht eine Bijektion zwischen der Menge der m-elementigen Basen von M und der allgemeinen linearen Gruppe $GL(n, R)$, die jeder Basis $C = B \cdot A$ die Basiswechselmatrix A zuordnet.*

Beweis. Zunächst sei $C = (c_j)_j$ eine weitere Basis von M. Dann erhalten wir eine Basiswechselmatrix A wie in Bemerkung 2.76. Da die Basisabbildungen zu B und C Isomorphismen sind, ist $A = B^{-1} \circ C$ ebenfalls ein Isomorphismus, also ist die zugehörige Matrix nach Bemerkung 2.70 (2) invertierbar. Außerdem wird sie durch B und C eindeutig festgelegt.

Sei jetzt A eine invertierbare Matrix, dann ist $C = B \circ A \colon R^m \to M$ ein Isomorphismus. Die Bilder $c_1, \dots, c_m$ der Standardbasisvektoren $e_1, \dots, e_m$ von R^m bilden eine Basis von M, und C ist die zugehörige Basisabbildung. Denn sei $m \in M$ ein Element mit den B-Koordinaten $r_1, \dots, r_m$, dann folgt

$$m = B((r_j)_j) = \bigl(C \circ A^{-1}\bigr)((r_j)_j) = C\bigl(A^{-1} \cdot (r_j)_j\bigr) = \sum_{i=1}^{m} c_i \cdot s_i ,$$

wobei s_i die i-te Komponente der Spalte $A^{-1} \cdot (r_j)_j$ sei. Mithin lässt sich jedes Element von M als Linearkombination der c_i schreiben, das heißt, die Elemente $c_1, \dots, c_m$ erzeugen M.

Wenn eine dieser Linearkombinationen das Nullelement $0 \in M$ ergibt, folgt umgekehrt, dass

$$0 = \sum_{i=1}^{m} c_i \cdot s_i = B\bigl(A \cdot (s_i)_i\bigr) ,$$

also ist $A \cdot (s_i)_i = 0$ und wegen Invertierbarkeit von A auch $(s_i)_i$. Also ist das Tupel $(c_1, \dots, c_m)$ linear unabhängig und bildet daher eine Basis. □

2.78. Bemerkung. Wir können jetzt auch überlegen, wie sich die Abbildungsmatrix aus Proposition 2.75 verhält, wenn wir eine der beiden Basen durch eine andere ersetzen. Wir betrachten dazu die kommutativen Diagramme

$$\begin{tikzcd} & N \arrow[rr, "F"] & & M & \\ R^n \arrow[ur, "C"] \arrow[rr, "A"] & & R^m \arrow[ur, "B"] \arrow[rr, "P"] & & R^m \arrow[ul, "D"'] \end{tikzcd} \quad \text{und} \quad \begin{tikzcd} & N \arrow[rr, "F"] & & M & \\ R^n \arrow[ur, "E"] \arrow[rr, "Q"] & & R^n \arrow[ul, "C"'] \arrow[rr, "A"] & & R^m \arrow[ul, "B"'] \end{tikzcd} .$$

Hier ist D eine neue Basis von M und $P \in GL(m, R)$ die zugehörige Basiswechselmatrix, und E ist eine Basis von N und $Q \in GL(n, R)$ die zugehörige Basiswechselmatrix.

Es sei wieder $(\varepsilon_i)_i$ die Standardbasis des Raumes ${}^mR = \mathrm{Hom}_R(R^m, R)$ der Zeilen der Länge R.

2.79. Proposition. *Es sei R ein Ring mit Eins und M ein freier Modul mit Basis $B = (b_1, \dots, b_m)$. Dann bilden die einzelnen Komponentenfunktionen $\beta_i = \varepsilon_i \circ B^{-1} \colon M \to R$ der Koordinatenabbildung $B^{-1} \colon M \to R^m$ eine Basis $(\beta_i)_i$ des dualen Moduls B^*. Sie ist* dual *zur Basis B, das heißt, für alle i, j gilt*

$$\beta_i(b_j) = \delta_{ij} . \tag{1}$$

Beweis. Die Abbildungen β_i sind offensichtlich Elemente des dualen Moduls M^*. Da $b_j = B(e_j)$ gilt, folgt (1), denn

$$\beta_i(b_j) = \big(\varepsilon_i \circ B^{-1}\big)(B(e_j)) = \varepsilon_i(e_j) = \delta_{ij}$$

nach Bemerkung 2.72.

Aus (1) folgt, dass $(\beta_i)_i$ eine Basis von M^* ist. Sei etwa $F \in M^* = \mathrm{Hom}_R(M, R)$, und sei $m = B((r_j)_j) \in M$ ein Element mit den B-Koordinaten $(r_j)_j \in R^m$, dann gilt

$$\begin{aligned} F(m) = \sum_{j=1}^m F(b_j) \cdot r_j = \sum_{i,j=1}^m F(b_i) \cdot \delta_{ij} \cdot r_j \\ = \sum_{i,j=1}^m F(b_i) \cdot \beta_i(b_j) \cdot r_j = \Bigg(\sum_{i=1}^m F(b_i) \cdot \beta_i\Bigg)(m) , \end{aligned}$$

somit $F = \sum_{i=1}^m F(b_i) \cdot \beta_i$, und die Elemente β_i erzeugen M^*.

Sie sind auch linear unabhängig, denn wäre $\sum_{i=1}^m s_i \cdot \beta_i = 0$, so würde für alle j folgen, dass

$$s_j = \sum_{i=1}^m s_i \cdot \delta_{ij} = \sum_{i=1}^m s_i \cdot \beta_i(b_j) = 0 .$$

Also ist das Tupel $(\beta_1, \dots, \beta_m)$ linear unabhängig und bildet daher eine Basis.

$\square$

Zum Schluss betrachten wir als Spezialfall bestimmte Basen des $\mathbb{R}^n$, mit denen man besonders gut arbeiten kann. Mehr dazu erfahren Sie in Abschnitt 7.2 unten. Außerdem brauchen wir den Begriff der transponierten und der adjungierten Matrix.

2.80. Definition. Es sei $A = (a_{ij})_{i,j} \in M_{m,n}(R)$ eine Matrix, dann definieren wir die zu A *transponierte Matrix* $A^t \in M_{n,m}(R)$ durch $A^t = (a_{ij})_{ji}$. Falls $R = \mathbb{C}$ oder $\mathbb{H}$, definieren wir die zu A *adjungierte Matrix* $A^* \in M_{n,m}(R)$ durch $A^* = (\bar{a}_{ij})_{j,i}$.

Transponierten macht zum Beispiel aus Zeilen Spalten und umgekehrt. In Büchern wird häufig $(r_1, \dots, r_n)^t$ für die Spalte $\left(\begin{smallmatrix} r_1 \\ \vdots \\ r_n \end{smallmatrix}\right)$ geschrieben.

Wir können das Standardskalarprodukt auf $\mathbb{R}^n$ zweier Spaltenvektoren $v = (v_i)_i$ und $w = (w_i)_i$ aus Definition 1.51 (1) jetzt auch schreiben als

$$\langle v, w \rangle = \sum_{i=1}^n v_i \cdot w_i = v^t \cdot w \ .$$

2.81. Definition. Eine *Orthonormalbasis* des $\mathbb{R}^n$ ist ein Tupel $B = (b_1, \dots, b_n)$ von Vektoren im $\mathbb{R}^n$, so dass für alle i, j gilt

$$\langle b_i, b_j \rangle = \delta_{ij} \ .$$

2.82. Proposition. *Es sei $B = (b_1, \dots, b_n)$ eine Orthonormalbasis des $\mathbb{R}^n$. Dann ist B eine Basis, und für jeden Vektor $v \in \mathbb{R}^n$ gilt*

$$v = \sum_{i=1}^n b_i \,.\, \langle b_i, v \rangle \ .$$

Die Matrix B mit den Spalten b_1, …, b_n ist invertierbar mit $B^{-1} = B^t$.

Man beachte, dass die Matrix B jetzt tatsächlich die Matrix der Basisabbildung B ist, so dass die obige Merkregel aus Bemerkung 2.74 hier wirklich richtig ist. Im Spezialfall einer Orthonormalbasis wird die Koordinatenabbildung also gegeben durch $B^{-1} = B^t$. Im Allgemeinen lässt sich das Inverse einer Matrix nicht so leicht bestimmen, und allgemeine Verfahren zum Invertieren von Matrizen lernen wir erst in den nächsten zwei Kapiteln kennen.

Beweis. Wir schreiben $b_j = (b_{ij})_i$, so dass $B = (b_{ij})_{i,j}$. Dann gilt also

$$\langle e_i, b_j \rangle = \sum_{k=1}^n \delta_{ki} b_{kj} = b_{ij} \ .$$

Wir versuchen, den Vektor e_i als Linearkombination der b_j darzustellen. Dazu setzen wir

$$r_i = e_i - \sum_{j=1}^n b_j \,.\, \langle b_j, e_i \rangle \in \mathbb{R}^n \ .$$

Es folgt

$$\begin{aligned}
\|r_i\|^2 &= \Bigl\langle e_i - \sum_{j=1}^n b_j \,.\, \langle b_j, e_i\rangle, e_i - \sum_{k=1}^n b_k \,.\, \langle b_k, e_i\rangle \Bigr\rangle \\
&= \|e_i\|^2 - 2\sum_{j=1}^n \langle e_i, b_j\rangle \cdot \langle b_j, e_i\rangle + \sum_{j,k=1}^n \underbrace{\langle b_j, b_k\rangle}_{=\delta_{jk}} \cdot \langle b_j, e_i\rangle \cdot \langle b_k, e_i\rangle \\
&= 1 - \sum_{j=1}^n b_{ij}^2 \;.
\end{aligned}$$

Nun ist $\|r_i\|^2 \geq 0$ nach Bemerkung 1.52 (3). Auf der anderen Seite liefert Summieren über i, dass

$$\sum_{i=1}^n \|r_i\|^2 = n - \sum_{i=1}^n \sum_{j=1}^n b_{ij}^2 = n - \sum_{j=1}^n \sum_{i=1}^n b_{ij}^2 = n - \sum_{j=1}^n \langle b_j, b_j\rangle = 0 \;,$$

so dass $\|r_i\|^2 = 0$ und daher $r_i = 0$ für alle i, und somit gilt

$$e_i = \sum_{j=1}^n b_j \,.\, \langle b_j, e_i\rangle = \sum_{j=1}^n b_j \,.\, b_{ij} \;.$$

Wir stellen den Vektor $v = (r_i)_i$ in Koordinaten dar und erhalten

$$v = \sum_{i=1}^n e_i \,.\, r_i = \sum_{i,j=1}^n b_j \,.\, \langle b_j, e_i\rangle \,.\, r_i = \sum_{j=1}^n b_j \,.\, \langle b_j, v\rangle \;.$$

Also erzeugen die Vektoren $b_1, \dots, b_n$ den $\mathbb{R}^n$.

Sie sind auch linear unabhängig, denn aus $b_1 \,.\, r_1 + \dots + b_n \,.\, r_n = 0$ mit $r_1, \dots, r_n \in \mathbb{R}$ folgt für alle i, dass

$$r_i = \sum_{j=1}^n \delta_{ij} \cdot r_j = \Bigl\langle b_i, \sum_{i=1}^n b_j \,.\, r_j \Bigr\rangle = \langle b_i, 0\rangle = 0 \;.$$

Also bildet $(b_1, \dots, b_n)$ eine Basis.

Schließlich berechnen wir noch

$$E_n = (\delta_{ij})_{i,j} = (\langle b_i, b_j\rangle)_{i,j} = \Bigl(\sum_{k=1}^n b_{ki} b_{kj}\Bigr)_{i,j} = B^t \cdot B$$

und

$$\begin{aligned}
E_n &= (\delta_{ij})_{i,j} = (\langle e_i, e_j\rangle)_{i,j} \\
&= \Bigl(\Bigl\langle \sum_{k=1}^n b_k \,.\, b_{ik}, \sum_{\ell=1}^n b_\ell \,.\, b_{j\ell} \Bigr\rangle\Bigr)_{i,j} = \Bigl(\sum_{k=1}^n b_{ik} b_{jk}\Bigr)_{i,j} = B \cdot B^t
\end{aligned}$$

und schließen daraus, dass $B^{-1} = B^t$. □

Es folgt eine kurze Zwischenbilanz zum Ende des Abschnitts: In Abschnitt 2.1 haben wir Gruppen, Ringe und Körper kennengelernt. Uns interessieren dabei am meisten Ringe mit Eins, darunter fallen auch Körper und Schiefkörper.

Im Abschnitt 2.2 haben wir Moduln betrachtet. In Zukunft werden wir fast nur noch mit unitären Moduln arbeiten, dazu gehören auch die Vektorräume über Körpern und Schiefkörpern. In den Beispielen 2.30 und 2.31 haben wir die frei erzeugten Moduln $R^{(I)}$ und speziell den Raum R^n der Spalten kennengelernt kennengelernt. Freie Moduln werden besondern wichtig werden, vor allem, da Vektorräume immer freie Moduln sind.

Der Inhalt von Abschnitt 2.3 waren lineare Abbildungen. Wir haben gesehen, wie man lineare Abbildungen $R^n \to R^m$ und allgemeiner zwischen endlich erzeugten freien Moduln durch Matrizen darstellen kann. Außerdem haben wir Folgerung 2.42 mit Hilfe von Matrizen neu interpretiert. Als Spezialfall haben wir den dualen Modul aus Definition 2.43 betrachtet.

In Abschnitt 2.4 ging es um Unterräume, Quotienten und (direkte) Summen. Diese Konstruktionen schauen wir uns näher an, sobald wir mehr über Basen von Vektorräumen wissen.

Schließlich haben wir in Abschnitt 2.5 Matrixrechnung kennengelernt. Sie hat zweierlei Aufgaben: zum einen erlaubt sie es, mit linearen Abbildungen zu rechnen, indem man sie durch Systeme von Zahlen darstellt. Dieser Aspekt ist später zum Beispiel in der Numerik sehr wichtig. Dazu muss man zunächst für jedes Modul eine Basis wählen — im Fall R^m wird das häufig die Standardbasis sein. Zum anderen haben wir Matrizen aber auch benutzt, um den Raum aller linearen Abbildungen besser zu verstehen. Dieser Aspekt wird im Folgenden häufig wichtig sein.

KAPITEL 3

Vektorräume über Körpern und Schiefkörpern

In diesem Kapitel lernen wir typische Eigenschaften von Vektorräumen über (Schief-) Körpern kennen. Insbesondere hat jeder Vektorraum eine Basis, ist also als Modul frei. Außerdem lernen wir das Gauß-Verfahren zum Lösen linearer Gleichungssysteme kennen. Solche linearen Gleichungssystem treten sowohl in der Praxis als auch in der Theorie häufig auf. So können wir das Gauß-Verfahren auch benutzen, um festzustellen, ob eine Matrix invertierbar ist, und gegebenenfalls die inverse Matrix zu bestimmen.

Alles, was in diesem Abschnitt passiert, beruht darauf, dass wir in einem Schiefkörper dividieren können. Auf der anderen Seite benötigen wir das Kommutativgesetz in diesem Abschnitt (noch) nicht. Für den Rest dieses Kapitels sei $\Bbbk$ ein Schiefkörper, also zum Beispiel $\Bbbk = \mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ oder $\mathbb{Z}/p\mathbb{Z}$ für p prim. Wenn nichts anderes angegeben wird, seien alle $\Bbbk$-Vektorräume nach wie vor Rechts-Vektorräume, und alle Basen seien angeordnet wie in Definition 2.73.

3.1. Basen

Wir haben spätestens im Abschnitt 2.5 gesehen, dass wir in freien Moduln weitaus leichter rechnen können als in beliebigen. Und wir haben auch gesehen, dass wir dadurch die Struktur dieser Moduln und der linearen Abbildungen gut beschreiben und verstehen können. Das soll diesen Abschnitt motivieren, in dem wir uns Gedanken über die Existenz von Basen machen wollen. Die beiden Sätze von Steinitz gehören zu den wichtigsten Ergebnissen dieser Vorlesung.

Für das folgende Lemma führen wir noch folgende Notation ein. Es sei $(a_1, \dots, a_n)$ ein Tupel und $1 \le i \le n$. Dann erhalten wir ein neues Tupel durch Weglassen von a_i, das wir bezeichnen wollen als

$$\bigl(a_1, \dots, \widehat{a_i}, \dots, a_n\bigr) = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) \,.$$

3.1. Lemma. *Es sei V ein $\Bbbk$-Vektorraum und $(v_1, \dots, v_n)$ ein linear abhängiges Tupel von Vektoren aus V. Dann existiert ein $j \in \{1, \dots, n\}$, so dass sich v_j als Linearkombination der Vektoren $\bigl(v_1, \dots, \widehat{v_j}, \dots, v_n\bigr)$ darstellen lässt. Falls das Tupel $(v_1, \dots, v_r)$ linear unabhängig ist für ein $r < n$, können wir $j > r$ wählen.*

Beweis. Da das Tupel $(v_1, \dots, v_n)$ linear abhängig ist, existieren $k_1, \dots, k_n \in \Bbbk$, so dass

$$\sum_{i=1}^{n} v_i \, k_i = 0 \in V \,.$$

Wäre $k_{r+1} = \cdots = k_n = 0$, so erhielten wir eine nicht-triviale Linearkombination des Tupels $(v_1, \dots, v_r)$, die den Nullvektor darstellt. Da $(v_1, \dots, v_r)$ nach Voraussetzung linear unabhängig ist, ist das nicht möglich. Wir finden also $j > r$ mit $k_j \neq 0$. Aus der obigen Gleichung folgt jetzt

$$v_j = -\sum_{i \neq j} v_i \, (k_i \, k_j^{-1}) \; . \qquad \square$$

3.2. Bemerkung. Man beachte, dass wir im Beweis durch k_j dividiert haben. Wir können daher nicht erwarten, dass das Lemma für Moduln über beliebigen Ringen gilt. Als Gegenbeispiel betrachte $\mathbb{Z}$ als $\mathbb{Z}$-Modul. Das Tupel $(2, 3)$ ist linear abhängig, da $2 \cdot 3 - 3 \cdot 2 = 0$. Aber weder ist 2 eine Linearkombination, also ein Vielfaches, der 3, noch umgekehrt. Die Voraussetzung, dass $\Bbbk$ ein (Schief-) Körper ist, ist also notwendig. Für die meisten Aussagen in diesem und im nächsten Abschnitt finden wir Gegenbeispiele in Moduln über beliebigen Ringen.

3.3. Satz (Basisergänzungssatz von Steinitz)**.** *Es sei V ein $\Bbbk$-Vektorraum. Es sei $(v_1, \dots, v_r)$ ein Tupel linear unabhängiger Vektoren, und $\{w_1, \dots, w_s\} \subset V$ sei eine endliche Erzeugermenge. Dann gibt es $n \geq r$ und Zahlen $i(r + 1)$, …, $i(n) \in \{1, \dots, s\}$, so dass das Tupel*

$$(v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)})$$

eine Basis von V bildet.

Wir bezeichnen das Erzeugnis der Vektoren eines Tupels B mit $\langle B \rangle$, siehe Definition 2.24.

Beweis. Wir setzen zunächst $n = r$ und $i(r) = 0$, und starten mit dem Tupel $B = (v_1, \dots, v_r)$. Dann gehen wir die Vektoren w_i für $i = 1, \dots, s$ der Reihe nach durch.

Wenn wir uns um w_i kümmern, nehmen wir an, dass wir bereits ein linear unabhängiges Tupel

$$B = (v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)})$$

konstruiert haben mit $i(1) < \cdots < i(n) < i$, so dass

$$\{w_1, \dots, w_{i-1}\} \subset \langle B \rangle$$

für alle $j < i$.

Dann gibt es zwei Möglichkeiten. Falls das Tupel

$$(v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)}, w_i)$$

linear abhängig ist, ist nach Lemma 3.1 der Vektor w_i eine Linearkombination der Vektoren aus B, das heißt, es gilt $w_i \in \langle B \rangle$. Es folgt also

$$\{w_1, \dots, w_i\} \subset \langle B \rangle \; ,$$

und wir können den Vektor w_i überspringen.

Falls das obige Tupel linear unabhängig ist, wird es unser neues B, also

$$B = (v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)}, w_i) \;.$$

Selbstverständlich gilt nun auch $w_i \in \langle B \rangle$. Wir setzen also $i(n+1) = i$ und erhöhen anschließend n um 1.

Am Schluss erhalten wir ein lineares Tupel

$$B = (v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)})$$

mit der Eigenschaft, dass

$$\{w_1, \dots, w_s\} \subset \langle B \rangle \;.$$

Nach Voraussetzung ist jeder Vektor $v \in V$ eine Linearkombination der Vektoren $w_1, \dots, w_s$. Jeder dieser Vektoren ist wiederum eine Linearkombination der Vektoren aus B. Indem wir diese Darstellungen der w_j in die obige Darstellung von v einsetzen, erhalten wir v als Linearkombination der Vektoren aus B. Also ist B nun auch ein Erzeugendensystem, und somit eine Basis. □

3.4. Satz (Basisaustauschsatz von Steinitz)**.** *Es sei V ein $\Bbbk$-Vektorraum, es sei $(v_1, \dots, v_r)$ ein linear unabhängiges Tupel, und $(w_1, \dots, w_s)$ sei ein Erzeugensystem. Dann gilt $r \le s$.*

Beweis. Wir dürfen annehmen, dass $B_0 = (v_1, \dots, v_r)$ bereits eine Basis von V ist. Anderfalls ergänzen wir nach Satz 3.3 zu einer Basis

$$(v_1, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n)}) \;.$$

Das folgende Argument wird uns $n \le s$ liefern. Da $r \le n$, folgt erst recht $r \le s$.

Wir gehen die Indizes $j = 1, \dots, r$ der Reihe nach durch. Wenn wir j behandeln, nehmen wir an, dass

$$B_{j-1} = \bigl(v_j, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n_{j-1})}\bigr)$$

bereits eine Basis von V ist mit der Länge

$$(r - (j-1)) + (n_j - r) = n_{j-1} - (j-1) \ge r \;.$$

Durch Weglassen von v_j entsteht ein Tupel

$$B'_j = \bigl(v_{j+1}, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n_{j-1})}\bigr) \;,$$

das nach wie vor linear unabhängig ist. Wäre $v_j \in \langle B'_j \rangle$, also

$$v_j = \sum_{\ell=j+1}^{r} v_\ell \, k_\ell + \sum_{\ell=r+1}^{n_{j-1}} w_{i(\ell)} \, k_\ell$$

für geeignete $k_1, \dots, k_{n_{j-1}} \in \Bbbk$, dann erhielten wir eine nichttriviale Linearkombination

$$0 = v_j - \sum_{\ell=j+1}^{r} v_\ell \, k_\ell - \sum_{\ell=r+1}^{n_{j-1}} w_{i(\ell)} \, k_\ell \;,$$

was nicht möglich ist, da B_{j-1} nach Annahme linear unabhängig ist.

Es folgt $v_j \notin \langle B'_j \rangle$, also ist B'_j keine Basis. Nach Satz 3.3 können wir B'_j zu einer Basis

$$B_j = \bigl(v_{j+1}, \dots, v_r, w_{i(r+1)}, \dots, w_{i(n_j)}\bigr)$$

ergänzen, indem wir mindestens einen weiteren Vektor aus $\{w_1, \dots, w_s\}$ ergänzen. Es folgt also $n_j \geq n_{j-1} + 1$, und somit

$$r \leq n_{j-1} - (j-1) \leq n_j - j \ .$$

Zum Schluss erhalten wir eine Basis

$$B_r = \bigl(w_{i(r+1)}, \dots, w_{i(n_r)}\bigr)$$

der Länge $n_r - r \geq r$. Für $j \neq k$ folgt $i(j) \neq i(k)$, ansonsten erhielten wir die nichttriviale Linearkombination

$$0 = w_{i(j)} - w_{i(k)} \ .$$

Also enthält das ursprüngliche Tupel $(w_1, \dots, w_s)$ mindestens $n_r - r \geq r$ verschiedene Elemente, es folgt $r \leq s$. □

3.5. Folgerung. *Es sei V ein endlich erzeugter $\Bbbk$-Vektorraum. Dann existiert eine Basis $B = (v_1, \dots, v_n)$ von V, und alle anderen Basen von V haben ebenfalls n Elemente.*

Selbstverständlich gelten analoge Aussagen auch für Links-$\Bbbk$-Vektorräume.

Beweis. Es sei $(w_1, \dots, w_s)$ ein Erzeugendensystem von V. Der Basisergänzungssatz 3.3 liefert uns, ausgehend vom leeren Tupel $()$ mit $r = 0$, eine Basis $B = (v_1, \dots, v_n)$ von V, deren Länge $n \leq s$ endlich ist.

Sei nun $C \subset V$ eine beliebige ungeordnete Basis von V. Wir können jeden Vektor w_i als Linearkombination von Vektoren aus C darstellen, dazu benötigen wir aber nur endlich viele. Da $(w_1, \dots, w_s)$ Erzeugendensystem ist, ist jeder Vektor $v \in V$ als Linearkombination der w_i darstellbar. In diese Linearkombination setzen wir die obigen Darstellungen der w_i ein. Insgesamt erhalten wir v als Linearkombination der Vektoren aus C, wobei wir aber nur eine feste endliche Teilmenge $C_0 \subset C$ benötigen, nämlich nur diejenigen Elemente von C, die in einer der Darstellungen der w_i mit Koeffizient $\neq 0$ vorkommen. Alle anderen Vektoren $c \in C \setminus C_0$ lassen sich als Linearkombination der w_i darstellen, also auch als Linearkombination der Vektoren aus C_0. Wäre $C \neq C_0$, so wäre C insbesondere linear abhängig. Wir schließen also, dass die Basis C endlich ist.

Jetzt können wir C anordnen zu $(u_1, \dots, u_s)$. Indem wir B als linear unabhängiges Tupel und C als Erzeugendensystem auffassen, erhalten wir $n \leq s$ aus dem Basisaustauschsatz 3.4. Wir können die Rolle der beiden Basen auch vertauschen, und erhalten $s \leq n$. Also haben B und C gleich viele Elemente. □

3.6. Bemerkung. Man kann analoge Sätze auch für beliebige, nicht notwendig endlich erzeugte $\Bbbk$-Vektorräume beweisen. Dazu braucht man allerdings ein weiteres Axiom für die zugrundeliegende Mengenlehre, das *Auswahlaxiom.* Es ist äquivalent zum folgenden *Lemma von Zorn* (zuerst formuliert von Kuratowski):

Es sei M eine Menge mit einer Halbordnung $\preceq\ \subset M \times M$, siehe Definition 1.34. Wenn zu jeder total geordneten Teilmenge, also zu jeder Teilmenge $U \subset M$, für die die Einschränkung $\preceq \cap (U \times U)$ eine Ordnung ist, eine obere Schranke existiert, also ein Element $m \in M$ mit $u \preceq m$ für alle $u \in U$, dann gibt es ein maximales Element in M, also ein Element $m_0 \in M$, so dass $m_0 \preceq n$ für kein $n \in M \setminus \{m_0\}$ gilt.

Um jetzt beispielsweise den Basisergänzungssatz 3.3 zu verallgemeinern, starten wir mit einer linear unabhängigen Teilmenge $U \subset V$ und einer Erzeugermenge $W \subset V$. Wir betrachten die Menge

$$\mathcal{M} = \bigl\{\, A \subset V \bigm| A \text{ ist linear unabhängig und } U \subset A \subset U \cup W \,\bigr\} \subset \mathcal{P}(V)$$

mit der Halbordnung „$\subset$“. Sei $\mathcal{U} \subset \mathcal{M}$ eine total geordnete Teilmenge, dann betrachten wir

$$M = \bigcup \mathcal{U} = \bigl\{\, v \in V \bigm| \text{es gibt ein } A \in \mathcal{U} \text{ mit } v \in A \bigr\}\ .$$

Wenn eine Linearkombination von Elementen aus M den Nullvektor darstellt, gibt es nur endlich viele Elemente $a_1, \dots, a_n \in M$, deren Koeffizienten von 0 verschieden sind. Jeder Vektor a_i liegt in einer Menge $A_i \in \mathcal{U}$. Da $\mathcal{U}$ total geordnet ist, dürfen wir (nach Umnummerieren) annehmen, dass

$$A_1 \subset \dots \subset A_n \subset M\ .$$

Aber A_n ist linear unabhängig, also verschwinden auch alle Koeffizienten der Vektoren $a_1, \dots, a_n$ in der obigen Linearkombination. Das zeigt, dass M linear unabhängig ist, und damit eine obere Schranke für $\mathcal{U}$ in $\mathcal{M}$.

Jetzt wenden wir das Zornsche Lemma auf $\mathcal{M}$ an und erhalten ein maximales Element $B \in \mathcal{M}$. Also ist B eine linear unabhängige Teilmenge von V mit $U \subset B \subset U \cup W$. Maximalität bedeutet, dass die Hinzunahme eines weiteren Vektors $w \in W \setminus B$ die lineare Unabhängigkeit zerstört. Mit ähnlichen Argumenten wie in Lemma 3.1 folgt daraus, dass B bereits den Vektorraum V erzeugt.

Der Nachteil im obigen Beweis besteht darin, dass man im Allgemeinen keine Chance hat, eine Basis explizit anzugeben. Ein Beispiel dafür ist der Raum $\mathbb{R}^{\mathbb{N}}$ aller reellwertigen Folgen, siehe dazu den Kommentar nach Beispiel 2.30. Dennoch kann aus dem allgemeinen Basisergänzungssatz interessante Schlussfolgerungen ziehen, siehe unten.

3.2. Dimension und Rang

Wir benutzen die Basissätze, um ein paar interessante Aussagen über Vektorräume und ihre Unterräume, Quotienten und über lineare Abbildungen zu beweisen.

Aufgrund von Folgerung 3.5 ist die folgende Definition sinnvoll.

3.7. Definition. Es sei V ein endlich erzeugter $\Bbbk$-Vektorraum. Dann ist die *Dimension* $\dim V$ von V die Länge n einer Basis $(v_1, \dots, v_n)$ von V, und wir nennen V *endlichdimensional.* Wenn V keine Basis endlicher Länge besitzt, heißt V *unendlichdimensional.*

Die Begriffe „endlichdimensional“ und „endlich erzeugt“ für Vektorräume sind nach Folgerung 3.5 äquivalent.

3.8. Folgerung. *Sei endlichdimensionale $\Bbbk$-Vektorräume V und W sind genau dann isomorph, wenn $\dim V = \dim W$.*

Beweis. Zu „$\Longrightarrow$“ sei $F\colon V \to W$ ein Isomorphismus. Wir wählen eine Basis $C = (c_1, \dots, c_n)$ von V, wobei $n = \dim V$, und identifizieren wieder C mit der zugehörigen Basisabbildung. Dann ist die Abbildung $B = F \circ C\colon \Bbbk^n \to W$ ein Isomorphismus, und das Diagramm

$$\begin{tikzcd} V \arrow[r, "F"] & W \\ \Bbbk^n \arrow[u, "C"] \arrow[r, "\mathrm{id}_{\Bbbk^n}", "E_n"'] & \Bbbk^n \arrow[u, "B"'] \end{tikzcd}$$

kommutiert. Wie im Beweis von Proposition 2.77 überzeugt man sich leicht, dass $b_1 = B(e_1)$, $\dots$, $b_n = B(e_n)$ eine Basis von W bilden, so dass insbesondere $\dim W = n = \dim V$.

Zu „$\Longleftarrow$“ sei $n = \dim V = \dim W$. Wir wählen Basen von V und W mit Basisabbildungen $B\colon \Bbbk^n \to W$ und $C\colon \Bbbk^n \to V$. Nach Bemerkung 2.74 sind Basisabbildungen Isomorphismen. Wir erhalten also einen Isomorphismus $F = B \circ C^{-1}\colon V \to W$, so dass das obige Diagramm wieder kommutiert. □

Wir erinnern uns an die Begriffe „direkte Summe“ und „komplementärer Unterraum“ aus Definition 2.56.

3.9. Proposition. *Es sei V ein $\Bbbk$-Vektorraum von endlicher Dimension und $U \subset V$ ein Unterraum. Dann besitzt U ein Komplement $W \subset V$, und es gilt die Dimensionsformel*

$$\dim V = \dim U + \dim W\ .$$

Beweis. Wir beginnen mit einer Basis B von U. Da B als Tupel in V linear unabhängig ist, hat B eine Länge $r = \dim U \le n = \dim V$. Es sei also $B = (v_1, \dots, v_r)$. Wir ergänzen B zu einer Basis $(v_1, \dots, v_n)$ von V mit dem Basisergänzungssatz 3.3.

Es sei $W = \langle v_{r+1}, \dots, v_n \rangle$, dann ist das Tupel $(v_{r+1}, \dots, v_n)$ eine Basis von W, denn es erzeugt W und ist als Teil einer Basis von V auch linear unabhängig. Insbesondere gilt

$$\dim V = \dim U + \dim W\ .$$

Außerdem gilt

$$U + W = \langle v_1, \dots, v_n \rangle = V\ .$$

Sei nun $v \in U \cap W$. Dann existieren $k_1, \dots, k_r \in \Bbbk$ und $\ell_{r+1}, \dots, \ell_n \in \Bbbk$ mit

$$\sum_{i=1}^{r} v_i\,k_i = v = \sum_{j=r+1}^{n} v_j\,\ell_j\ .$$

Beides sind Darstellung als Linearkombination der Basis $(v_1, \dots, v_n)$. Nach Proposition 2.32 sind die Koordinaten von v eindeutig, also gilt $k_1 = \dots = k_r = 0 = \ell_{r+1} = \dots = \ell_n$. Insbesondere folgt $U \cap W = \{0\}$, also $V = U \oplus W$. ☐

3.10. Folgerung. *Es seien U und W zwei Unterräume eines endlichdimensionalen $\Bbbk$-Vektorraums V. Dann sind äquivalent*

(1) $$V = U \oplus W\ ,$$

(2) $$V = U + W \quad \textit{und} \quad \dim U + \dim W \le \dim V\ ,$$

(3) $$U \cap W = \{0\} \quad \textit{und} \quad \dim U + \dim W \ge \dim V\ .$$

Beweis. Die Richtungen „(1) $\Longrightarrow$ (2)“ und „(1) $\Longrightarrow$ (3)“ folgen sofort aus der Definition 2.56 der direkten Summe und Proposition 3.9.

In den Übungen beweisen Sie die Dimensionsformel für Summen

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W)\ .$$

Aus (2) schließen wir, dass

$$0 \le \dim(U \cap W) = \dim U + \dim W - \dim V \le 0\ ,$$

aber also wird $U \cap W$ von einer Basis der Länge 0 erzeugt, das heißt $U \cap W = \{0\}$, und es folgt (1).

Aus (3) schließen wir, dass

$$\dim V \ge \dim(U + W) = \dim U + \dim W - \dim\{0\} \ge \dim V\ ,$$

also hat $U + W$ eine Basis der Länge $\dim V$. Wäre $U + W$ eine echte Teilmenge von V, so könnten wir zu einer Basis von V der Länge $\ge \dim V + 1$ ergänzen, im Widerspruch zu Folgerung 3.5. Also gilt $U + W = V$, und wieder folgt (1). ☐

3.11. Folgerung. *Es sei V ein $\Bbbk$-Vektorraum von endlicher Dimension und $U \subset V$ ein Unterraum. Dann gilt*

$$\dim(V/U) = \dim V - \dim U\ .$$

Beweis. Wir wählen einen zu U komplementären Unterraum $W \subset V$. Aus den Propositionen 2.58 und 3.9 folgt

$$\dim(V/U) = \dim W = \dim V - \dim U\ .$$ ☐

3.12. Bemerkung. Wenn V unendlichdimensional ist, können wir mit dem allgemeineren Basisergänzungssatz aus Bemerkung 3.6 immer noch zu jedem Unterraum einen komplementären Unterraum konstruieren. Da man aber unendliche Dimensionen nicht subtrahieren kann, ist die Dimensionsformel in Proposition 3.9 nicht geeignet, um die Dimension des Komplements zu bestimmen.

Als Beispiel betrachten wir den Raum $V = \mathbb{R}^{(\mathbb{N})}$ der endlichen reellwertigen Folgen mit der Basis $(e_j)_{j\in\mathbb{N}}$, wobei wieder $e_j = (\delta_{ij})_{i\in\mathbb{N}}$, siehe dazu den Kommentar nach Beispiel 2.30. Wir betrachten zwei unendlichdimensionale Unterräume

$$U = \langle e_r, e_{r+1}, e_{r+2}, \dots \rangle \qquad \text{und} \qquad W = \langle e_0, e_2, e_4, \dots \rangle \;.$$

Beide sind als Vektorräume isomorph, denn wir können einen Isomorphismus $F\colon U \to W$ angeben mit $F(e_{r+j}) = e_{2j}$ für alle $j \in \mathbb{N}$. Aber U besitzt ein endlichdimensionales Komplement $\langle e_0, \dots, e_{r-1}\rangle$, während W ein unendlichdimensionales Komplement $\langle e_1, e_3, e_5, \dots \rangle$ hat. Und da nach Proposition 2.58 alle Komplemente von U zu V/U isomorph sind, und alle Komplemente von W zu V/W, können wir die Dimension des Komplementes nun nicht mehr aus der Dimension der Räume selbst ablesen.

Übrigens hat auch $\mathbb{R}^{(\mathbb{N})}$ selbst im Raum $\mathbb{R}^{\mathbb{N}}$ aller reellwertigen Folgen ein Komplement. Da wir das aber wieder mit Hilfe des Zornschen Lemma beweisen müssen, können wir das Komplement nicht explizit angeben.

Mit den gleichen Methoden wie oben können wir auch lineare Abbildungen studieren. Unter einer *Blockmatrix* verstehen wir eine Matrix, die durch das Neben- und Untereinanderschreiben von Matrizen passender Größe gebildet wird. Seien etwa $A \in M_{p,r}(\mathbb{k})$, $B \in M_{p,s}(\mathbb{k})$, $C \in M_{q,r}(\mathbb{k})$ und $D \in M_{q,s}(\mathbb{k})$, dann ist

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1r} & b_{11} & \dots & b_{1s} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{p1} & \dots & a_{pr} & b_{p1} & \dots & b_{ps} \\ c_{11} & \dots & c_{1r} & d_{11} & \dots & d_{1s} \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{q1} & \dots & c_{qr} & d_{q1} & \dots & d_{qs} \end{pmatrix} \in M_{p+q,r+s}(\mathbb{k}) \;.$$

3.13. Satz (Rangsatz)**.** *Es seien V und W endlich-dimensionale $\mathbb{k}$-Vektorräume, und es sei $F\colon V \to W$ linear. Dann existieren Basen B von W und C von V, so dass die Abbildungsmatrix A von F bezüglich dieser Basen die* Normalform

$$A = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

als Blockmatrix hat, wobei $r = \dim \operatorname{im} F$. Insbesondere gilt die Dimensionsformel

$$\dim \ker F + \dim \operatorname{im} F = \dim V \;.$$

Beweis. Es sei $n = \dim V$ und $r = n - \dim \ker F$. Wir wählen zunächst eine Basis $(c_{r+1}, \dots, c_n)$ von $\ker F$ und ergänzen dann zu einer Basis $(c_1, \dots, c_n)$ von V. Dann ist $U = \langle c_1, \dots, c_r\rangle$ ein Komplement von $\ker F$ in V. Nach dem Homomorphiesatz 2.54 und Proposition 2.58 erhalten wir einen Isomorphismus

$$\begin{tikzcd} U \arrow[r, "\cong"] \arrow[rr, bend right, "F|_U"'] & V/\ker F \arrow[r, "\cong"] & \operatorname{im} F \end{tikzcd}$$

einen Isomorphismus $U \to V/\ker F$. Somit induziert die Basis $(c_1, \dots, c_r)$ von U eine Basis $(b_1, \dots, b_r)$ von $\operatorname{im} F$ mit $b_i = F(c_i)$ für alle $1 \le i \le r$. Schließlich ergänzen wir zu einer Basis $(b_1, \dots, b_m)$ von W. Für die Abbildung F gilt also

$$F(c_j) = \begin{cases} b_j & \text{falls } j \le r, \text{ und} \\ 0 & \text{falls } j > r. \end{cases}$$

Daraus ergibt sich die angegebene Form der Abbildungsmatrix. Außerdem folgt

$$\dim V = \dim \ker F + \dim U = \dim \ker F + \dim \operatorname{im} F \ . \qquad \square$$

3.14. Definition. Es sei $F\colon V \to W$ linear, dann definieren wir den *Rang* von F durch $\operatorname{rg} F = \dim \operatorname{im} F$, falls $\operatorname{im} F$ endlichdimensional ist, ansonsten nennen wir F *von unendlichem Rang.*

Es sei $A \in M_{m,n}(\Bbbk)$ eine Matrix mit den Spalten $a_1, \dots, a_n \in \Bbbk^m$, dann definieren wir den *Spaltenrang* von A durch $\operatorname{rg}_S A = \dim\langle a_1, \dots, a_n\rangle$. Analog definieren wir den *Zeilenrang* von A durch $\operatorname{rg}_Z A = \operatorname{rg}_S(A^t)$.

Da wir eine Matrix $A \in M_{m,n}(\Bbbk)$ auch als lineare Abbildung $A\colon \Bbbk^n \to \Bbbk^m$ auffassen können, ist auch $\operatorname{rg} A$ definiert. Manchmal heißt auch die folgende Proposition „Rangsatz“.

3.15. Proposition. *Es sei $A \in M_{m,n}(\Bbbk)$.*

(1) *Der Rang von A ändert sich nicht, wenn man von links oder rechts mit einer invertierbaren Matrix multipliziert.*
(2) *Es gilt* $\operatorname{rg}_S A = \operatorname{rg} A = \operatorname{rg}_Z A$.

Beweis. Es sei zunächst $B \in GL(m, \Bbbk)$ eine invertierbare Matrix. Die zugehörige lineare Abbildung $B\colon \Bbbk^n \to \Bbbk^n$ ist also ein Automorphismus, insbesondere also bijektiv. Es gilt

$$\operatorname{im}(B \circ A) = \operatorname{im}(B|_{\operatorname{im} A}) \ .$$

Die Abbildung $B|_{\operatorname{im} A}\colon \operatorname{im} A \to \operatorname{im}(B \circ A)$ ist sicherlich immer noch injektiv und linear. Sie ist auch surjektiv, da wir das Bild entsprechend eingeschränkt haben. Somit sind $\operatorname{im} A$ und $\operatorname{im}(B \circ A)$ isomorph, und es folgt

$$\operatorname{rg}(B \circ A) = \dim \operatorname{im}(B \circ A) = \dim \operatorname{im} A = \operatorname{rg} A \ .$$

Sei jetzt $C \in GL(n, \Bbbk)$ invertierbar, insbesondere ist $\operatorname{im} C = \Bbbk^n$. Dann gilt

$$\operatorname{im}(A \circ C) = \operatorname{im}(A|_{\operatorname{im} C}) = \operatorname{im}(A|_{\Bbbk^n}) = \operatorname{im} A \ ,$$

und es folgt

$$\operatorname{rg}(A \circ C) = \dim \operatorname{im}(A \circ C) = \dim \operatorname{im} A = \operatorname{rg} A \ .$$

Damit ist (1) bewiesen.

Es seien wieder $a_1, \dots, a_n \in \Bbbk^m$ die Spalten von A. Dann gilt

$$\langle a_1, \dots, a_n\rangle = \langle A(e_1), \dots, A(e_n)\rangle = \operatorname{im}(A|_{\langle e_1, \dots, e_n\rangle}) = \operatorname{im} A \ ,$$

also auch

$$\operatorname{rg}_S A = \dim\langle a_1, \dots, a_n\rangle = \dim \operatorname{im} A = \operatorname{rg} A \ .$$

Insbesondere ist also auch der Spaltenrang invariant unter Multiplikation mit invertierbaren Matrizen von links oder rechts.

Sei wieder $B \in GL(m, \Bbbk)$ invertierbar, und sei $D \in GL(m, \Bbbk)$ die Inverse. Aus den Übungen wissen wir, wie sich das Matrixprodukt unter Transposition verhält. Insbesondere gilt

$$B^t \cdot D^t = (D \cdot B)^t = E_m^t = E_m = (B \cdot D)^t = D^t \cdot B^t \;,$$

das heißt, die Transponierte B^t ist ebenfalls invertierbar mit Inverser D^t. Seien also $B \in GL(m, \Bbbk)$ und $C \in GL(n, \Bbbk)$, dann gilt für den Zeilenrang

$$\operatorname{rg}_Z(B \cdot A \cdot C) = \operatorname{rg}_S(C^t \cdot A^t \cdot B^t) = \operatorname{rg}_S(A^t) = \operatorname{rg}_Z(A) \;,$$

genau wie für den Rang und den Spaltenrang.

Wir wählen jetzt Basen B von $\Bbbk^m$ und C von $\Bbbk^n$ wie in Satz 3.13 und erhalten

$$\begin{aligned} \operatorname{rg}_S(A) = \operatorname{rg}_S(B^{-1} \cdot A \cdot C) &= \operatorname{rg}_S \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = r \\ &= \operatorname{rg}_Z \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \operatorname{rg}_Z(B^{-1} \cdot A \cdot C) = \operatorname{rg}_Z A \;. \end{aligned}$$

Damit ist auch (2) bewiesen. □

3.16. Folgerung. *Es seien $F\colon V \to W$ und $G\colon X \to Y$ zwei lineare Abbildungen zwischen endlich-dimensionalen $\Bbbk$-Vektorräumen. Dann gibt es genau dann Isomorphismen $P\colon V \to X$ und $\colon W \to Y$, so dass das Diagramm*

(1)
$$\begin{tikzcd} V \arrow[r, "F"] \arrow[d, "P"', "\cong"] & W \arrow[d, "Q", "\cong"'] \\ X \arrow[r, "G"] & Y \end{tikzcd}$$

kommutiert, wenn

(2) $\dim V = \dim X \;,\qquad \dim W = \dim Y \;,\qquad$ *und* $\qquad \operatorname{rg} F = \operatorname{rg} G \;.$

Beweis. Zu „$\Longrightarrow$“ nehmen wir an, dass Isomorphismen P, Q existieren, so dass das Diagramm (1) kommutiert. Dann folgt die Gleichheit der Dimensionen in (2) bereits aus Folgerung 3.8. Außerdem gilt

$$\operatorname{im} G = \operatorname{im}(G \circ P) = \operatorname{im}(Q \circ F) = \operatorname{im}(Q|_{\operatorname{im} F}) \;,$$

und $Q|_{\operatorname{im} F}\colon \operatorname{im} F \to \operatorname{im}(Q|_{\operatorname{im} F}) = \operatorname{im} G$ ist ein Isomorphismus ist. Also folgt

$$\operatorname{rg} G = \dim \operatorname{im} G = \dim \operatorname{im} F = \operatorname{rg} F \;.$$

Zu „$\Longleftarrow$“ nehmen wir an, dass alle Gleichungen in (2) gelten. Dann wählen wir Basen B von W, C von V, D von Y und E von X wie im Rangsatz 3.13, so dass F und G jeweils durch die gleiche Blockmatrix

$$A = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{m,n}(\Bbbk)$$

dargestellt werden, wobei $m = \dim W = \dim Y$, $n = \dim V = \dim X$ und $r = \operatorname{rg} F = \operatorname{rg} G$. Indem wir wieder Basen und Basisabbildungen mit dem gleichen Buchstaben bezeichnen, erhalten wir das kommutative Diagramm

$$\begin{tikzcd}
V \arrow[r, "F"] \arrow[dd, bend right=60, dashed, "P"'] & W \arrow[dd, bend left=60, dashed, "Q"] \\
\Bbbk^n \arrow[u, "C", "\cong"'] \arrow[r, "A"] \arrow[d, "E", "\cong"'] & \Bbbk^m \arrow[u, "B"', "\cong"] \arrow[d, "D"', "\cong"] \\
X \arrow[r, "G"] & Y\,.
\end{tikzcd}$$

Es folgt (1) für $P = E \circ C^{-1}$ und $Q = D \circ B^{-1}$. $\square$

3.17. Bemerkung. Anhand dieser Folgerung können wir gut erklären, was *Normalformen* und *vollständige Invarianten* sind. Wir geben uns eine Klasse von Objekten vor, in unserem Falle lineare Abbildungen zwischen endlich-dimensionalen Vektorräumen. Außerdem sagen wir, wann zwei Objekte „isomorph“ sein sollen, in unserem Falle dann, wenn (1) aus Folgerung 3.16 gilt. Jetzt suchen wir in jeder Isomorphieklasse ein möglichst einfaches Objekt, in unserem Fall die lineare Abbildung $A\colon \Bbbk^n \to \Bbbk^m$ aus dem Rangsatz 3.16. Das heißt, wir bringen eine lineare Abbildung $F\colon V \to W$ „in Normalform“, indem wir die isomorphe Abbildung vom Typ aus Satz 3.13 bestimmen. Dabei kommt es nur darauf an, dass diese Normalform eindeutig bestimmt ist; die benötigten Isomorphismen müssen nicht eindeutig sein. Manchmal ist die Normalform durch eine vollständige Invariante festgelegt, in unserem Fall durch das Tripel

$$(\dim V, \dim W, \dim F) \in \left\{\, (m,n,r) \;\middle|\; m,\ n,\ r \in \mathbb{N} \text{ und } r \le \min(m,n) \,\right\}.$$

Wenn wir den Wertebereich unser Invarianten wie oben vorgeben, existiert zu jedem möglichen Wert der Invarianten genau eine lineare Abbildung in Normalform.

Ein weiteres Beispiel sind endlich-dimensionale $\Bbbk$-Vektorräume V: hier wäre die „Normalform“ der Spaltenraum $\Bbbk^n$; hier ist die vollständige Invariante die Dimension $\dim V \in \mathbb{N}$. Jeder Vektorraum V ist zu einem eindeutigen $\Bbbk^n$ isomorph, und der Isomorphismus ist die Basisabbildung. Nach Proposition 2.77 ist die Basis nicht eindeutig, sondern die Menge aller Basen von V steht in Bijektion zu $GL(n, \Bbbk)$. Das bedeutet insbesondere, dass man nicht sagen kann, welche Spalte $x \in \Bbbk^n$ einem vorgegebenen Vektor v in der Normalform entspricht, da die Koordinaten x von v von der Wahl der Basis abhängen.

Ein noch einfacheres Beispiel ist die Klasse der endlichen Mengen, siehe Definition 1.30. Wir nennen zwei endliche Mengen M und N *gleichmächtig,* falls es eine bijektive Abbildung $f\colon M \to N$ gibt. Als Normalform erhalten wir die Mengen $\underline{n} = \{\underline{0}, \dots, \underline{n-1}\}$ aus Bemerkung 1.29 für $n \in \mathbb{N}$ (man könnte auch die Mengen $\{1, \dots, n\}$ nehmen), und die zugehörige vollständige Invariante ist die Mächtigkeit $\#M$. Die Analogie zwischen Mächtigkeit und Dimension geht relativ weit, beispielsweise gilt für zwei Unterräume U, $W \subset V$ eine ähnliche Formel für $\dim(U+W)$ wie für die Mächtigkeit der Vereinigung zweier endlicher Mengen.

3.3. Lineare Gleichungssysteme

3.18. Definition. Es sei V ein $\Bbbk$-Vektorraum. Eine Teilmenge $A \subset V$ heißt *affiner Unterraum* von V, wenn es einen Untervektorraum $U \subset V$ und ein Element $a_0 \in A$ gibt, so dass

$$A = a_0 + U = \left\{\, a_0 + u \;\middle|\; u \in U \,\right\} .$$

Ein affiner Unterraum $A = a+U$ heißt *endlichdimensional* mit $\dim A = \dim U$, wenn U endlichdimensional ist, sonst *unendlichdimensional.* Seien U, $W \subset V$ Untervektoräume, dann heißen zwei affine Unterräume $a+U$ und $b+W$ *parallel,* wenn $U = W$.

Man beachte, dass in manchen Büchern auch die leere Menge $\emptyset$ als affiner Unterraum der Dimension $\dim \emptyset = -\infty$ betrachtet wird. Wir wollen die leere Menge hier separat betrachten.

3.19. Bemerkung. Ein affiner Unterraum ist also das Bild eines Untervektorraums unter der Verschiebung um a_0.

(1) In der Definition kommt es nicht darauf an, welches $a_0 \in A$ wir wählen. Denn sei $a_1 = a_0 + u_1 \in A$, dann gilt nach dem Unterraumaxiom (U2), dass
$$a_1 + U = a_0 + (u + U) = a_0 + U .$$
(2) Ein affiner Unterraum ist genau dann ein Untervektorraum, wenn $0 \in A$. Die Richtung „$\Longrightarrow$“ folgt aus (U1), und „$\Longleftarrow$“ folgt aus (1), denn aus $0 \in A$ folgt $A = 0 + U = U$ für einen Untervektorraum $U \subset V$. Insbesondere ist jeder Untervektorraum auch ein affiner Unterraum.
(3) Es sei $U \subset V$ ein Untervektorraum. Die Menge aller zu U parallelen affinen Unterräume von V ist gerade der Quotientenraum V/U aus Definition 2.47.
(4) In der Euklidischen Geometrie betrachtet man affine Unterräume des $\mathbb{R}^3$ der Dimensionen 0 (Punkte), 1 (Geraden) und 2 (Ebenen).

Wir kommen zu *linearen Gleichungssystemen*. Gegeben eine Matrix $A \in M_{m,n}(\Bbbk)$, die sogenannte *linke Seite* und einen Vektor $b \in \Bbbk^m$, die *rechte Seite*, suchen wir alle Vektoren $x \in \Bbbk^n$, so dass $A \cdot x = b$. Das heißt, wir suchen die *Lösungsmenge*

$$L = \left\{\, x \in \Bbbk^n \;\middle|\; A \cdot x = b \,\right\} .$$

Wenn wir die Gleichung $A \cdot x = b$ ausschreiben, erhalten wir tatsächlich ein System linearer Gleichungen, nämlich

$$(*) \qquad \begin{array}{ccccccc} a_{11} \cdot x_1 & + & \dots & + & a_{1n} \cdot x_n & = & b_1 \,, \\ \vdots & & & & \vdots & & \vdots \\ a_{m1} \cdot x_1 & + & \dots & + & a_{mn} \cdot x_n & = & b_m \,. \end{array}$$

Wir nennen das Gleichungssystem (*) *homogen,* wenn $b = 0$, und *inhomogen,* wenn $b \neq 0$. Das zu $A \cdot x = b$ gehörige homogene Gleichungssystem ist also $A \cdot x = 0$.

Etwas allgemeiner können wir eine lineare Abbildung $F\colon V \to W$ und eine „rechte Seite“ $w \in W$ betrachten, und nach der „Lösungsmenge“

$$L = \bigl\{\, v \in V \bigm| F(v) = w \,\bigr\} = F^{-1}(\{w\}) ,$$

also dem Urbild von w unter F, fragen. Wenn V und W endlichdimensional sind, können wir Basen wählen und F als Matrix schreiben, und erhalten ein lineares Gleichungssystem vom obigen Typ.

3.20. Bemerkung. Lineare Gleichungssysteme treten zum Beispiel beim Lösen der folgenden Probleme auf.

(1) Betrachte $A \in M_{m,n}(\Bbbk)$, dann ist der Kern $\ker A$ von A gerade die Lösungsmenge des homogenen Gleichungssystems $A \cdot x = 0$.
(2) Sei A wie oben, dann liegt $b \in \Bbbk^m$ genau dann im Bild $\operatorname{im} A$ von A, wenn das Gleichungssystem $A \cdot x = b$ eine Lösung hat.
(3) Es sei $B \in M_n\Bbbk$ eine Basis des $\Bbbk^n$. Um die Koordinaten x eines Vektors $v \in \Bbbk^n$ bezüglich B zu bestimmen, müssen wir nach Bemerkung 2.74 das lineare Gleichungssystem $B \cdot x = v$ lösen. Für Orthonormalbasen geht es einfacher, siehe Proposition 2.82.
(4) Eine quadratische Matrix $A \in M_n(\Bbbk)$ ist genau dann invertierbar, wenn eine Matrix $B \in M_n(\Bbbk)$ mit $A \cdot B = E_n$ existiert (Übung). Um die Spalten $b_1, \dots, b_n$ von B zu bestimmen, müssen wir die n Gleichungssysteme $A \cdot b_i = e_i$ lösen.
(5) Das Bestimmen von Schnittpunkten von Geraden und Ebenen im Euklidischen Raum führt oft auf lineare Gleichungssysteme. Seien etwa eine Gerade G und eine Ebene $E \subset \mathbb{R}^3$ gegeben durch

$$E = \left\{ \begin{pmatrix}2\\0\\0\end{pmatrix} + \begin{pmatrix}1\\-1\\0\end{pmatrix} . r + \begin{pmatrix}1\\0\\-1\end{pmatrix} . s \;\middle|\; r, s \in \mathbb{R} \right\}$$

$$\text{und} \qquad G = \left\{ \begin{pmatrix}3\\2\\1\end{pmatrix} + \begin{pmatrix}2\\1\\1\end{pmatrix} . t \;\middle|\; t \in \mathbb{R} \right\} ,$$

dann bestimmen wir $G \cap E$ durch Lösen des Gleichungssystems

$$\begin{aligned} 2 + r + s &= 3 + 2t \\ -r \phantom{{}+s} &= 2 + t \\ -s &= 1 + t \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} r + s - 2t &= 1 , \\ -r \phantom{{}+s} - t &= 2 , \\ -s - t &= 1 . \end{aligned}$$

Die einzige Lösung dieses Systems ist $r = -1$, $s = 0$, $t = -t$; sie führt auf den einzigen Schnittpunkt

$$\begin{pmatrix}2\\0\\0\end{pmatrix} - \begin{pmatrix}1\\-1\\0\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}3\\2\\1\end{pmatrix} - \begin{pmatrix}2\\1\\1\end{pmatrix} .$$

(6) In der Numerik approximiert man Funktionen, indem man ihre Werte nur an endlich vielen Stützstellen vorgibt. Anschließend nähert man Gleichungen mit zahlreichen unterschiedlichen Operationen (darunter

Multiplikation mit anderen Funktionen und Differentiation) durch lineare Gleichungssysteme in den endlich vielen gesuchten Funktionswerten an und erhält so lineare Gleichungssysteme mit sehr vielen Variablen und Gleichungen. In einer Simulation sind unter Umständen für jeden Zeitschritt mehrere solcher Gleichungssysteme zu lösen. Diese Gleichungssysteme zeichnen sich dadurch aus, dass in jeder Zeile und jeder Spalte der linken Seite A nur sehr wenige Einträge von 0 verschieden sind. Für diese Gleichungssysteme benötigt man schnelle, approximative Lösungsverfahren, die wir hier nicht besprechen werden.

Es folgen einfache, grundätzliche Überlegungen zum Lösungsverhalten linearer Gleichungssysteme.

3.21. Proposition. *Es sei $A \in M_{m,n}(\Bbbk)$ und $b \in \Bbbk^m$.*

(1) *Die Lösungsmenge des homogenen Gleichungssystems $A \cdot x = 0$ ist gerade* $\ker A$.
(2) *Das inhomogene Gleichungssystem $A \cdot x = b$ hat genau dann Lösungen, wenn $b \in \operatorname{im} A$.*
(3) *Es sei $A \cdot x_0 = b$, dann ist die Lösungsmenge des inhomogenen Gleichungssystems $A \cdot x = b$ der affine Unterraum*

$$\left\{\, x \in \Bbbk^n \;\middle|\; A \cdot x = b \,\right\} = x_0 + \ker A \;.$$

Beweis. Die Aussagen (1) und (2) sind gerade die Punkte (1) und (2) aus Bemerkung 3.20. Zu (3) beachten wir, dass aus $A \cdot x_0 = b$ folgt, dass

$$A \cdot x = b \quad\Longleftrightarrow\quad A \cdot (x - x_0) = b - b = 0 \quad\Longleftrightarrow\quad x - x_0 \in \ker A \;. \qquad \square$$

Punkt (3) wird gern so umformuliert: Die *allgemeine Lösung* x des inhomogenen Gleichungssystems $A \cdot x + b$ ist die Summe aus einer *speziellen Lösung* x_0 des inhomogenen Gleichungssystems und der allgemeinen Lösung $v = x - x_0$ des zugehörigen homogenen Gleichungssystems $A \cdot v = 0$.

3.22. Proposition. *Es seien $A \in M_{m,n}(\Bbbk)$ und $b \in \Bbbk^m$. Die Lösungsmenge des linearen Gleichungssystems $A \cdot x = b$ verändert sich nicht, wenn man A und b von links mit der gleichen invertierbaren Matrix $B \in GL(m, \Bbbk)$ multipliziert.*

Beweis. Es sei $x \in \Bbbk^n$ mit $A \cdot x = b$, dann folgt

$$(B \cdot A) \cdot x = B \cdot (A \cdot x) = B \cdot b \;.$$

Gelte umgekehrt $(B \cdot A) \cdot x = B \cdot b$, und sei B^{-1} die Inverse von B, dann folgt

$$A \cdot x = B^{-1} \cdot (B \cdot A) \cdot x = B^{-1} \cdot B \cdot b = b \;.$$

Also haben das alte und das neue Gleichungssystem die gleichen Lösungen. $\square$

3.23. Bemerkung. Wir betrachten jetzt besonders einfache invertierbare Matrizen, die sogenannten *Elementarmatrizen.* Dazu seien $i,\ j \in \{1, \dots, m\}$ mit $i \neq j$ und $k \in \Bbbk^\times = \Bbbk \setminus \{0\}$. Außerdem sei $A \in M_{m,n}(\Bbbk)$.

(1) Wir betrachten die Matrix

$$P_{ij} = \begin{array}{cc} & \begin{array}{ccccccccccc} &&&i&&&&j&&& \end{array} \\ \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} & \begin{pmatrix} 1 & 0 & & & & \cdots & & & & & 0 \\ 0 & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & 0 & & & & 1 & & & \\ & & & & 1 & & & & & & \\ \vdots & & & & & \ddots & & & & & \vdots \\ & & & & & & 1 & & & & \\ & & & 1 & & & & 0 & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & 0 \\ 0 & & & & & \cdots & & & & 0 & 1 \end{pmatrix} \end{array} .$$

Linksmultiplikation mit P_{ij} vertauscht die i-te und die j-te Zeile von A. Diese Matrix ist zu sich selbst invers, also $P_{ij}^{-1} = P_{ij}$.

(2) Als nächstes betrachten wir die Matrix

$$M_i(k) = \begin{array}{cc} & \begin{array}{ccccccc} &&&i&&& \end{array} \\ \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ \end{array} & \begin{pmatrix} 1 & 0 & & \cdots & & & 0 \\ 0 & \ddots & & & & & \\ & & 1 & & & & \\ \vdots & & & k & & & \vdots \\ & & & & 1 & & \\ & & & & & \ddots & 0 \\ 0 & & & \cdots & & 0 & 1 \end{pmatrix} \end{array} .$$

Linksmultiplikation mit $M_i(k)$ multipliziert die i-te Zeile von A mit k und lässt alle anderen Zeilen unverändert. Das Inverse dieser Matrix ist $M_i(k)^{-1} = M_i(k^{-1})$.

(3) Zu guter Letzt betrachten wir

$$E_{ij}(k) = \begin{array}{cc} & \begin{array}{cccc} &&j& \end{array} \\ \begin{array}{c} \\ i \\ \\ \\ \end{array} & \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & k & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \end{array} .$$

Linksmultiplikation mit $M_i(k)$ addiert das k-fache der j-ten Zeile von A zur i-ten Zeile. Das Inverse dieser Matrix ist $E_{ij}(-k)$.

Multiplikation mit einer der obigen Elementarmatrizen P_{ij}, $M_i(k)$ beziehungsweise $E_{ij}(k)$ heißt daher eine *elementare Zeilenumformung* vom Typ (1), (2) oder (3).

3.24. Definition. Eine Matrix $A \in M_{m,n}(\Bbbk)$ heißt *in Zeilenstufenform,* wenn es ganze Zahlen $0 \le r \le m$ und $1 \le j_1 < \cdots < j_r \le n$ gibt, so dass

(1) $a_{ij} = 0$ falls $i > r$ oder falls $i \le r$ und $j < j_i$, und
(2) $a_{ij_i} = 1$ für alle $1 \le i \le r$.

Wir sagen, M sei *in strenger Zeilenstufenform,* wenn darüberhinaus

(3) $a_{ij_p} = 0$ für alle $1 \le p \le r$ und $i \ne p$.

Ein Gleichungssystem $A \cdot x = b$ heißt *in (strenger) Zeilenstufenform,* wenn die Matrix A in (strenger) Zeilenstufenform ist.

Eine Matrix $A = (a_{ij})_{i,j}$ in Zeilenstufenform hat also folgende Gestalt:

$$\begin{array}{cc}
 & \begin{array}{ccccccccccccc} & j_1 & & & & j_2 & & & & j_r & & & \end{array} \\
r & \begin{pmatrix}
0 \cdots 0 & 1 & a_{1,j_1+1} & \cdots & a_{1,j_2-1} & * & a_{1,j_2+1} & \cdots & a_{1,j_r-1} & * & a_{1,j_r-1} & \cdots & a_{1,n} \\
0 & & \cdots & & 0 & 1 & a_{2,j_2+1} & \cdots & a_{2,j_r-1} & * & a_{2,j_r+1} & \cdots & a_{2,n} \\
\vdots & & & & & & & \ddots & \vdots & \vdots & \vdots & & \vdots \\
0 & & & \cdots & & & & & a_{r-1,j_r-1} & * & a_{r-1,j_r+1} & \cdots & a_{r-1,n} \\
0 & & & & \cdots & & & & 0 & 1 & a_{r,j_r+1} & \cdots & a_{r,n} \\
0 & & & & & & \cdots & & & & & & 0 \\
\vdots & & & & & & & & & & & & \vdots \\
0 & & & & & & \cdots & & & & & & 0
\end{pmatrix}
\end{array} .$$

Die „$*$“ sind beliebig, verschwinden aber, wenn A in *strenger* Zeilenstufenform ist. Die Zahlen r und $j_1, \dots, j_r$ sind durch A eindeutig bestimmt. Wir sehen in Proposition 3.26 unten, dass man bei einem Gleichungssystem in Zeilenstufenform die Lösunsmenge leicht ablesen kann.

3.25. Satz (Gauß-Verfahren)**.** *Jedes lineare Gleichungssystem lässt sich mit Hilfe elementarer Zeilenumformungen in (strenge) Zeilenstufenform bringen.*

Andere Namen sind *Gauß-Algorithmus* oder *Gauß-Elimination.*

Beweis. Das Gauß-Verfahren ist ein induktiver Algorithmus, bei man eine Reihe elementarer Zeilenumformungen auf die Matrix A und die rechte Seite b anwendet und so die Matrix A Spalte für Spalte in strenge Zeilenstufenform bringt.

Induktionsannahme. Es seien $r \geq 0$ und $1 \leq j_1 < \dots < j_r \leq n$ sowie q mit $j_r \leq q \leq n$ (beziehungsweise $q \geq 0$, falls $r = 0$) gegeben, so dass die Bedingungen (1) und (2) (beziehungsweise (1)–(3) für strenge Zeilenstufenform) in Definition 3.24 für alle $i \leq n$ und für alle $j \leq q$ gelten. Das heißt, die Matrix A ist bis einschließlich Spalte q bereits in strenger Zeilenstufenform.

Induktionsanfang. Wir beginnen mit $q = r = 0$. Dann sind die obigen Annahmen trivialerweise erfüllt.

Induktionsschritt. Falls $r = m$ oder $q = n$ gilt, sind wir fertig. Ansonsten setzen wir $j = q + 1 \leq n$ und unterscheiden zwei Fälle.

1. Fall: Falls es kein i mit $r < i \leq m$ und $a_{ij} \neq 0$ gibt, ist die Matrix bereits bis zur j-ten Spalte in strenger Zeilenstufenform. In diesem Fall erhöhen wir q um 1, so dass $q = j$, und führen den nächsten Induktionsschritt durch.

2. Fall: Ansonsten gibt es ein kleinstes $i > r$ mit $a_{ij} \neq 0$.

Schritt 1 („Tauschen“): Falls $i \neq r+1$, vertauschen wir die i-te und die $(r+1)$-te Zeile mit einer elementaren Zeilenumformung vom Typ (1). Anschließend erhöhen wir r um 1, so dass jetzt also $a_{rj} \neq 0$.

Schritt 2 („Normieren“): Falls $a_{rj} \neq 1$, multiplizieren wir die r-te Zeile mit a_{rj}^{-1}, so dass anschließend $a_{rj} = 1$, das ist eine elementare Zeilenumformung vom Typ (2). Jetzt setzen wir $j_r = j$, so dass jetzt $a_{rj_r} = 1$, das heißt, Punkt (2) in Definition 3.24 ist für $i = r$ erfüllt.

Schritt 3 („Ausräumen"): Schließlich subtrahieren wir von der i-ten Zeile das a_{ij_r}-fache der r-ten Zeile für alle $i > r$ (beziehungsweise für alle $i \neq r$ für die strenge Zeilenstufenform), das ist eine elementare Zeilenumformung vom Typ (3), so dass hinterher $a_{ij_r} = 0$ für alle $i > r$ (beziehunsweise für alle $i \neq r$). Wir erhöhen q um 1, so dass jetzt $q = j$, und haben nun auch Punkt (1) (und gegebenenfalls auch (3)) in Definition 3.24 für alle $j \leq q$ erfüllt. Anschließend wiederholen wir den Induktionsschritt.

Am Ende erhalten wir eine Matrix in Zeilenstufenform, beziehungsweise in strenger Zeilenstufenform, je nachdem, ob wir in Schritt 3 die gesamte Spalte oder nur unterhalb vom jeweiligen r ausgeräumt haben. □

Man beachte, dass wir in einem Schritt eine ganze Zeile durch a_{rj_r} dividieren mussten, um $a_{rj_r} = 1$ zu erreichen. Aus diesem Grund lässt sich das Gauß-Verfahren nicht auf Matrizen über Ringen anwenden, in denen nicht alle Elemente außer 0 invertierbar sind.

3.26. Proposition. *Sei $A \in M_{m,n}(\Bbbk)$ eine Matrix in Zeilenstufenform, und sei $b \in \Bbbk^m$.*

(1) *Eine Basis des Bildes* $\operatorname{im} A = \Bbbk^r \times \{0\} \subset \Bbbk^m$ *von A besteht aus den Spalten $a_{j_i} = A(e_{j_i})$ für $i = 1, \dots, r$, insbesondere ist* $\operatorname{rg} A = r$.

(2) *Das Gleichungssystem (*) ist genau dann lösbar, wenn $b_{r+1} = \cdots = b_m = 0$; in diesem Fall hat die Lösungsmenge die Gestalt*

$$\begin{aligned}&\left\{\, x \in \Bbbk^n \;\middle|\; A \cdot x = b \,\right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \Bbbk^n \;\middle|\; x_{j_i} = b_i - \sum_{j=j_i+1}^{n} a_{ij} x_j \text{ für alle } i = 1, \dots, r \right\},\end{aligned}$$

jede Lösung ist also eindeutig bestimmt durch die Angabe der Koordinaten x_j für alle $j \in \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$.

(3) *Es sei A in strenger Zeilenstufenform, und es sei $\{k_{r+1}, \dots, k_n\} = \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$ eine Aufzählung der restlichen Spaltenindizes, dann erhalten wir eine Basis $(c_{r+1}, \dots, c_n)$ von* $\ker A$ *aus Vektoren der Form*

$$c_\ell = e_{k_\ell} - \sum_{i=1}^{r} e_{j_i} \,.\, a_{ik_\ell} \in \ker A \subset \Bbbk^n \qquad \text{für } \ell = r+1, \dots, n\,,$$

mit $c_{i\ell} \in \Bbbk$ für alle $i = 1, \dots, r$.

Für die Basis von $\ker A$ in (2) benutzen wir die gleichen Buchstaben wir im Beweis des Rangsatzes 3.13.

Beweis. Zu Aussage (1) überlegen wir uns zunächst, dass $\operatorname{im} A \subset \Bbbk^r \times \{0\} \subset \Bbbk^n$, da alle Spalten von A in diesem Unterraum liegen.

Sei umgekehrt $b \in \Bbbk^r \times \{0\}$, dann hat die Lösungsmenge die in (2) angegebene Gestalt. Wenn wir x_j für $j \in \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$ beliebig vorgeben, bestimmen die Zeilen $i = r, \dots, 1$ in umgekehrter Reihenfolge die fehlenden Koordinaten $x_{j_r}, \dots, x_{j_1}$ eindeutig. Daraus folgt (2) sowie im $A = \Bbbk^r \times \{0\}$ und insbesondere $r = \operatorname{rg} A$, also gilt auch (1).

Zu (3) wählen wir $b = 0$ und bestimmen Elemente c_ℓ der Lösungsmenge $\ker A$, indem wir für $p = r + 1, \dots, n$ die Koordinaten $x_{k_p} = \delta_{\ell p}$ vorgeben. Wenn A in strenger Zeilenstufenform ist, ist die i-te Gleichung äquivalent zu

$$x_{j_i} = -\sum_{p=r+1}^{n} a_{ik_p} \,.\, x_{k_p} = -a_{ik_\ell} \,. \qquad \square$$

Wenn man das Gauß-Verfahren konkret anwendet, schreibt man gern die jeweilige linke Seite als Matrix ohne runde Klammern, macht rechts daneben einen senkrechten Strich, und schreibt die rechte Seite rechts neben diesen Strich. Dann führt man den obigen Algorithmus durch, wobei man sich nur an der linken Seite orientiert, aber alle Zeilenumformungen immer auf die linke und die rechte Seite simultan anwendet. Dabei reicht es, für jeden Induktionsschritt ein neues System aufzuschreiben. Unter Umständen kann es sinnvoll sein, auf der rechten Seite mehr als nur einen Vektor stehen zu haben, zum Beispiel, wenn man ein Gleichungssystem simultan für mehrere rechte Seiten zu lösen hat.

3.27. Bemerkung. Das Gauß-Verfahren kann benutzt werden, um viele verschiedene Probleme zu lösen. Einige davon haben wir in Bemerkung 3.20 bereits angeführt.

(1) Um das Gleichungssystem (*), also $A \cdot x = b$ zu lösen, bringen wir es zunächst mit dem Gauß-Verfahren in Zeilenstufenform. Nach Bemerkung 3.23 entsprechen elementare Zeilenumformungen gerade der Multiplikation mit invertierbaren Matrizen von links. Da wir alle Zeilenumformungen sowohl auf die linke als auch auf die rechte Seite des Gleichungssystems angewandt haben, ist das neue Gleichungssystem nach Proposition 3.22 zum alten äquivalent, und wir können die Lösungsmenge nach Proposition 3.26 (2) ablesen.

    Zur Sicherheit sei daran erinnert, dass man ein Gleichungssystem löst, indem man die *gesamte* Lösungsmenge angibt (eventuell, indem man feststellt, dass diese leer ist), und nicht nur eine einzelnes Element der Lösungsmenge. Wenn die Lösungsmenge nicht leer ist, reicht es allerdings nach Proposition 3.21 (3), eine spezielle Lösung x_0 und den Unterraum $\ker A$ zu bestimmen, da die Lösungsmenge dann gerade $x_0 + \ker A$ ist.

(2) Es sei $A \in M_{m,n}(\Bbbk)$, dann können wir Basen von $\ker A \subset \Bbbk^n$ und $\operatorname{im} A \subset \Bbbk^m$ bestimmen. Wir bringen dazu A mit dem Gauß-Verfahren in strenge Zeilenstufenform. Sei $B \in GL(m, \Bbbk)$ das Produkt der elementaren Zeilenumformungen entsprechenden Elementarmatrizen, so dass $B \cdot A$ in strenger Zeilenstufenform ist. Mit Proposition 3.26 (3) bestimmen wir zunächst eine Basis von $\ker(B \cdot A) = \ker A$.

    Seien r und $j_1, \dots, j_r$ wie in Definition 3.24 zur Matrix $B \cdot A$, dann bilden die Vektoren $(B \cdot A)(e_{j_i})$ für $i = 1, \dots, r$ eine Basis von $\operatorname{im}(B \cdot A)$

nach Proposition 3.26 (1). Da B einen Isomorphismus

$$B|_{\operatorname{im} A}\colon \operatorname{im} A \xrightarrow{\cong} \operatorname{im}(B \circ A)$$

induziert, erhalten wir als Basis von $\operatorname{im} A \subset \Bbbk^m$ gerade

$$\big(A(e_{j_1}), \dots, A(e_{j_r})\big) ,$$

insbesondere gilt $\operatorname{rg} A = \operatorname{rg}(B \cdot A) = r$, siehe auch Proposition 3.15 (1).

Übrigens können wir die Basis $(c_{r+1}, \dots, c_n)$ von $\ker A$ wie im Beweis des Rangsatzes 3.13 zu einer Basis $(c_1, \dots, c_n)$ von $\Bbbk^n$ mit $c_i = e_{j_i}$ für $i = 1, \dots, r$ ergänzen. Wenn wir wie dort fortfahren, erhalten wir ebenfalls die obige Basis $(A(e_{j_1}), \dots, A(e_{j_r}))$ von $\operatorname{im} A$.

(3) Es seien Vektoren $v_1, \dots, v_n \in \Bbbk^m$ gegeben. Wir möchten wissen, ob diese Vektoren linear unabhängig sind, und ob sie $\Bbbk^m$ erzeugen. Dazu schreiben wir die Vektoren als Spalten in eine Matrix A und bringen A in Zeilenstufenform. Dann bilden $(v_1, \dots, v_n)$ genau dann ein Erzeugendensystem, wenn $r = \operatorname{rg} A = m$ gilt.

Und sie sind linear unabhängig, wenn $A \cdot x = 0$ nur eine Lösung besitzt. Nach Proposition 3.26 (2) ist das genau dann der Fall, wenn $\{j_1, \dots, j_r\} = \{1, \dots, n\}$, das heißt, wenn $r = \operatorname{rg} A = n$ gilt.

(4) Um eine Matrix $A \in M_n(\Bbbk)$ zu invertieren, wenden wir das Gauß-Verfahren diesmal mit der rechten Seite E_n an, das heißt, wir lösen n lineare Gleichungssysteme mit der gleichen linken Seite simultan. Wenn wir während des Verfahrens nie eine Spalte überspringen (Fall 1 im Beweis tritt nicht ein) und A in strenge Zeilenstufenform bringen, dann gilt $j_i = i$ für alle $i = 1, \dots, n$. Also bleibt auf der linken Seite die Einheitsmatrix E_n stehen.

Rechts steht das Produkt B aller Elementarmatrizen, die wir im Laufe des Verfahrens angewendet haben, also

$$A \mid E_n \quad \rightsquigarrow \quad E_n \mid B .$$

Es gilt also $B \cdot A = E_n$. Da beide Matrizen quadratisch waren, ist A invertierbar, und B ist die inverse Matrix; dazu interpretiere A und B als lineare Abbildungen und wende eine Übungsaufgabe an.

Falls wir im Laufe des Gauß-Verfahrens eine Spalte überspringen, so dass $j_{i_0+1} > j_{i_0} + 1$ für ein i_0 (oder $j_1 > 1$ für $i_0 = 0$), folgt $i < j_i$ für alle $i > i_0$, insbesondere $r < j_r \leq n$, so dass $\operatorname{rg} A < n$ gilt und A daher nicht invertierbar sein kann. Das bedeutet, dass wir das Verfahren abbrechen können, sobald Fall 1 eintritt, und feststellen können, dass A nicht invertierbar ist.

(5) Für sehr große Matrizen ist das Gauß-Verfahren zu rechenaufwendig. Es gibt aber noch ein anderes Problem, sobald man nicht mit exakten Zahlen rechnet, sondern in jedem Zwischenschritt nach einer bestimmten Anzahl von Dual- oder Dezimalstellen rundet oder abschneidet: Sobald man zwei annähernd gleich große Zahlen mit kleinen prozentualen Fehlern voneinander abzieht, erhält man einen wesentlich größeren prozentualen Fehler im Ergebnis. Um dieses Problem so gut wie

möglich zu umgehen, kann man ein Verfahren anwenden, dass man Pivotisierung nennt. Dabei tauscht man in jedem Schritt 1 die Zeile $r+1$ mit derjenigen Zeile i, für die das Element a_{ij} in der gerade aktuellen Spalte betragsmäßig am größten ist.

3.28. Bemerkung. Die strenge Zeilenstufenform ist wieder eine Normalform. Diesmal betrachten wir als Objekte Matrizen $A \in M_{m,n}(\Bbbk)$ und nennen zwei Objekte A, $A' \in M_{m,n}(\Bbbk)$ „linksäquivalent“, wenn es eine invertierbare Matrix $B \in GL(m, \Bbbk)$ gibt, so dass $A' = B \cdot A$. Mit dem Gauß-Verfahren 3.25 sehen wir, dass jede Matrix zu einer Matrix in strenger Zeilenstufenform linksäquivalent ist.

Mit ein bisschen zusätzlichem Aufwand kann man zeigen, dass zwei Matrizen in strenger Zeilenstufenform genau dann linksäquivalent sind, wenn sie gleich sind. Also gibt es in jeder Linksäquivalenzklasse genau eine Matrix in strenger Zeilenstufenform. Da es von diesen Matrizen offensichtlich sehr viele gibt, erhalten wir keine schöne vollständige Invariante für dieses Problem, außer der besagten Matrix in strenger Zeilenstufenform selbst.

KAPITEL 4

Determinanten

Wir wollen Endomorphismen von Vektorräumen beziehungsweise freien R-Moduln V verstehen, also lineare Abbildungen $F\colon V \to V$. Endomorphismen endlich erzeugter freier Moduln werden durch quadratische Matrizen $A \in M_n(R)$ dargestellt. In diesem Kapitel lernen wir eine wichtige Invariante quadratischer Matrizen kennen, die Determinante.

Über den reellen Zahlen hat die Determinante etwas mit Volumina von Parallelotopen zu tun, und etwas mit Orientierung. Über den meisten anderen Körpern und Ringen lassen sich diese Aspekte nicht voneinander trennen. Wir beginnen in Abschnitt 4.1 mit der Beschreibung von Volumina, benutzen die dort gewonnenen Erkenntnisse in Abschnitt 4.2 zur Definition der Determinante, und führen in Abschnitt 4.3 den Begriff der Orientierung ein.

Im ganzen Kapitel benötigen wir das Kommutativgesetz für die Multiplikation. Insbesondere wird R in diesem Kapitel immer einen kommutativen Ring mit Eins und $\Bbbk$ immer einen Körper bezeichnen. Warum wir das Kommutativgesetz brauchen, erklären wir in Bemerkung 4.5, und was ansonsten schiefgehen kann, sehen Sie in Beispiel 4.22.

4.1. Volumina und Determinantenfunktionen

In Bemerkung 1.69 (2) haben wir die Volumina von Parallelotopen im $\mathbb{R}^3$ ausgerechnet. Im $\mathbb{R}^n$ wollen wir entsprechend das n-dimensionale Volumen

$$\operatorname{vol}(v_1, \dots, v_n)$$

eins von Vektoren $v_1, \dots, v_n \in \mathbb{R}^n$ aufgespannten Parallelotops bestimmen. Wir möchten, dass dieser Volumenbegriff zwei Eigenschaften hat, nämlich *positive Homogenität* und *Scherungsinvarianz:* Für alle n-Tupel $(v_1, \dots, v_n)$, alle i, $j \in \{1, \dots, n\}$ mit $i \neq j$ und alle $k \in \mathbb{R}$ soll gelten

$$\operatorname{vol}(v_1, \dots, v_{i-1}, v_i \,.\, k, v_{i+1}, \dots, v_n) = \operatorname{vol}(v_1, \dots, v_n) \cdot |k| \ ; \tag{1}$$

$$\operatorname{vol}(v_1, \dots, v_{i-1}, v_i + v_j \,.\, k, v_{i+1}, \dots, v_n) = \operatorname{vol}(v_1, \dots, v_n) \ . \tag{2}$$

Bedingung (2) lässt sich mit dem Cavalierischen Prinzip begründen: die Querschnitte von beiden Parallelotopen mit affinen Unterräumen parallel zu $\langle v_1, \dots, \widehat{v_i}, \dots, v_n \rangle$ haben jeweils dasselbe Volumen, wenn man v_i um ein Vielfaches von v_j abändert. Da allgemeine Körper nicht angeordnet sind, ist

Bedingung (1) im allgemeinen nicht sinnvoll. Wir ersetzen sie daher durch eine Art Homogenität und erhalten ein „Volumen mit Vorzeichen“ mit

(1’) $$\omega(v_1, \dots, v_{i-1}, v_i \,.\, k, v_{i+1}, \dots, v_n) = \omega(v_1, \dots, v_n) \cdot k \;.$$

Falls $\Bbbk = \mathbb{R}$ oder $\mathbb{Q}$ und ω die Bedingungen (1’) und (2) erfüllt, erfüllt $\mathrm{vol} = |\omega|$ die Bedingungen (1) und (2) und liefert daher einen Volumen.

Soviel zur Motivation. Wir wollen jetzt Volumina mit Vorzeichen betrachten, und zwar zunächst über kommutativen Ringen R. Wir beginnen mit beliebigen R-Moduln M und Zahlen $k \in \mathbb{N}$.

4.1. Definition. Es sei M ein R-Modul, $k \in \mathbb{N}$, und $\alpha\colon M^k \to R$ eine Abbildung. Dann heißt α *multilinear,* wenn für alle $i \in \{1, \dots, k\}$ und alle v_1, …, v_{i-1}, v_{i+1}, …, $v_k \in M$ die Abbildung

(1) $$M \to R \qquad \text{mit} \qquad w \longmapsto \alpha(v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_k)$$

linear ist. Sie heißt *alternierend* oder auch *alternierende Form,* wenn für alle $i = 1$, …, $k-1$ gilt, dass

(2) $$\alpha(v_1, \dots, v_k) = 0 \qquad \text{falls } v_{i+1} = v_i \;.$$

Die Menge aller alternierenden multilinearen Abbildungen $\alpha\colon M^k \to R$ wird mit $\Lambda^k M^*$ bezeichnet. Falls V ein n-dimensionaler $\Bbbk$-Vektorraum ist, heißt eine alternierende multilineare Abbildung $\omega\colon V^n \to \Bbbk$ eine *Determinantenfunktion*.

Man beachte, dass wir für $k = 1$ gerade den dualen Modul $\Lambda^1 M^* = M^*$ erhalten. Für $k = 0$ setzt man sinnvollerweise $\Lambda^0 M^* = R$.

4.2. Beispiel. Wir betrachten das Spatprodukt $\mathbb{R}^3 \to \mathbb{R}$ mit $(x, y, z) \mapsto \langle x \times y, z\rangle$ aus Satz 1.68. Wegen Bemerkungen 1.52 (1) und 1.67 (1), (1’) ist das Spatprodukt multilinear, und wegen Bemerkung 1.67 (2) und Satz 1.68 (1) ist es alternierend. Also ist das Spatprodukt eine Determinantenfunktion.

4.3. Proposition. *Es sei M ein R-Modul und $\alpha\colon M^k \to R$ multilinear. Dann sind die folgenden Aussagen äquivalent.*

(1) *Die Abbildung α ist alternierend,*
(2) *Es gilt $\alpha(v_1, \dots, v_k) = 0$, wenn $(v_1, \dots, v_k)$ linear abhängig sind.*

Die Aussagen (1) und (2) implizieren die folgende Eigenschaft, die zu (1) und (2) über Körpern $R = \Bbbk$ der Charakteristik $\chi(\Bbbk) \neq 2$ äquivalent ist.

(3) *Die Abbildung α ist* antisymmetrisch, *das heißt, für alle $(v_1, \dots, v_k) \in M^k$ und alle i, $j \in \{1, \dots, k\}$ mit $i < j$ gilt*

$$\alpha(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_k) = -\alpha(v_1, \dots, v_k) \;.$$

Beweis. Die Richtung „(2) $\Longrightarrow$ (1) ist klar, denn falls $v_{i+1} = v_i$, sind $(v_1, \dots, v_k)$ linear abhängig.

Zu „(1) $\Longrightarrow$ (3) betrachten wir zunächst den Fall $j = i + 1$. Dann folgt

$$\alpha(v_1, \dots, v_k) = \alpha(v_1, \dots, v_k) + \underbrace{\alpha(v_1, \dots, v_i, v_i, v_{i+2}, \dots, v_k)}_{=0}$$

$$\begin{aligned}
&= \alpha(v_1, \dots, v_i, v_i + v_{i+1}, v_{i+2}, \dots, v_k) \\
&\qquad - \underbrace{\alpha(v_1, \dots, v_{i-1}, v_i + v_{i+1}, v_i + v_{i+1}, v_{i+2}, \dots, v_k)}_{=0} \\
&= \alpha(v_1, \dots, v_{i-1}, -v_{i+1}, v_i + v_{i+1}, v_{i+2}, \dots, v_k) \\
&\qquad + \underbrace{\alpha(v_1, \dots, v_{i-1}, -v_{i+1}, -v_{i+1}, v_{i+2}, \dots, v_k)}_{=0} \\
&= -\alpha(v_1, \dots, v_{i-1}, v_{i+1}, v_i, v_{i+2}, \dots, v_k) \;.
\end{aligned}$$

Also ändert sich das Vorzeichen, wenn man zwei benachbarte Vektoren vertauscht. Der allgemeine Fall folgt durch Induktion über $p = j - i$, denn

$$\begin{aligned}
\alpha(v_1, \dots, v_k) &= -\alpha(v_1, \dots, v_{i-1}, v_{j-1}, v_{i+1}, \dots, v_{j-2}, v_i, v_j, v_{j+1}, \dots, v_k) \\
&= \alpha(v_1, \dots, v_{i-1}, v_{j-1}, v_{i+1}, \dots, v_{j-2}, v_j, v_i, v_{j+1}, \dots, v_k) \\
&= -\alpha(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-2}, v_{j-1}, v_i, v_{j+1}, \dots, v_k) \;.
\end{aligned}$$

Dabei haben wir nur Argumente im Abstand von weniger als p vertauscht.

Zu „(1) $\Longrightarrow$ (2)“ benutzen wir (3) und zeigen, dass aus (1) Scherungsinvarianz folgt. Für alle i, $j \in \{1, \dots, k\}$ mit $i \neq j$ und alle $r \in R$ gilt, dass

$$\begin{aligned}
\alpha(v_1, \dots, v_k) &= -\alpha(v_1, \dots, v_{i-1}, v_{j-1}, v_{i+1}, \dots, v_{j-2}, v_i, v_j, v_{j+1}, \dots, v_k) \\
&\qquad - \underbrace{\alpha(v_1, \dots, v_{i-1}, v_{j-1}, v_{i+1}, \dots, v_{j-2}, v_j, v_j, v_{j+1}, \dots, v_k)}_{=0} \cdot r \\
&= -\alpha(v_1, \dots, v_{i-1}, v_{j-1}, v_{i+1}, \dots, v_{j-2}, v_i + v_j \,.\, r, v_j, v_{j+1}, \dots, v_k) \\
&= \alpha(v_1, \dots, v_{i-1}, v_i + v_j \,.\, r, v_{i+1}, \dots, v_k) \;,
\end{aligned}$$

insbesondere ist α scherungsinvariant. Seien jetzt die Vektoren $(v_1, \dots, v_k)$ linear abhängig. Nach Lemma 3.1 existiert ein $i \in \{1, \dots, k\}$, so dass v_i als Linearkombination der restlichen Vektoren dargestellt werden kann, also

$$v_i = \sum_{j \neq i} v_j \,.\, r_j$$

mit $r_j \in R$. Nur mit Hilfe der obigen Scherungsinvarianz folgt daraus

$$\begin{aligned}
\alpha(v_1, \dots, v_k) &= \alpha\Bigl(v_1, \dots, v_{i-1}, \sum_{j \neq i} v_j \,.\, r_j, v_{i+1}, \dots, v_k\Bigr) \\
&= \alpha(v_1, \dots, v_{i-1}, 0, v_{i+1}, \dots, v_k) = 0 \;.
\end{aligned}$$

Schließlich sei $R = \Bbbk$ ein Körper der Charakteristik $\chi(\Bbbk) \neq 2$, und es gelte $v_i = v_j$ für i, $j \in \{1, \dots, k\}$ mit $i < j$. Aus (3) folgt

$$\begin{aligned}
\alpha(v_1, \dots, v_k) &= \frac{1}{2}\,\alpha(v_1, \dots, v_k) \\
&\qquad - \frac{1}{2}\,\alpha(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_k) = 0 \;. \qquad \square
\end{aligned}$$

4.4. Bemerkung. Wir haben in der Motivation von einem „Volumen mit Vorzeichen“ Homogenität (1’) und Scherungsinvarianz gefordert. Multilineare Abbildungen sind insbesondere homogen, und im obigen Beweis zu „(1) $\Longrightarrow$ (2)“

haben wir gesehen, dass alternierende multilineare Abbildungen auch scherungsinvariant sind.

Umgekehrt sei V jetzt ein n-dimensionaler $\mathbb{k}$-Vektorraum und $\omega\colon V^n \to \mathbb{k}$ sei homogen und scherungsinvariant. Wir haben im Schritt „(1) $\Longrightarrow$ (2)“ gesehen, dass aus Scherungsinvarianz folgt, dass $\omega(v_1, \dots, v_n) = 0$, wenn $(v_1, \dots, v_n)$ linear abhängig sind.

Wir wollen jetzt zeigen, dass

$$\omega(u + w, v_2, \dots, v_n) = \omega(u, v_2, \dots, v_n) + \omega(w, v_2, \dots, v_n) ,$$

dann ist $\omega(v_1, \dots, v_n)$ linear in v_1. Die Linearität in den anderen Argumenten folgt genauso. Wir unterscheiden zwei Fälle: wenn $(v_2, \dots, v_n)$ linear abhängig sind, reduziert sich die obige Gleichung zu $0 = 0 + 0$.

Wir dürfen also annehmen, dass $(v_2, \dots, v_n)$ linear unabhängig sind, und ergänzen zu einer Basis $(v_1, \dots, v_n)$. Dann existieren k_j, $\ell_j \in \mathbb{k}$, so dass

$$u = \sum_j v_j \,.\, k_j \qquad \text{und} \qquad w = \sum_j v_j \,.\, \ell_j \,.$$

Aus Scherungsinvarianz und Homogenität folgt

$$\begin{aligned}
\omega(u + w, v_2, \dots, v_n) &= \omega\biggl(\sum_j v_j \,.\, (k_j + \ell_j), v_2, \dots, v_n\biggr) \\
&= \omega(v_1 \,.\, (k_1 + \ell_1), v_2, \dots, v_n) \\
&= \omega(v_1, v_2, \dots, v_n) \cdot k_1 + \omega(v_1, v_2, \dots, v_n) \cdot \ell_1 \\
&= \omega\biggl(\sum_j v_j \,.\, k_j, v_2, \dots, v_n\biggr) + \omega\biggl(\sum_j v_j \,.\, \ell_j, v_2, \dots, v_n\biggr) \\
&= \omega(u, v_2, \dots, v_n) + \omega(w, v_2, \dots, v_n) \,.
\end{aligned}$$

Also entsprechen Determinantenfunktionen genau unseren „Volumina mit Vorzeichen.“

4.5. Bemerkung. Wir überlegen uns leicht, dass die Summe zweier Determinantenfunktionen und auch ein skalares Vielfaches einer Determinantenfunktion wieder eine solche ist. Also ist $\Lambda^n M^*$ ein R-Modul. An dieser Stelle braucht man Kommutativität von R, siehe dazu die Bemerkung vor Definition 2.41. Aber man braucht Kommutativität von R bereits, um überhaupt multilineare Abbildungen mit zwei oder mehr Argumenten zu bekommen, wie die folgende Rechnung zeigt:

$$\begin{aligned}
\alpha(v_1, \dots, v_k) \cdot r \cdot s &= \alpha(v_1 \,.\, r, v_2, \dots, v_k) \cdot s = \alpha(v_1 \,.\, r, v_2 \,.\, s, v_3, \dots, v_k) \\
&= \alpha(v_1, v_2 \,.\, s, v_3, \dots, v_k) \cdot r = \alpha(v_1, \dots, v_k) \cdot s \cdot r \,.
\end{aligned}$$

Dass es überhaupt verschiedene Determinantenfunktionen auf demselben Modul oder Vektorraum gibt, sollte uns nicht erstaunen; schließlich kann man auch das Volumen im „uns umgebenden $\mathbb{R}^3$“ verschieden messen — etwa in Litern, Kubikmetern, flüssigen Unzen, Fässern, etc.

Wir wollen jetzt für alle Ringe R (kommutativ, mit Eins) ein spezielles Element $\omega_n \in \Lambda^n(R^n)^*$, die *Standard-Determinantenfunktion,* durch Induktion über $n \in \mathbb{N}$ konstruieren. Für $n = 0$ setzen wir $\omega_0() = 1 \in R$ und sind fertig.

Sei ω_{n-1} bereits konstruiert. Wir fassen Vektoren $x \in R^n$ durch Weglassen der letzten Koordinate als Vektoren $x' \in R^{n-1}$ auf, und nennen die letzte Koordinate $\varepsilon_n(x)$. Wir definieren ω_n rekursiv durch

$$(*) \qquad \omega_n(x_1, \dots, x_n) = \sum_{i=1}^{n} (-1)^{i+n}\, \varepsilon_n(x_i)\, \omega_{n-1}(x'_1, \dots, \widehat{x'_i}, \dots, x'_n) \in R \,,$$

wobei ein Dach über einem Eintrag wie zu Beginn von Abschnitt 3.1 gerade „Weglassen“ bedeutet. Diese Konstruktion liefert zugleich ein erstes Verfahren zur Berechnung von ω_n, die Laplace-Entwicklung, siehe Satz 4.12 unten.

4.6. Proposition. *Es sei R ein kommutativer Ring mit Eins, dann ist die oben konstruierte Abbildung $\omega_n \colon R^n \to R$ alternierend, multilinear, und erfüllt*

$$\omega_n(e_1, \dots, e_n) = 1 \,.$$

Beweis. Wir beweisen die Aussage wieder durch Induktion über n. Für $n = 1$ ist die Behauptung klar.

Sei die Proposition für ω_{n-1} bereits bewiesen. Linearität von ω_n an der i-ten Stelle folgt für den i-ten Summand in (*) aus der Linearität von ε_n, für die restlichen Summanden aus der Multilinearität von ω_{n-1}.

Sei jetzt $x_{i+1} = x_i$ für ein $i \in \{1, \dots, n-1\}$. Dann sind der i-te und der $(i+1)$-te Summand in (*) bis auf das Vorzeichen gleich und heben sich weg, bei allen anderen Summanden werden zwei gleiche Vektoren nebeneinander in ω_{n-1} eingesetzt, was nach Induktionsvoraussetzung 0 ergibt.

Außerdem ist $\varepsilon_n(e_i) = 0$ für $i < n$, und die Vektoren e'_i für $i < n$ sind gerade die Standardbasisvektoren des R^n. Also gilt

$$\omega_n(e_1, \dots, e_n) = (-1)^{n+n}\, \varepsilon_n(e_n)\, \omega_{n-1}(e'_1, \dots, e'_{n-1}) = 1 \,. \qquad \square$$

Als nächstes überlegen wir uns, dass der Raum $\Lambda^n V^*$ genau eindimensional ist.

4.7. Proposition. *Es sei R ein kommutativer Ring mit Eins, $r \in R$ und M ein freier R-Modul mit Basis $B = (b_1, \dots, b_n)$. Dann existiert genau eine Determinantenfunktion $\omega \in \Lambda^n M^*$ mit*

$$\omega(e_1, \dots, e_n) = r \,.$$

Sei ω_B die obige Determinantenfunktion zu $r = 1$, dann ist $\Lambda^b M^*$ ein freier R-Modul mit Basis (ω_B).

Beweis. Zur Eindeutigkeit bestimmen wir den Wert von $\omega(v_1, \dots, v_n)$ für Modulelemente

$$v_j = \sum_{i=1}^{n} b_i \,.\, a_{ij} \in R^n \,.$$

Als erstes schließen wir aus Multilinearität, dass

$$\begin{aligned}\omega(v_1,\dots,v_n) &= \omega\Bigl(\sum_{i=1}^n b_i\,.\,a_{i1},\dots,\sum_{i=1}^n b_i\,.\,a_{in}\Bigr)\\ &= \sum_{i_1=1}^n\cdots\sum_{i_n=1}^n \omega(b_{i_1},\dots,b_{i_n})\cdot a_{i_11}\cdots a_{i_nn}\;.\end{aligned}$$

Als nächstes dürfen wir wegen 4.3 (2) wir alle Summanden weglassen, bei denen $i_j = i_k$ für $j \neq k$, und erhalten

$$\omega(v_1,\dots,v_n) = \sum_{\substack{i_1,\dots,i_n\in\{1,\dots,n\}\\ \{i_1,\dots,i_n\}=\{1,\dots,n\}}} \omega\bigl(b_{i_1},\dots,b_{i_n}\bigr)\cdot a_{i_1,1}\cdots a_{i_n,n}\;.$$

Schließlich geht $\omega\bigl(b_{i_1},\dots,b_{i_n}\bigr)$ nach Proposition 4.11 unten aus $\omega(b_1,\dots,b_n)$ hervor, indem man p-mal einzelne Argumente vertauscht, wobei die Zahl $p = p(i_1,\dots,i_n)$ vom Tupel $(i_1,\dots,i_n)$ abhängt. Wegen Proposition 4.3 (3) gilt

$$\omega\bigl(b_{i_1},\dots,b_{1_n}\bigr) = (-1)^p\cdot\omega(b_1,\dots,b_n) = (-1)^p\,r\;.$$

Also ist ω eindeutig bestimmt durch

$$\omega(v_1,\dots,v_n) = \sum_{\substack{i_1,\dots,i_n\in\{1,\dots,n\}\\ \{i_1,\dots,i_n\}=\{1,\dots,n\}}} (-1)^{p(i_1,\dots,i_n)}\,r\cdot a_{i_1,1}\cdots a_{i_n,n}\;.$$

Es sei $B\colon R^n \to M$ die Basisabbildung, siehe Bemerkung 2.74, und es sei $v_j = B(a_j)$ mit $a_j = (a_{ij})_i \in R^n$. Wir definieren zunächst eine alternierende multilineare Abbildung ω_B mit

$$\omega_B(v_1,\dots,v_n) = \omega_n(a_1,\dots,a_n)\;,\qquad\text{so dass}\qquad \omega_B(b_1,\dots,b_n) = 1\;.$$

Nach Proposition 4.6 ist ω_n multilinear und alternierend, also auch ω_B. Wir erhalten die gesuchte Form ω als

$$\omega = \omega_B\cdot r = \omega_B\,.\,\omega(b_1,\dots,b_n)\;.$$

Aufgrund der obigen Eindeutigkeitsaussage sind alle $\omega \in \Lambda^n M^*$ von dieser Gestalt, also bildet (ω_B) eine Basis. $\square$

Der obige Beweis liefert uns eine zweite Berechnungsmethode der Standard-Determinantenfunktion ω_n, die sogenannte Leibniz-Formel, siehe Satz 4.13 unten.

4.2. Die Determinante

Ausgehend von den Überlegungen im letzten Kapitel führen wir jetzt Determinanten von Endomorphismen und quadratischen Matrizen ein. Während Determinantenfunktionen dazu dienen, Volumina von Parallelotopen in einem Vektorraum zu beschreiben, misst die Determinante, um welchen Faktor ein Endomorphismus das Volumen einzelner Parallelotope vergrößert oder verkleinert.

4.8. Bemerkung. Es seien M, N Moduln über R und $F\colon M \to N$ linear, dann definieren wir für alle k eine Abbildung $F^*\colon \Lambda^k N^* \to \Lambda^k M^*$ durch

$$(F^*(\alpha))(v_1, \dots, v_k) = \alpha\bigl(F(v_1), \dots, F(v_k)\bigr) \tag{1}$$

für alle $\alpha \in \Lambda^k N^*$ und alle $v_1, \dots, v_k$. Die rechte Seite ist sinnvoll, da wir α auf k Elemente $F(v_1), \dots, F(v_k)$ von N anwenden, und entsprechend erhalten wir eine Abbildung $F^*(\alpha)\colon M^k \to R$. Man nennt $F^*(\alpha)$ auch die mit F *zurückgeholte Form.*

Wir zeigen, dass $F^*(\alpha)$ im ersten Argument linear ist; für die anderen Argumente zeigt man Linearität genauso. Es seien x, y und $v_2, \dots, v_k \in M$ und r, $s \in R$, dann folgt

$$\begin{aligned}
&(F^*(\alpha))(x \,.\, r + y \,.\, s, v_2, \dots, v_k)\\
&\qquad = \alpha\bigl(F(x \,.\, r + y \,.\, s), F(v_2), \dots, F(v_k)\bigr)\\
&\qquad = \alpha\bigl(F(x) \,.\, r + F(y) \,.\, s, F(v_2), \dots, F(v_k)\bigr)\\
&\qquad = \alpha\bigl(F(x), F(v_2), \dots, F(v_k)\bigr) \cdot r + \alpha\bigl(F(y), F(v_2), \dots, F(v_k)\bigr) \cdot s\\
&\qquad = (F^*(\alpha))(x, v_2, \dots, v_k) \cdot r + (F^*(\alpha))(y, v_2, \dots, v_k) \cdot s\,.
\end{aligned}$$

Also ist $F^*(\alpha)$ multlinear.

Und $F^*(\alpha)$ ist auch alternierend, denn

$$(F^*(\alpha))(v_1, \dots, v_i, v_i, \dots, v_k) = \alpha\bigl(F(v_1), \dots, F(v_i), F(v_i), \dots, F(v_k)\bigr) = 0\,.$$

Es folgt $F^*(\alpha) \in \Lambda^k M^*$ wie behauptet.

In Bemerkung 4.5 haben wir uns überlegt, dass $\Lambda^k M^*$ und $\Lambda^k N^*$ Moduln über R sind. Die Abbildung $F^*\colon \Lambda^k N^* \to \Lambda^k M^*$ ist linear, denn für alle α, $\beta \in \Lambda^k N^*$, alle r, $s \in R$ und alle $v_1, \dots, v_k \in M$ gilt

$$\begin{aligned}
&\bigl(F^*(\alpha \,.\, r + \beta \,.\, s)\bigr)(v_1, \dots, v_k)\\
&\qquad = (\alpha \,.\, r + \beta \,.\, s)\bigl(F(v_1), \dots, F(v_k)\bigr)\\
&\qquad = \alpha\bigl(F(v_1), \dots, F(v_k)\bigr) \cdot r + \beta\bigl(F(v_1), \dots, F(v_k)\bigr) \cdot s\\
&\qquad = \bigl(F^*(\alpha) \,.\, r + F^*(\beta) \,.\, s\bigr)(v_1, \dots, v_k)\,.
\end{aligned} \tag{2}$$

Schließlich seien $F\colon M \to N$ und $G\colon L \to M$ lineare Abbildungen, dann gilt $(F \circ G)^* = G^* \circ F^*\colon \Lambda^k N^* \to \Lambda^k L^*$, denn

$$\begin{aligned}
\bigl((F \circ G)^*(\alpha)\bigr)(\ell_1, \dots, \ell_k) &= \alpha\bigl(F(G(\ell_1)), \dots, F(G(\ell_k))\bigr)\\
&= (F^*(\alpha))\bigl(G(\ell_1), \dots, G(\ell_k)\bigr) = \bigl(G^*(F^*(\alpha))\bigr)(\ell_1, \dots, \ell_k)\,.
\end{aligned} \tag{3}$$

Es sei M ein freier R-Modul mit Basis $(b_1, \dots, b_n)$. In Proposition 4.7 haben wir gesehen, dass $\Lambda^n V^n$ ein eindimensionaler Vektorraum ist, erzeugt von dem Element ω_B mit $\omega_B(b_1, \dots, b_n) = 1$. Sei jetzt $F \in \operatorname{End}_R(M)$, dann ist $F^* \in \operatorname{End}_R(\Lambda^n M^*)$ nach der obigen Bemerkung, aber $\operatorname{End}_R(\Lambda^n M^*) \cong \operatorname{End}_R(R) = R$, da $\Lambda^n V^* \cong R$. Also existiert zu jedem $F \in \operatorname{End} M$ ein Skalar $a = \det F \in R$, so dass

$$F^*\omega = \omega \,.\, a \qquad \text{für alle } \omega \in \Lambda^n M^* \;.$$

Um a zu bestimmen, wählen wir eine Basis $(b_1, \dots, b_n)$, definieren ω_B wie in Proposition 4.7, und überlegen uns, dass

$$\begin{aligned}\omega_B\bigl(F(b_1), \dots, F(b_n)\bigr) &= (F^*(\omega_B))(b_, \dots, b_n)\\ &= (\omega_B \,.\, a)(b_, \dots, b_n) = (\omega_B)(b_, \dots, b_n) \cdot a = a \;.\end{aligned}$$

Im Spezialfall $M = R^n$ mit der Standardbasis sind die Vektoren $F(e_1)$, …, $F(e_n)$ nach Folgerung 2.75 genau die Spalten der Abbildungsmatrix $A \in M_n(R)$ von F, und ω_B ist gerade die Standarddeterminantenfunktion ω_n aus Proposition 4.6. Das motiviert die folgende Definition.

4.9. Definition. Es sei R ein kommutativer Ring mit Eins, M ein freier R-Modul mit einer n-elementigen Basis, wobei $n \geq 1$, und $F \in \operatorname{End}_R(M)$ ein Endomorphismus. Dann ist die *Determinante* von F der eindeutige Skalar $\det F \in R$, so dass

$$F^*\omega = \omega \,.\, \det F \qquad \text{für alle } \omega \in \Lambda^n M^* \;. \tag{1}$$

Wir definieren die *Determinante* einer Matrix $A \in M_n(R)$ mit den Spalten a_1, …, $a_n \in R^n$ durch

$$\det A = \omega_n(a_1, \dots, a_n) \;. \tag{2}$$

Im Falle $n = 0$ folgt $\det() = 1$, da $\omega_0() = 1$. In Gleichung (1) haben wir für jeden Endomorphismus $F \in \operatorname{End}_R M$ die Determinante definiert, ohne eine Basis fixiert und F als Matrix geschrieben zu haben; diese Definition ist also *basisunabhängig.* Wenn wir eine Basis B wählen und A die Abbildungsmatrix von F bezüglich der Basis B (sowohl vom Definitions- als auch vom Wertebereich) darstellen, ist die Determinante von A durch (2) definiert. Unsere obige Vorüberlegung besagt, dass

$$\det A = \det F \;.$$

Auf diese Weise hängen (1) und (2) zusammen.

Wir wollen jetzt verschiedene Berechnungsverfahren für Determinanten angeben. Zunächst erinnern wir uns an die Automorphismengruppe $\operatorname{Aut}(M)$ einer Menge aus Beispiel 2.5.

4.10. Definition. Es sei $n \in \mathbb{N}$. Die *symmetrische Gruppe S_n in n Elementen* ist definiert als $S_n = \operatorname{Aut}(\{1, \dots, n\})$, ihre Elemente $S_n \ni \sigma \colon \{1, \dots, n\} \to \{1, \dots, n\}$ heißen *Permutationen.*

Es sei R ein kommutativer Ring mit Eins. Einer Permutation $\sigma \in S_n$ ordnen wir die *Permutationsmatrix* $P_\sigma = (\delta_{i,\sigma(j)})_{i,j} \in M_n(R)$ und das *Vorzeichen* oder auch *Signum*

$$\operatorname{sign}(\sigma) = \det(P_\sigma) = \det\bigl((\delta_{i,\sigma(j)})_{i,j}\bigr) \;.$$

Unter einer *Transposition* verstehen wir eine Permutation $\tau \in S_n$, die nur zwei Elemente i und j mit $1 \le i < j \le n$ vertauscht, also

$$\tau(k) = \begin{cases} j & \text{falls } k = i, \\ i & \text{falls } k = j, \text{ und} \\ k & \text{sonst.} \end{cases}$$

Die Permutationsmatrix einer Transposition ist gerade die Elementarmatrix $P_{ij} = P_{\tau_{ij}}$ aus Bemerkung 3.23 (1).

4.11. Proposition. *Jede Permutation $\sigma \in S_n$ kann als Produkt $\sigma = \tau_1 \circ \dots \circ \tau_k$ von Transpositionen $\tau_1, \dots, \tau_k \in S_n$ geschrieben werden, und es gilt*

$$\operatorname{sign}(\sigma) = (-1)^k . \tag{1}$$

Für ρ, $\sigma \in S_n$ gilt

$$\operatorname{sign}(\rho \circ \sigma) = \operatorname{sign}(\rho) \cdot \operatorname{sign}(\sigma) \qquad \textit{und} \qquad \operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma) . \tag{2}$$

Dabei fassen wir die Identität als „leeres Produkt“ mit $k = 0$ auf. Nach (1) ist das Vorzeichen einer Permutation stets 1 oder -1, unabhängig vom Ring R. Allerding könnte $1 = -1$ in R gelten (Beispiel: $R = \mathbb{Z}/2\mathbb{Z}$); in diesem Fall verliefert das Vorzeichen seine Information. Der Beweis unten benutzt keine Determinanten, um Permutationen als Produkte von Transpositionen darzustellen, so dass wir im Beweis von Proposition 4.7 keinen Zirkelschluss erhalten.

Beweis. Wir beweisen die erste Aussage durch Induktion über n. Für $n = 1$ gibt es nur eine Permutation, die Identität, mit

$$\operatorname{sign}(\mathrm{id}_{\{1\}}) = \det(E_1) = 1 .$$

Sei die Aussage für alle $\sigma' \in S_{n-1}$ bewiesen, und sei $\sigma \in S_n$. Falls $\sigma(n) = n$, sei $\sigma' = \sigma|_{\{1,\dots,n-1\}} \in S_{n-1}$. Da σ' ein Produkt von Transpositionen aus S_{n-1}, ist σ das Produkt von Transpositionen aus S_n, die jeweils die gleichen Elemente vertauschen. Falls $\sigma(n) \ne n$, sei τ die Transposition, die $\sigma(n)$ und n vertauscht, so dass

$$(\tau \circ \sigma)(n) = n .$$

Nach dem obigen Argument ist $\tau \circ \sigma$ ein Produkt von Transpositionen $\tau_1 \circ \dots \circ \tau_k$. Da $\tau = \tau^{-1}$, folgt

$$\sigma = \tau \circ \tau_1 \circ \dots \circ \tau_k .$$

Sei $\sigma = \tau_1 \circ \dots \circ \tau_k \in S_n$, dann geht P_σ aus der Einheitsmatrix hervor, indem man nacheinander k-mal je zwei Spalten vertauscht. Aus Proposition 4.3 (3) folgt (1), denn

$$\operatorname{sign}(\sigma) = \det(P_\sigma) = (-1)^k .$$

Zu (2) stellen wir ρ und σ als Produkte von j und k Transpositionen dar, dann erhalten wir eine Darstellung von $\rho \circ \sigma$ als Produkt von $j + k$ Transpositionen, und die erste Behauptung folgt. Die letzte ergibt sich dann aus

$$\operatorname{sign}(\sigma) \cdot \operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma \cdot \sigma^{-1}) = \operatorname{sign}(\mathrm{id}) = 1 . \qquad \square$$

Sei $A = (a_{ij})_{i,j} \in M_n(R)$ eine Matrix, dann bezeichnen wir die Matrix A ohne die i-te Zeile und die j-te Spalte mit

$$A_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{nn} \end{pmatrix} \in M_{n-1}(R) \; .$$

Es folgen zwei Sätze zur Berechnung von Determinanten.

4.12. Satz (Laplace-Entwicklung)**.** *Es sei $A \in M_n(R)$ mit $n \geq 1$.*
Entwicklung nach der i-ten Zeile. *Für alle $i \in \{1, \dots, n\}$ gilt*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A_{ij}) \; . \tag{1}$$

Entwicklung nach der j-ten Spalte. *Für alle $j \in \{1, \dots, n\}$ gilt*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A_{ij}) \; . \tag{2}$$

Beweis. Wir betrachten die durch die rechte Seite von Formel (1) induktiv definierte Abbildung $\omega \colon M_n(R) \to R$ und zeigen wie im Beweis von Proposition 4.6, dass sie multilinear und alterniered in den Spalten von A ist. Aufgrund der Eindeutigkeitsaussage in Proposition 4.7 und Definition 4.9 reicht es zu zeigen, dass die rechte Seite für die Einheitsmatrix den Wert 1 annimmt, um die Behauptung (1) zu beweisen.

Es sei $n \geq 1$ und $i \in \{1, \dots, n\}$, und $\omega(A)$ bezeichne die rechte Seite von (1). Linearität von ω an der k-ten Stelle folgt für den k-ten Summand, da a_{ik} linear von $a_k \in R^n$ abhängt. Für die restlichen Summanden folgt sie, da $\det A_{ij}$ multilinear in den Spalten von A_{ij} ist.

Sei jetzt $a_{k+1} = a_k$ für ein $k \in \{1, \dots, n-1\}$. Dann sind der k-te und der $(k+1)$-te Summand in (1) bis auf das Vorzeichen gleich und heben sich weg; bei allen anderen Summanden stimmen zwei benachbarte Spalten von A_{ij} überein, so dass $\det(A_{ij}) = 0$. Also ist ω multilinear und alternierend.

Für die Einheitsmatrix erhalten wir

$$\omega(E_n) = \sum_{j=1}^{n} (-1)^{i+j} \, \delta_{ij} \cdot \det\bigl((E_n)_{ij}\bigr) = \det(E_{n-1}) = 1 \; ,$$

da nur der Summand mit $i = j$ beiträgt, und da nach Streichen der i-ten Spalte und Zeile aus der Einheitsmatrix E_n die Einheitsmatrix E_{n-1} wird. Damit ist (1) bewiesen.

Wir beweisen (2), indem wir die Transponierte A^t in (1) einsetzen. In Folgerung 4.15 unten zeigen wir, dass $\det A^t = \det A$ ohne Benutzung der Laplace-Entwicklung, so dass dann (2) aus (1) folgt. $\square$

4.13. Satz (Leibniz-Formel). *Für jede Matrix $A \in M_n(R)$ mit $n \geq 1$ gilt*

$$\det A = \sum_{\rho \in S(n)} \operatorname{sign}(\sigma) \cdot \prod_{j=1}^{n} a_{\rho(j),j} = \sum_{\sigma \in S(n)} \operatorname{sign}(\sigma) \cdot \prod_{i=1}^{n} a_{i,\sigma(i)} \, .$$

Beweis. Wir gehen vor wie im Beweis von Proposition 4.7 und erhalten

$$\begin{aligned} \det A = \omega_n(a_1, \dots, a_n) &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \omega_n(e_{i_1}, \dots, e_{i_n}) \cdot a_{i_1 1} \cdots a_{i_n n} \\ &= \sum_{\rho \in S_n} \omega_n(e_{\rho(1)}, \dots, e_{\rho(n)}) \prod_{j=1}^{n} a_{\rho(j),j} = \sum_{\rho \in S_n} \det\bigl((\delta_{i\rho(j)})_{i,j}\bigr) \prod_{j=1}^{n} a_{\rho(j),j} \\ &= \sum_{\rho \in S_n} \operatorname{sign}(\rho) \prod_{j=1}^{n} a_{\rho(j),j} \, . \end{aligned}$$

Dabei haben wir alle Tupel $(i_1, \dots, i_n)$ mit gleichen Einträgen aussortiert und die verbleibenden injektiven (und daher bijektiven) Abbildungen $j \mapsto i_j$ als Permutationen $\rho \in S_n$ aufgefasst.

Es sei schließlich $\sigma = \rho^{-1}$. Nach Proposition 4.11 (2) gilt $\operatorname{sign}(\sigma) = \operatorname{sign}(\rho)$. Indem wir über $i = \rho(j)$ summieren, erhalten wir

$$\det A = \sum_{\rho \in S_n} \operatorname{sign}(\rho) \prod_{j=1}^{n} a_{\rho(j),j} = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \, . \qquad \square$$

4.14. Bemerkung. Permutationen $\sigma \in S_n$ werden oft als $2 \times n$-Matrix

$$\sigma = \begin{pmatrix} 1 & \cdots & n \\ \sigma(1) & \dots & \sigma(n) \end{pmatrix}$$

geschrieben.

Für $n = 2$ gibt es genau zwei Permutationen

$$\mathrm{id} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \qquad \text{und} \qquad \tau = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \, .$$

Da τ eine Transposition ist, gilt $\operatorname{sign}(\tau) = -1$, und es folgt die einfache Formel

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{1,\mathrm{id}(1)} \cdot a_{2,\mathrm{id}(2)} - a_{1,\tau(1)} \cdot a_{2,\tau(2)} = a_{11}\, a_{22} - a_{12}\, a_{21} \, .$$

Für $n = 3$ gibt es schon sechs Permutationen

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$

wobei die erste Reihe Vorzeichen 1 und die zweite Reihe Vorzeichen -1 hat. Hiermit erhalten wir für 3×3-Matrizen die *Sarrussche Regel:*

$$\begin{aligned}
\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} &= \begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \\
&= a_{11}\, a_{22}\, a_{33} + a_{12}\, a_{23}\, a_{31} + a_{13}\, a_{21}\, a_{32} \\
&\qquad - a_{11}\, a_{23}\, a_{32} - a_{12}\, a_{21}\, a_{33} - a_{13}\, a_{22}\, a_{31} \;.
\end{aligned}$$

Hierbei werden die Elemente entlang der drei durchgezogenen Linien jeweils aufmultipliziert und zusammenaddiert, und die Elemente entlang der unterbrochenen Linien werden ebenfalls aufmultipliziert und danach subtrahiert.

Für $n = 4$ gibt es bereits $4! = 24$ Permutationen; zuviele, um die Leibniz-Formel durch ein einprägsames Rechenschema darzustellen.

Zur Berechnung größerer Determinanten ist die Leibniz-Formel nicht zu empfehlen (Übung). Sie erlaubt aber einige interessante Schlussfolgerungen.

4.15. Folgerung. *Es sei R ein kommutativer Ring mit Eins und $A \in M_n(R)$.*

(1) *Es gilt* $\det(A^t) = \det A$.
(2) *Die Determinante* $\det A$ *ist multilinear und alternierend in den Zeilen der Matrix A.*
(3) *Die Determinante* $\det A$ *verschwindet, wenn die Zeilen von A linear abhängig sind.*
(4) *Die Determinante* $\det A$ *ändert sich nicht, wenn man ein Vielfaches einer Zeile zu einer anderen dazuaddiert.*
(5) *Die Determinante* $\det A$ *wechselt das Vorzeichen, wenn man zwei Zeilen vertauscht.*

Beweis. Aussage (1) ergibt sich durch Vergleich der Leibniz-Formeln aus Satz 4.13 für A und A^t, denn

$$\det(A^t) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot \prod_{i=1}^{n} a_{\sigma(i),i} = \det(A) \;.$$

Jetzt folgen (2)–(5) für A jeweils aus Definition 4.1, Proposition 4.3 und Bemerkung 4.4, angewandt auf die Matrix A^t. $\square$

4.16. Definition. Eine Matrix $A = (a_{ij})_{i,j} \in M_n(\Bbbk)$ heißt *in oberer (unterer) Dreiecksgestalt,* oder kurz *obere (untere) Dreiecksmatrix,* wenn $a_{ij} = 0$ für alle i, $j \in \{1, \dots, n\}$ mit $i > j$ $(i < j)$. Eine Matrix heißt in *strikter* oberer/unterer Dreiecksgestalt, wenn zusätzlich $a_{ii} = 0$ für alle $i \in \{1, \dots, n\}$.

Somit ist die linke Matrix unten eine obere Dreiecksmatrix, und die rechte sogar in strikter Dreiecksgestalt:

$$\begin{pmatrix} a_{11} & & \dots & a_{1n} \\ 0 & a_{22} & & \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}, \qquad \begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ & & & a_{n-1,n} \\ 0 & & \dots & 0 \end{pmatrix}.$$

Außerdem erinnern wir uns an Blockmatrizen, siehe Satz 3.13.

4.17. Folgerung. *Es sei R ein kommutativer Ring mit Eins.*

(1) *Seien $A \in M_k(R)$, $B \in M_{k,\ell}(R)$, $C \in M_{\ell,k}(R)$ und $D \in M_{\ell,\ell}(R)$. Dann gilt*

$$\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A) \cdot \det(D) = \det \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}$$

und

$$\det \begin{pmatrix} B & A \\ D & 0 \end{pmatrix} = (-1)^{k\ell} \det(A) \cdot \det(D) = \det \begin{pmatrix} 0 & A \\ D & C \end{pmatrix}.$$

(2) *Es sei A eine obere oder untere Dreiecksmatrix, dann gilt*

$$\det(A) = \prod_{i=1}^{n} a_{ii}.$$

Beweis. Die symmetrische Gruppe $S_{k+\ell}$ enthält eine Teilmenge

$$\begin{aligned} U &= \left\{ \sigma \in S_{k+\ell} \;\middle|\; \sigma(i) \le k \text{ für } i \le k \text{ und } \sigma(i) > k \text{ für } i > k \right\} \\ &= \operatorname{Aut}(\{1, \dots, k\}) \times \operatorname{Aut}(\{k+1, \dots, k+\ell\}) \cong S_k \times S_\ell . \end{aligned}$$

Für $\pi \in S_k$ und $\rho \in S_\ell$ sei $\sigma = (\pi, \rho) \in U$ gegeben durch

$$\sigma(i) = (\pi, \rho)(i) = \begin{cases} \pi(i) & \text{falls } i \le k, \text{ und} \\ \rho(j) + k & \text{falls } i = k + j > k. \end{cases}$$

Schreibt man π und ρ als Produkt von Transpositionen, dann erhält man σ als Produkt all dieser Transpositionen, wobei jede zu ρ Transposition, die eigentlich i und $j \le \ell$ vertauscht, durch eine Transposition ersetzt, die stattdessen $k+i$ und $k + j$ vertauscht. Es folgt

$$\operatorname{sign}(\pi, \rho) = \operatorname{sign}(\pi) \cdot \operatorname{sign}(\rho).$$

Wir bezeichnen die gesamte Matrix mit $M = (m_{ij}) \in M_{k+\ell}(R)$. Wir beweisen die erste Gleichung in (1), das heißt, wir nehmen an, dass $m_{ij} = 0$, falls $j \le k < i$. Sei nun $\sigma \in S_{k+\ell} \setminus U$, dann gibt es entweder ein $i > k$ mit $\sigma(i) \le k$ und in dem zugehörigen Summand der Leibniz-Formel taucht das Element $m_{i,\sigma(i)} = 0$ auf; oder es gibt ein $j \le k$ mit $\sigma(j) > k$, aber in diesem Fall muss es auch ein i wie oben geben, da σ bijektiv ist. Mit dieser und den vorangegangenen Überlegungen lässt sich die Leibniz-Formel aus 4.13 vereinfachen zu

$$\begin{aligned}
\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} &= \sum_{\sigma \in U} \operatorname{sign}(\sigma) \prod_{i=1}^{k+\ell} m_{i,\sigma(i)} \\
&= \sum_{\pi \in S_k} \sum_{\rho \in S_\ell} \operatorname{sign}(\pi)\ \operatorname{sign}(\rho) \prod_{i=1}^{k} m_{i,\sigma(i)} \prod_{j=1}^{\ell} m_{j+k,\rho(j)+k} \\
&= \sum_{\pi \in S_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} a_{i,\sigma(i)} \cdot \sum_{\rho \in S_\ell} \operatorname{sign}(\rho) \prod_{j=1}^{\ell} d_{j,\rho(j)} = \det A \cdot \det D\ .
\end{aligned}$$

Genauso zeigt man die zweite Gleichung in der ersten Zeile von (1). Für die zweite Zeile vertauscht man erst jede der k hinteren Spalten mit jeder der ℓ vorderen, bis die Matrizen wieder die gleiche Gestalt wie in der ersten Zeile haben. Die $k \cdot \ell$ Vertauschungen ergeben den zusätzlichen Faktor $(-1)^{k\ell}$.

Wir beweisen (2) für obere Dreiecksmatrizen durch Induktion. Für $n = 1$ ist die Aussage klar. Wenn wir sie für $n - 1$ bereits bewiesen haben, schreiben wir A als Blockmatrix

$$A = \det \begin{pmatrix} A' & b \\ 0 & a_{nn} \end{pmatrix}\ ,$$

dabei ist $A' \in M_{n-1}(R)$ wieder eine obere Dreiecksmatrix und $b \in R^{n-1}$. Aus (1) folgt, dass

$$\det A = \det \begin{pmatrix} A' & b \\ 0 & a_{nn} \end{pmatrix} = \det A' \cdot a_{nn} = \prod_{i=1}^{n} a_{ii}\ . \qquad \square$$

4.18. Bemerkung. In Folgerung 4.15 haben wir gesehen, wie sich die Determinante unter Zeilenumformungen verhält, also können wir Determinanten jetzt auch mit dem Gauß-Verfahren aus Satz 3.25 berechnen. Wegen Folgerung 4.17 müssen wir unsere Matrix nicht auf strenge Zeilenstufenform bringen; es reicht obere Dreiecksgestalt. In den Übungen sehen Sie, dass das Gauß-Verfahren weniger Rechenaufwand verursacht als Leibniz-Formel und Laplace-Entwicklung, es sei denn, die Matrix enthielte viele Nullen. Hauptnachteil des Gauß-Verfahrens: es funktioniert nur über Körpern.

Wir modifizieren das im Beweis von Satz 3.25 beschriebene Verfahren, angewandt auf eine Matrix A, wie folgt. Wir beginnen mit einem Vorfaktor $a_0 = 1$ und erhalten nach dem r-ten Schritt

$$\det A = \cdots = a_r \cdot \det \begin{array}{c} \\ \\ r \\ \\ \\ \\ \end{array} \begin{pmatrix}
1 & a_{12} & & \cdots & & a_{1,n} \\
0 & \ddots & \ddots & & & \vdots \\
 & \ddots & 1 & a_{r,r+1} & \cdots & a_{r,n} \\
\vdots & & 0 & a_{r+1,r+1} & \cdots & a_{r+1,n} \\
 & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & a_{n,r+1} & \cdots & a_{n,n}
\end{pmatrix}$$

$$= a_r \cdot \det \begin{pmatrix} a_{r+1,r+1} & \cdots & a_{r+1,n} \\ \vdots & & \vdots \\ a_{n,r+1} & \cdots & a_{n,n} \end{pmatrix}$$

Hierbei haben wir Folgerung 4.17 (1) und (2) ausgenutzt und geschlossen, dass nur der untere rechte Block einen Beitrag leistet. Sie müssen also bei einer größeren Matrix gegen Ende des Verfahrens nicht mehr die ganze Matrix mitschleppen.

Falls wir im Laufe des Verfahrens eine Spalte überspringen („1. Fall“ im Beweis von Satz 3.25) ist am Ende des Verfahrens die letzte Zeile 0, folgt $\det A$ aus Folgerung 4.15 (3), und wir können das Gauß-Verfahren an dieser Stelle abbrechen. Genauso sind wir beim Invertieren in Bemerkung 3.27 (4) verfahren.

Ansonsten ändern wir dann, wenn wir im ersten Schritt tauschen müssen, das Vorzeichen der Vorfaktors wegen Folgerung 4.15 (5). Beim Normieren multiplizieren wir den Vorfaktor mit a_{rr} und erhalten unser neues a_r wegen Folgerung 4.15 (2). Anschließend räumen wir unterhalb der aktuellen Zeile aus, wobei sich der Vorfaktor wegen Folgerung 4.15 (4) nicht ändert.

Am Schluss des Verfahrens erhalten wir einen Vorfaktor a_n, multipliziert mit der Determinante einer oberen Dreicksmatrix mit Einsen auf der Diagonalen (also $a_{11} = \cdots = a_{nn} = 1$). Nach Folgerung 4.17 ist diese Determinante 1, also ist a_n die Determinante der ursprünglichen Matrix.

Unsere Definition der Determinante auf dem Umweg über das Zurückziehen von Determinantenfunktionen hat Vorteile: sie ist basisunabhängig und erlaubt es uns, relativ einfach die Multiplikativität der Determinante zu verstehen.

4.19. Satz. *Sei V ein freier R-Modul mit einer n-elementigen Basis, und es seien F, $G \in \operatorname{End} V$, dann gilt*

$$\det(F \circ G) = \det F \cdot \det G\ , \tag{1}$$

d.h., die Determinante ist multiplikativ. *Insbesondere ist* $\det\colon \operatorname{Aut} V \to R^\times$ *ein Gruppen-Homomorphismus. Für Matrizen A, $B \in M_n(R)$ gilt entsprechend*

$$\det(A \cdot B) = \det A \cdot \det B\ . \tag{2}$$

Beweis. Die Multiplikativität von det über einem Körper $\Bbbk$ folgt direkt aus der Kompositionsregel in Bemerkung 4.8 (3), denn für alle $\omega \in \Lambda^n V^*$ gilt

$$\omega \,.\, \det(F \circ G) = (F \circ G)^*\omega = G^* \circ F^*\omega = F^*\omega \,.\, \det G = \omega \,.\, \det G \cdot \det F\ .$$

Indem wir $\omega = \omega_n \neq 0$ wählen, folgt (1). Sei $F \in \operatorname{Aut} V$, dann ist F invertierbar nach Definition 2.41, also existiert eine Umkehrabbildung F^{-1} mit

$$\det F \cdot \det F^{-1} = \det(F \circ F^{-1}) = \det(\mathrm{id}_V) = 1\ .$$

Insbesondere folgt $\det F \in \Bbbk^\times = \Bbbk \setminus \{0\}$ mit $(\det F)^{-1} = \det(F^{-1})$, und außerdem ist $\det\colon \operatorname{Aut} V \to \Bbbk^\times$ ein Gruppenhomomorphismus.

Wir erhalten (2) als Spezialfall für $M = R^n$, da $\operatorname{End}_R M = M_n(R)$. □

In der folgenden Definition verwenden wir die Matrizen A_{ij} aus der Lagrange-Entwicklung, siehe Satz 4.12.

4.20. Definition. Es sei $A \in M_n(R)$. Die *Adjunkte* von A ist definiert als

$$\operatorname{adj} A = \bigl((-1)^{i+j} \det(A_{ji})\bigr)_{i,j} \in M_n(R)\ .$$

Trotz der ähnlichen Namen hat die Adjunkte nichts mit der adjungierten Matrix aus Definition 2.80 zu tun. Die nächste Folgerung ergibt sich aus dem Laplaceschen Entwicklungssatz.

4.21. Folgerung (Cramersche Regeln)**.** *Es sei R ein kommutativer Ring mit Eins. Eine Matrix $A \in M_n(R)$ ist genau dann invertierbar, wenn*

$$\det A \in R^\times = \bigl\{\, r \in R \bigm| \text{es gibt ein } s \in R \text{ mit } rs = 1 \,\bigr\}\ ,$$

und in diesem Fall gilt

$$A^{-1} = (\det A)^{-1} \operatorname{adj} A\ . \tag{1}$$

Wenn $\det A \in R^\times$*, ist das Gleichungssystem $A \cdot x = b$ für alle $b \in R^n$ eindeutig lösbar mit*

$$x_i = \frac{\det A_i}{\det A}\ , \qquad \text{wobei } A_i = (a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n) \in M_n(R)\ . \tag{2}$$

Wir nennen $R^\times$ auch die *Einheitengruppe* von R, und ihre Elemente *Einheiten.*

Beweis. Sei $A'_{ij} \in M_n(\Bbbk)$ diejenige Matrix, die wir erhalten, indem wir in A die i-te Spalte durch eine Kopie der j-ten ersetzen. Außerdem sei $\operatorname{adj} A = (c_{ij})_{i,j}$. Wir berechnen

$$\begin{aligned}\operatorname{adj} A \cdot A &= \Biggl(\sum_{k=1}^n c_{ik} a_{kj}\Biggr)_{i,j} = \Biggl(\sum_{k=1}^n (-1)^{i+k} \det(A_{ki}) \cdot a_{kj}\Biggr)_{i,j} \\ &= (\det A'_{ij})_{i,j} = \det A \cdot E_n\ .\end{aligned}$$

Im letzten Schritt haben wir zum einen ausgenutzt, dass A'_{ij} zwei gleiche Spalten hat und daher $\det A'_{ij} = 0$, falls $i \neq j$. Zum anderen ist $A'_{ii} = A$ für alle i, und die obige Formel folgt aus der Laplace-Entwicklung nach Satz 4.12 (2).

Wenn $A \in M_n(R)$ in R invertierbar ist, folgt $\det A \cdot \det A^{-1} = 1$ aus Satz 4.19 (2), also ist $\det A$ in R invertierbar. Umgekehrt, wenn $\det A$ in R invertierbar ist, existiert nach obiger Rechnung eine Inverse A^{-1} wie in (1).

Zu (2) multiplizieren wir b mit der Inversen A^{-1} aus (1) und erhalten mit der Laplace-Entwicklung nach der i-ten Spalte insbesondere

$$x_i = \det A^{-1} (\operatorname{adj} A \cdot b)_i = \frac{1}{\det A} \sum_{j=1}^n (-1)^{i+j} \det(A_{ji}) \cdot b_j = \frac{\det A_j}{\det A}\ . \qquad \square$$

4.22. Beispiel. Das Inverse einer 2×2-Matrix ist nach der 1. Cramerschen Regel gerade

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} .$$

Bereits für $n \geq 3$ empfiehlt es sich jedoch nicht mehr, Matrizen über Körpern mit der Cramerschen Regel zu invertieren. Das Gauß-Verfahren aus Bemerkung 3.27 (4) ist schneller. Nur über Ringen funktioniert das Gauß-Verfahren in der Regel nicht.

Anhand der obigen Formel sieht man ein Problem mit Determinanten über Schiefkörpern: die quaternionische Matrix $A = \left(\begin{smallmatrix} 1+i & 1+j \\ 1-j & 1-i \end{smallmatrix}\right)$ ist invertierbar (Übung), aber es gilt

$$ad - bc = (1+i)(1-i) - (1+j)(1-j) = 2 - 2 = 0 .$$

4.23. Bemerkung. Mit Hilfe von Determinantenfunktionen können wir (ähnlich wie in Folgerung 3.5 über Schiefkörpern) zeigen, dass alle Basen von endlich erzeugten freien R-Moduln die gleiche Länge haben.

Seien nämlich $B = (b_1, \dots, b_m)$ und $C = (c_1, \dots, c_n)$ Basen, und sei $A \in M_{m,n}(R)$ die Basiswechselmatrix wie in Proposition 2.77, so dass

$$c_j = \sum_{i=1}^{m} b_i \,.\, a_{ij} \qquad \text{für alle } j \in \{1, \dots, n\} .$$

Wir konstruieren ω_C wie in Proposition 4.7, dann gilt

$$1 = \omega_C(c_1, \dots, c_n) = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \omega_C(b_{i_1}, \dots, b_{i_n}) \cdot a_{i_1 1} \cdots a_{i_n n}$$

Wäre $m < n$, so würde auf der rechten Seite mindestens ein Index mehr als einmal vorkommen, also $i_j = i_k$ für $j \neq k$, und die gesamte rechte Seite wäre 0 wegen Proposition 4.3 (2). Da das nicht sein kann, folgt $m \geq n$. Indem wir die Rollen von B und C vertauschen, erhalten wir auch $n \geq m$, und somit schließlich $n = m$.

Also ist die Zahl der Basiselemente eines freien R-Moduls M mit einer endlichen Basis eine Invariante des Moduls M, der *Rang* $\operatorname{rg} M \in \mathbb{N}$ von M. Im Falle eines Vektorraums ist der Rang gerade die Dimension. Für freie R-Moduln mit endlicher Basis bis auf Isomorphie ist der Rang wieder eine vollständige Invariante , das heißt, je zwei freie R-Moduln vom gleichen Rang sind isomorph, und die zugehörige Normalform ist wieder der Modul R^n der Spaltenvektoren.

Als letztes wollen wir die Ableitung der Determinante berechnen.

4.24. Definition. Es sei $A \in M_n(R)$, dann definieren wir die *Spur* von A durch

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii} \in R .$$

4.25. Folgerung. *Es sei* $\Bbbk = \mathbb{R}$ *oder* $\mathbb{C}$ *und* $A\colon (a,b) \to M_n(\Bbbk)$ *eine differenzierbare Abbildung, dann gilt*

$$(\det A)' = \operatorname{tr}(A' \cdot \operatorname{adj} A) = \det A \cdot \operatorname{tr}(A' \cdot A^{-1}) \;.$$

Der letzte Ausdruck ist wegen der Cramerschen Regel 4.21 (1) offensichtlich nur sinnvoll, wenn $\det A \neq 0$.

Beweis. Es sei $t \in (a,b)$. Wir setzen $A = A(t)$ und $B = A'(t)$. Mit Hilfe der Leibniz-Formel aus Satz 4.13 sehen wir, dass die Determinante eine Summe von Produkten ist, die aus jeder Spalte genau einen Matrixeintrag enthalten. Nach der Produktregel müssen wir in jedem Produkt jeden einzelnen Faktor einmal ableiten und mit den anderen Faktoren zusammenmultiplizieren. Sei wieder $\operatorname{adj} A = (c_{ij})_{i,j}$, dann gilt

$$\begin{aligned}(\det A)'(t) &= \sum_{j=1}^n \det((a_1, \dots, a_{j-1}, b_j, a_{j+1}, \dots, a_n)) \\ &= \sum_{i,j=1}^n (-1)^{i+j} b_{ij} \cdot \det(A_{ij}) = \sum_{i,j=1}^n b_{ij}\, c_{ji} = \operatorname{tr}(B \cdot \operatorname{adj} A) \;.\end{aligned}$$

Dabei haben in der zweiten Zeile nach der j-ten Spalte entwickelt und dann die Definition der Adjunkten ausgenutzt. Mit der Cramerschen Regel 4.21 (1) folgt auch die zweite Behauptung. □

4.3. Orientierung reeller Vektorräume

Eine einfache Folgerung aus Satz 4.19 ist die Möglichkeit, endlich erzeugte Vektorräume zu „orientieren". Wir lassen nur Körper $\Bbbk \subset \mathbb{R}$ zu, damit wir vom „Vorzeichen" eines Elements von $\Bbbk$ sprechen können.

4.26. Definition. Sei $\Bbbk \subset \mathbb{R}$ ein Körper, und sei V ein n-dimensionaler $\Bbbk$-Vektorraum. Seien $(x_1, \dots, x_n)$ und $(y_1, \dots, y_n)$ zwei Basen von V mit $y_j = \sum_{i=1}^n a_{ij} x_i$. Dann heißen die Basen *gleich orientiert,* wenn die Basiswechselmatrix $A = (a_{ij})_{i,j} \in \operatorname{End}(\Bbbk^n)$ positive Determinante hat.

4.27. Folgerung. *Sei* V *ein* $\Bbbk$*-Vektorraum der Dimension* $n \geq 1$*. Der Begriff „gleich orientiert" definiert eine Äquivalenzrelation mit zwei Äquivalenzklassen auf der Menge aller Basen von* V*.*

Sei $0 \neq \omega \in \Lambda^n V^*$ *eine Determinantenfunktion, dann bestehen diese Äquivalenzklassen aus allen Basen* $(b_1, \dots, b_n)$ *für die* $\omega(b_1, \dots, b_n) > 0$ *beziehungsweise* $\omega(b_1, \dots, b_n) < 0$ *gilt.*

Beweis. Es seien B, C Basen von V. Wir betrachten den Basiswechsel

$$\begin{tikzcd} & V & \\ \Bbbk^n \arrow[ur, "C"] \arrow[rr, "A"] & & \Bbbk^n \arrow[ul, "B"'] \end{tikzcd} \;.$$

Da Basiswechsel nach Proposition 2.77 invertierbar sind, hat A ein Inverses $A^{-1} \in \operatorname{End}(\Bbbk^n)$. Also folgt $\det A \neq 0$.

Wenn wir $\omega \in \Lambda^n V^* \neq 0$, dann folgt

$$\omega(c_1, \dots, c_n) = (A^*\omega)(b_1, \dots, b_n) = \det A \cdot \omega(b_1, \dots, b_n) .$$

Also sind die Basen B und C genau dann gleich orientiert, wenn $\omega(b_1, \dots, b_n)$ und $\omega(c_1, \dots, c_n)$ das gleiche Vorzeichen haben. Da „hat das gleiche Vorzeichen wie" eine Äquivalenzrelation auf $\Bbbk^\times = \Bbbk \setminus \{0\} \subset \mathbb{R}^\times$ definiert, erhalten wir die gesuchte Äquivalenzrelation auf der Menge aller Basen.

Da es nur zwei mögliche Vorzeichen gibt, finden wir höchstens zwei Äquivalenzklassen. Dass es zwei gibt, sieht man daran, dass $(-b_1, b_2, \dots, b_n)$ und $(b_1, \dots, b_n)$ verschieden orientiert sind. □

4.28. Definition. Sei $\Bbbk \subset \mathbb{R}$ ein Körper. Eine *Orientierung* eines endlich erzeugten $\Bbbk$-Vektorraums V ist eine Äquivalenzklasse gleich orientierter Basen. Sei $\omega \neq 0$ eine Determinantenfunktion, die genau auf dieser Äquivalenzklasse positiv ist, dann heißt ω *positiv* bezüglich der gegebenen Orientierung, und umgekehrt heißt obige Orientierung *durch ω induziert.*

Ein Automorphismus $F \in \operatorname{Aut} V$ heißt *orientierungserhaltend* (*orientierungsumkehrend*), wenn $\det F > 0$ ($\det F < 0$).

Aus dem obigen Beweis folgt, dass die Begriffe „orientierungserhaltend" und „orientierungsumkehrend" nicht von der Wahl einer Orientierung auf V abhängen.

4.29. Beispiel. Auf dem Vektorraum $\mathbb{R}^n$ definieren wir die *Standard-Orientierung* so, dass die Standard-Basis $e_1, \dots, e_n$ positiv orientiert ist. Für die Standard-Determinantenfunktion gilt

$$\omega_n(e_1, \dots, e_n) = 1 > 0 ,$$

also ist sie positiv bezüglich der Standard-Orientierung.

In Bemerkung 1.69 haben wir eine geometrische Interpretation des Kreuz- und des Spatproduktes gegeben. Nur das Vorzeichen hatten wir nicht klären können. Mit Hilfe der Sarrusschen Regel können wir nachrechnen, dass

$$\omega_3(u, v, w) = \langle u \times v, w \rangle$$

gilt. Also ist das Spatprodukt nach 1.69 (2) die (eindeutige) positive Determinantenfunktion, deren Absolutbetrag das Volumen von Parallelotopen angibt. Da

$$\omega_3(u, v, u \times v) = \|u \times v\|^2 \geq 0$$

gilt, ist das Kreuzprodukt $u \times v$ nach 1.69 (1) der (eindeutige) Vektor im $\mathbb{R}^3$, der senkrecht auf u und v steht, dessen Länge den Flächeninhalt des von u und v aufgespannten Parallelogramms angibt, und der (falls u und v nicht linear abhängig sind) mit u und v eine positiv orientierte Basis des $\mathbb{R}^3$ bildet.

4.30. Bemerkung. In den Übungen haben Sie die *orthogonale Gruppe* $O(n)$ der linearen Isometrien des $\mathbb{R}^n$ kennengelernt, das heißt, der linearen Abbildungen, die das Standardskalarprodukt $\langle\,\cdot\,,\,\cdot\,\rangle$ aus Definition 1.51 erhalten. Für alle $A \in O(n)$ gilt $\det A \in \{\pm 1\}$, da

$$O(n) = \left\{ A \in M_n(\mathbb{R}) \;\middle|\; A^t \cdot A = E_n \right\} ,$$

siehe auch Proposition 2.82. Außerdem haben wir die *spezielle orthogonale Gruppe* $SO(n)$ der Elemente $A \in O(n)$ mit $\det A = 1$ definiert, das ist also die Untergruppe der orientierungserhaltenden Isometrien.

Genauso haben wir die Untergruppe $SL(n, \mathbb{R}) \subset GL(n, \mathbb{R})$ der Elemente mit Determinante 1 kennengelernt. Wir betrachten zunächst die Gruppe

$$GL(n, \mathbb{R})^+ = \left\{ A \in GL(n, \mathbb{R}) \;\middle|\; \det A > 0 \right\} \subset GL(n, \mathbb{R})$$

der orientierungserhaltenden Automorphismen. Als nächstes gibt es auch eine Untergruppe

$$\left\{ A \in GL(n, \mathbb{R}) \;\middle|\; |\det A| = 1 \right\} \subset GL(n, \mathbb{R})$$

der *volumenerhaltenden Automorphismen.* Dabei erinnern wir uns daran, dass das Volumen durch den Absolutbetrag einer Determinantenfunktion gemessen wird, siehe dazu den Beginn von Abschnitt 4.1. Der Durchschnitt der beiden obigen Untergruppen ist genau $SL(n, \mathbb{R})$, somit ist $SL(n, \mathbb{R})$ die Gruppe der orientierungs- und volumenerhaltenden Automorphismen des $\mathbb{R}^n$. Wie bereits am Anfang von Abschnitt 4.1 gesagt, ist über anderen Körpern wie $\mathbb{C}$ oder $\mathbb{Z}/p\mathbb{Z}$ nicht möglich, Volumina „ohne Vorzeichen“ zu erklären. Aus dem gleichen Grund ist von den obigen Untergruppen der $GL(n, \mathbb{k})$ nur $SL(n, \mathbb{k})$ für alle $\mathbb{k}$ sinnvoll definiert.

KAPITEL 5

Eigenwerte

Wir versuchen, Endomorphismen durch möglichst einfache Matrizen darzustellen, im Idealfall durch Diagonalmatrizen. Dazu studieren wir Eigenwerte und Eigenvektoren. Wir lernen feinere Invarianten von Endomorphismen endlich-dimensionaler Vektorräume kennen, das charakteristische und das Minimalpolynom. Beide helfen, Eigenwerte zu finden und die Struktur von Endomorphismen besser zu verstehen. Allgemeine Struktursätze beweisen wir im nächsten Kapitel. Wir benötigen nach wie vor das Kommutativgesetz für die Multiplikation, und arbeiten daher meist über Körpern oder über kommutativen Ringen mit Eins.

5.1. Eigenvektoren

5.1. Definition. Es sei V ein $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$ ein Endomorphismus. Ein Vektor $v \in V$ heißt *Eigenvektor* von F zum Eigenwert $\lambda \in \Bbbk$, wenn $v \neq 0$ und $F(v) = v \,.\, \lambda$. Sei $\lambda \in \Bbbk$, dann heißt die Menge

$$V_\lambda = \left\{\, v \in V \;\middle|\; F(v) = v \,.\, \lambda \,\right\}$$

Eigenraum von F zum Eigenwert λ. Ein Element $\lambda \in \Bbbk$ heißt *Eigenwert* von F, wenn es einen Eigenvektor $v \in V \setminus \{0\}$ zum Eigenwert λ gibt, das heißt, wenn $V_\lambda \neq \{0\}$. Genauso definieren wir Eigenvektoren, Eigenräume und Eigenwerte quadratischer Matrizen über $\Bbbk$.

Zum Beispiel betrachte

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \in M_2(\mathbb{Q}) \;, \qquad v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{Q}^2 \qquad \text{und} \qquad w = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{Q}^2 \;,$$

dann ist v ein Eigenvektor von A zum Eigenwert 3 und w ein Eigenvektor zum Eigenwert 1, wie man leicht nachrechnet.

Auch Endomorphismen unendlich-dimensionaler Vektorräume können Eigenvektoren haben. Sei etwa $C^\infty(\mathbb{R})$ der Raum der unendlich oft differenzierbaren reellwertigen Funktionen auf $\mathbb{R}$, dann ist die Ableitung $\frac{d}{dx}$ ein Endomorphismus von $C^\infty(\mathbb{R})$, und jedes $\lambda \in \mathbb{R}$ ist Eigenwert; die zugehörigen Eigenvektoren („Eigenfunktionen") haben die Form

$$x \mapsto c\, e^{\lambda x} \qquad \text{mit } c \neq 0 \;.$$

Der Eigenraum V_λ besteht somit aus dem Nullvektor und allen Eigenvektoren zu λ. Man beachte, dass der Nullvektor als Element des Eigenraumes zugelassen ist, aber selbst nicht als Eigenvektor betrachtet wird. Das liegt daran, dass die Gleichung $F(0) = 0 \,.\, \lambda$ für alle λ und alle F erfüllt ist und somit keine Information über F und λ enthält.

5.2. Bemerkung. Da $\Bbbk$ ein Körper ist, gilt $v \,.\, \lambda = (\lambda \cdot \mathrm{id}_V)(v)$, so dass wir den Eigenraum V_λ auch schreiben können als

$$V_\lambda = \bigl\{\, v \in V \bigm| (F - \lambda \cdot \mathrm{id}_V)(v) = 0 \,\bigr\} = \ker(F - \lambda \cdot \mathrm{id}_V)\;.$$

Insbesondere ist $V_\lambda \subset V$ ein Unterraum nach Proposition 2.52 (1).

Wir nehmen jetzt an, dass V endlich-dimensional ist. Um den Eigenraum zu einem vorgegebenen $\lambda \in \Bbbk$ zu berechnen, müssen wir also nur eine Basis B von V wählen, die Abbildungsmatrix A von F bezüglich der Basis B (sowohl für den Definitions- als auch für den Wertebereich V) bestimmen, und dann das Gleichungssystem

$$(A - \lambda\, E_n)(x) = 0$$

lösen. Die Lösungsmenge liefert genau die B-Koordinaten der Elemente des Eigenraums V_λ.

Wir könnten Eigenvektoren auch für Endomorphismen von Moduln über kommutativen Ringen definieren. Wenn wir auf die Kommutativität der Multiplikation verzichten, ist es sinnvoller, anstelle von Eigenräumen eindimensionale Unterräume $U \subset V$ mit $F(U) \subset U$ zu betrachten (Übung).

Wenn ein Endomorphismus $F \in \mathrm{End}_\Bbbk V$ genug Eigenvektoren hat, können wir unter Umständen eine Basis von V finden, bezüglich der F eine besonders einfache Abbildungsmatrix hat. Wir erinnern uns an den Begriff der Dreieckmatrix aus Definition 4.16.

5.3. Definition. Eine *Diagonalmatrix* ist eine quadratische Matrix $A = (a_{ij})_{i,j} \in M_n(R)$, so dass $a_{ij} = 0$ für alle $i, j \in \{1, \dots, n\}$ mit $i \neq j$.

Sei V ein n-dimensionaler $\Bbbk$-Vektorraum. Ein Endomorphismus $F \in \mathrm{End}_\Bbbk V$ heißt *diagonalisierbar,* wenn es eine Basis von V gibt, bezüglich der die Abbildungsmatrix von F eine Diagonalmatrix ist. Wir nennen F *trigonalisierbar,* wenn es eine Basis gibt, bezüglich der die Abbildungsmatrix von F eine Dreiecksmatrix ist.

Eine Matrix $A \in M_n(R)$ heißt trigonalisierbar (diagonalisierbar), wenn es eine invertierbare Matrix $G \in GL(n, R)$ gibt, so dass $G^{-1} \cdot A \cdot G$ eine Dreiecks- (Diagonal-) matrix ist.

Eine typische Diagonalmatrix hat also die Gestalt

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} \in M_n(R)\;.$$

Ein Endomorphismus F ist diagonalisierbar (trigonalisierbar), wenn es eine Basis B mit Basisabbildung $B\colon \Bbbk^n \to V$ und eine Diagonalmatrix (obere Dreiecksmatrix) $A \in M_n(\Bbbk)$ gibt, so dass das folgende Diagramm kommutiert:

$$\begin{tikzcd} V \arrow[r, "F"] & V \\ \Bbbk^n \arrow[u, "B"] \arrow[r, "A"] & \Bbbk^n. \arrow[u, "B"'] \end{tikzcd}$$

Da man mit Diagonalmatrizen besonders einfach rechnen kann, wäre es schön, wenn jeder Endomorphismus diagonalisierbar wäre. Das ist aber leider nicht der Fall.

Da jede Diagonalmatrix insbesondere eine Dreiecksmatrix ist, folgt trigonalisierbar aus diagonalisierbar. Übrigens spielt es keine Rolle, ob wir Trigonalisierbarkeit mit oberen oder mit unteren Dreiecksmatrizen definieren: sei die Abbildungsmatrix $A = (a_{ij})_{i,j}$ von F bezüglich $B = (b_1, \dots, b_n)$ eine obere Dreiecksmatrix, dann ist die Abbildungsmatrix $(a_{n+1-i,n+1-j})_{i,j}$ von F bezüglich der Basis $(b_n, \dots, b_1)$ eine untere Dreiecksmatrix, und umgekehrt.

5.4. Proposition. *Es sei V ein $\Bbbk$-Vektorraum mit Basis $B = (b_1, \dots, b_n)$, und sei $F \in \operatorname{End}_{\Bbbk} V$ ein Endomorphismus mit Abbildungsmatrix A bezüglich B. Dann ist b_j genau dann ein Eigenvektor von F zum Eigenwert λ_j, wenn die j-te Spalte von A gerade $e_j \,.\, \lambda$ ist. Insbesondere ist F genau dann diagonalisierbar, wenn es eine Basis aus Eigenvektoren gibt.*

Beweis. Die j-te Spalte von A enthält die B-Koordinaten von $F(b_j)$ nach Folgerung 2.75. Also ist b_j genau dann ein Eigenvektor zum Eigenwert λ_j, wenn $F(b_j) = b_j \,.\, \lambda_j$, und somit die j-te Spalte a_j von A die Zahl λ_j an der j-ten Stelle und sonst nur Nullen enthält, das heißt, wenn $a_j = e_j \,.\, \lambda_j$. Es folgt die erste Behauptung. Die zweite ist jetzt offensichtlich. □

Um eine Basis aus Eigenvektoren zu bekommen, brauchen wir also eine Familie von Eigenvektoren, die linear unabhängig sind und V erzeugen. Lineare Unabhängigkeit garantiert uns in Spezialfällen die folgende Überlegung.

5.5. Proposition. *Eigenvektoren eines Endomorphismus zu verschiedenen Eigenwerten sind linear unabhängig.*

Beweis. Es sei V ein $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$ ein Endomorphismus. Es sei $(v_i)_{i\in I}$ eine Familie in $V \setminus \{0\}$ und $(\lambda_i)_{i\in I}$ eine Familie in $\Bbbk$ mit $\lambda_i \neq \lambda_j$ für alle i, $j \in I$ mit $i \neq j$, so dass v_i jeweils Eigenvektor zum Eigenwert λ_i ist. Zu zeigen ist, dass die Familie $(v_i)_{i\in I}$ linear unabhängig ist.

Wir beginnen mit dem Fall $I = \{1, \dots, k\}$, das heißt, wir betrachten nur endlich viele Eigenvektoren. In diesem Fall verläuft der Beweis durch vollständige Induktion über k. Im Fall $k = 0$ ist nichts zu zeigen.

Sei $k \geq 1$ und seien $a_i \in \Bbbk$ gegeben, so dass

$$0 = \sum_{i=1}^{k} v_i \,.\, a_i \;.$$

Dann dürfen wir den Endomorphismus $F - \lambda_k \,\mathrm{id}_V$ anwenden und erhalten

$$\begin{aligned} 0 &= (F - \lambda_k \,\mathrm{id}_V)\Bigl(\sum_{i=1}^{k} v_i \,.\, a_i\Bigr) = \sum_{i=1}^{k} v_i \,.\, (\lambda_i - \lambda_k) \,.\, a_i \\ &= \sum_{i=1}^{k-1} v_i \,.\, \bigl(\underbrace{(\lambda_i - \lambda_k)}_{\neq 0} \cdot a_i\bigr) \;. \end{aligned}$$

Nach Induktionsvoraussetzung verschwinden die Zahlen $(\lambda_i - \lambda_k) \cdot a_i \in \Bbbk$, es folgt $a_i = 0$ für $i = 1, \dots, k-1$. Aus der ursprünglichen Gleichung wird also

$$0 = v_k \,.\, a_k \;,$$

und da $v_k \neq 0$ folgt $a_k = 0$. Also sind $(v_1, \dots, v_k)$ linear unabhängig.

Es bleibt der Fall einer unendlichen Indexmenge. Nach Definition (2.28) müssen wir Linearkombinationen

$$0 = \sum_{i \in I} v_i \,.\, a_i$$

betrachten, bei denen $a_i = 0$ für fast alle $i \in I$ gilt. Also bleibt eine endliche Linearkombination stehen, und das obige Argument zeigt wieder, dass $a_i = 0$ für alle $i \in I$. □

Wir erinnern uns an die direkte Summe von Unterräumen aus Abschnitt 2.4 (vor Definition 2.61). Insbesondere heißt eine Summe von Unterräumen U_i für $i \in I$ direkt, wenn wenn $U_i \cap \sum_{j \in I, j \neq i} U_j = \{0\}$ für alle $i \in I$.

5.6. Folgerung. *Es sei V ein n-dimensionaler $\Bbbk$-Vektorraum und $F \in \mathrm{End}_{\Bbbk} V$ ein Endomorphismus.*

(1) *Dann hat F höchstens n verschiedene Eigenwerte.*
(2) *Wenn es n verschiedene Eigenwerte gibt, besitzt V eine Basis aus Eigenvektoren; somit ist F dann diagonalisierbar.*
(3) *Seien $V_{\lambda_1}, \dots, V_{\lambda_k}$ Eigenräume von F zu verschiedenen Eigenwerten von F, dann gilt*

$$V_{\lambda_1} + \cdots + V_{\lambda_k} = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k} \subset V \;.$$

(4) *Der Endomorphismus F ist genau dann diagonalisierbar, wenn V eine direkte Summe aus Eigenräumen von F ist.*

Beweis. Nach dem Basisaustauschsatz 3.4 kann es kein Tupel aus mehr als n linear unabhängigen Vektoren geben, und (1) folgt aus Proposition 5.5.

Seien $v_1, \dots, v_n$ Eigenvektoren, dann sind sie linear unabhängig nach Proposition 5.5. Nach einer Übung bilden je n linear unabhängige Vektoren eine

Basis eines n-dimensionalen Vektorraums. Nach Proposition 5.4 ist F also diagonalisierbar, und es folgt (2).

Seien jetzt $V_{\lambda_1}, \dots, V_{\lambda_k}$ Eigenräume zu verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_k$ von F. Zu zeigen ist

$$V_{\lambda_i} \cap \sum_{j \neq i} V_{\lambda_j} = \{0\}$$

für alle $i = 1, \dots, k$. Sei also

$$v_i = \sum_{j \neq i} v_j \in V_{\lambda_i} \cap \sum_{j \neq i} V_{\lambda_j} \qquad \text{mit} \qquad v_j \in V_{\lambda_j} \text{ für alle } j \;,$$

dann gilt

$$0 = v_i - \sum_{j \neq i} v_j \in V \;.$$

Wäre $v_i \neq 0$, so ergäbe dies eine lineare Abhängigkeit zwischen Eigenvektoren zu verschiedenen Eigenwerten (dabei betrachten wir nur diejenigen $j \in \{1, \dots, k\}$ mit $v_j \neq 0$), im Widerspruch zu Proposition 5.5. Also ist die Summe direkt, und es folgt (3).

Als nächstes beweisen wir (4), „$\Longrightarrow$“. Sei also $B = (b_1, \dots, b_n)$ eine Basis von V, so dass F bezüglich B durch eine Diagonalmatrix A dargestellt wird. Nach Proposition 5.5 ist jeder Basisvektor ein Eigenvektor. Seien $\lambda_1, \dots, \lambda_k$ die Eigenwerte von F. Zu jedem Eigenwert λ_j von F sei $U_{\lambda_j} \subset V_{\lambda_j}$ der Unterraum, der von den Basisvektoren b_i mit $F(b_i) = b_i \,.\, \lambda_j$ aufgespannt wird. Dann folgt aus (3) und der Tatsache, dass B eine Basis ist und jeder Basisvektor in einem U_{λ_j} vorkommt, dass

$$V = \sum_{j=1}^{k} U_{\lambda_j} \subset \bigoplus_{j=1}^{k} V_{\lambda_j} \subset V \;.$$

Daher gilt „$=$“ anstelle von „$\subset$“ in der obigen Ungleichung, insbesondere ist V die direkte Summe der Eigenräume von F.

Sei zu (4), „$\Longleftarrow$“, schließlich $V = \sum_{i=1}^{n} V_{\lambda_i}$, dann ist die Summe direkt nach (3). Wir wählen Basen $(b_1^i, \dots, b_{r_i}^i)$ von V_{λ_i}. Nach Voraussetzung bildet das Tupel $(b_1^1, b_2^1, \dots, b_{r_1}^1, b_1^2, \dots, b_{r_k}^k)$ ein Erzeugendensystem von V. Da die Summe direkt ist, bildet es sogar eine Basis. Denn seien jetzt $a_j^i \in \Bbbk$, so dass

$$0 = \sum_{i=1}^{k} \sum_{j=1}^{r_i} b_j^i \,.\, a_j^i \;.$$

Für jedes $\ell \in \{1, \dots, k\}$ folgt daraus

$$\sum_{j=1}^{r_\ell} b_j^\ell \,.\, a_j^\ell = - \sum_{\substack{i=1 \\ i \neq \ell}}^{k} \sum_{j=1}^{r_i} b_j^i \,.\, a_j^i \;.$$

Da die Summe direkt ist, sind beide Seiten 0. Da $(b_1^\ell, \dots, b_{r_\ell}^\ell)$ als Basis von V_{λ_ℓ} linear unabhängig ist, gilt $a_1^\ell = \dots = a_{r_\ell}^\ell = 0$. Insgesamt ist das Tupel $(b_1^1, b_2^1, \dots, b_{r_1}^1, b_1^2, \dots, b_{r_k}^k)$ eine Basis aus Eigenvektoren, also ist F diagonalisierbar. □

5.7. Folgerung. *Es seien V und W zwei n-dimensionale $\Bbbk$-Vektorräume, $F \in \operatorname{End}_\Bbbk V$ und $G \in \operatorname{End}_\Bbbk W$. Außerdem seien $\lambda_1, \dots, \lambda_n \in \Bbbk$ paarweise verschiedene Eigenvektoren von F. Dann existiert genau dann ein Isomorphismus $P\colon V \to W$ mit $P \circ F = G \circ P$, wenn G die gleichen Eigenwerte wie F hat.*

Beweis. Übung. □

5.8. Bemerkung. Als Fazit dieses Abschnitts halten wir fest, dass es sich lohnt, Eigenwerte und Eigenvektoren von Endomorphismen zu bestimmen, um eine Abbildungsmatrix in möglichst einfacher Form zu finden. Wir werden dieses Ziel später noch weiter verfolgen.

Außerdem zeigt Folgerung 5.7, dass Eigenwerte etwas mit dem „Normalformproblem“ für Endomorphismen zu tun haben. Hierbei nennen wir zwei Endomorphismen $F \in \operatorname{End}_\Bbbk V$ und $G \in \operatorname{End}_\Bbbk W$ *isomorph,* wenn ein Isomorphismus $P\colon V \to W$ existiert, so dass das Diagramm

$$\begin{tikzcd} V \arrow[r, "P"] \arrow[d, "F"'] & W \arrow[d, "G"] \\ V \arrow[r, "P"] & W \end{tikzcd}$$

kommutiert. Nach Folgerung 5.6 (2) bilden Diagonalmatrizen die zugehörige Normalform für die dort betrachtete Klasse von Abbildungen, und die Menge der Eigenwerte eine vollständige Invariante im Sinne von Bemerkung 3.17. Da nicht alle Abbildungen den Voraussetzungen der Folgerung genügen, erhalten wir noch keine allgemeine Aussage.

Der entscheidende Unterschied zu Folgerung 3.16 besteht darin, dass wir anstelle zweier verschiedenener Isomorphismen P und Q denselben Isomorphismus benutzen wollen für Definitions- und Wertebereich, denn bei einem Endomorphismus sind ja Definitions- und Wertebereich identisch.

5.2. Polynome

In diesem Abschnitt führen wir den Polynomring über einem gegebenen Ring ein und beweisen einige wichtige Sätze, insbesondere eine „eindeutige Primfaktorzerlegung“ für normierte Polynome.

Zur Motivation betrachten wir Bemerkung 5.2. Sei V ein n-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$ ein Endomorphismus. Eine Zahl $\lambda \in \Bbbk$ ist genau dann ein Eigenwert von F, wenn $\ker(F - \lambda \operatorname{id}_V) \neq \{0\}$, wenn also $F - \lambda \operatorname{id}_V$ nicht invertierbar ist, das heißt nach Folgerung 4.21 (1) genau dann, wenn $\det(F - \lambda\,\mathrm{id}_V) = 0$. Wir können also die Funktion

$$\chi_F \colon \Bbbk \to \Bbbk \qquad \text{mit} \qquad \chi_F(\lambda) = \det(F - \lambda\,\mathrm{id}_V)$$

betrachten und ihre Nullstellen suchen, um Eigenwerte von F zu finden. Ohne etwas über die Struktur dieser Funktion zu wissen, kann es allerdings schwierig werden, Nullstellen zu finden. Wir wollen die Funktion χ_F daher etwas algebraischer betrachten, was uns in vieler Hinsicht mehr Informationen liefert.

Wir erinnern uns an die Menge $R^{(\mathbb{N})}$ der endlichen R-wertigen Folgen, siehe Beispiel 2.30. Für jede Zahl $r \in R$ in einem Ring mit Eins definieren wir $r^0 = 1$, siehe Definition 1.37.

5.9. Definition. Es sei R ein kommutativer Ring mit Eins. Ein *Polynom* über R in der Variablen X ist ein Ausdruck der Form

$$P = P(X) = \sum_{i=0}^{\infty} p_i\, X^i \tag{1}$$

mit $(p_i)_i \in R^{(\mathbb{N})}$. Die Menge aller Polynome über R in der Variablen X bezeichnen wir mit $R[X]$.

Der größte Index $i \in \mathbb{N}$ mit $p_i \neq 0$ heißt der *Grad* $\deg P$ von P; falls $p_i = 0$ für alle $i \in \mathbb{N}$, setzen wir $\deg P = -\infty$. Polynome $P \in R[X]$ vom Grad $\deg P \leq 0$ heißen *konstant.* Wir identifizieren die Menge der konstanten Polynome mit R, so dass das konstante Polynom $P(X) = p_0\, X^0 \in R[X]$ gerade dem Element $p_0 \in R$ entspricht. Polynome $P \in R[X]$ vom Grad $\deg P \leq 1$ heißen *linear.*

Es sei $P(X) = \sum_{i=0}^{k} p_i\, X^i \in R[X] \setminus \{0\}$ ein Polynom vom Grad $\deg P \geq 0$, dann heißt $p_{\deg P}$ der *Leitkoeffizient* von P. Ein *normiertes Polynom* ist ein Polynom $P \neq 0$ mit Leitkoeffizient $p_{\deg P} = 1$.

Es sei $Q = \sum_{j=0}^{\infty} q_j\, X^j$ ein weiteres Polynom, dann definieren wir die *Summe* und das *Produkt* von P und Q durch

$$(P + Q)(X) = \sum_{i=0}^{\infty} (p_i + q_i)\, X^i \tag{2}$$

$$\text{und} \qquad (P \cdot Q)(X) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} (p_{i-j} \cdot q_j)\, X^i\,. \tag{3}$$

Außerdem definieren wir die *Auswertungsabbildung* $\mathrm{ev} \colon R[X] \times R \to R$ durch

$$\mathrm{ev}(P, r) = P(r) = \sum_{i=0}^{\infty} p_i \cdot r^i = \sum_{i=0}^{\infty} p_i \cdot \underbrace{r \cdots r}_{i \text{ Faktoren}}\,. \tag{4}$$

Wir identifizieren Polynome über R mit Funktionen $P \colon R \to R$, so dass $P(r) = \mathrm{ev}(P, r)$.

Man beachte, dass wir bei einem Polynom die Variable X in der Notation $P = P(X)$ mitschreiben oder weglassen dürfen. In der Regel lassen wir die

Variable X nur dann weg, wenn klar ist, dass P ein Polynom in der Variablen X ist.

Wir fassen die Menge R^R aller Funktionen von R nach R als Ring mit punktweiser Addition und Multiplikation auf, also

$$(f+g)(r) = f(r) + g(r) \qquad \text{und} \qquad (f \cdot g)(r) = f(r) \cdot g(r)$$

für alle f, $g\colon R \to R$. Unter der Abbildung $R[X] \to R^R$ werden die konstanten Polynome auf konstante Funktionen abgebildet.

5.10. Definition. Es seien R und S Ringe. Ein *Ringhomomorphismus* ist eine Abbildung $f\colon R \to S$, die für alle r, $t \in R$, die Axiome

(H1) $$f(r+t) = f(r) + f(t) \qquad \text{und}$$

(H2) $$f(r \cdot t) = f(r) \cdot f(t)$$

erfüllt. Seien R und S Ringe mit Eins, dann heißt f *unitär,* wenn zusätzlich

(H3) $$f(1_R) = 1_S\ .$$

Da 0 das einzige Element ist mit $0 = 0+0$, folgt aus (H1) bereits $f(0_R) = 0_S$ für alle Ringhomomorphismen.

5.11. Proposition. *Es sei R ein kommutativer Ring mit Eins $1 \in R$.*

(1) *Die Polynome bilden einen kommutativen Ring $(R[X], +, \cdot\,)$ mit Eins*
$$1_{R[X]} = 1 \cdot X^0 = \sum_{i=0}^{k} \delta_{i0}\, X^i\ .$$
(2) *Für alle $r \in R$ ist die Auswertung $\operatorname{ev}(\,\cdot\,, r)\colon R[X] \to R$ mit $P \mapsto P(r)$ ein unitärer Ringhomomorphismus.*
(3) *Die Abbildung $R[X] \to R^R$ ist ebenfalls ein unitärer Ringhomomorphismus.*

Für Rechnungen mit Polynomen gelten nach (1) also die gleichen Regeln wie im Ring R, nämlich Assoziativ-, Kommutativ- und Distributivgesetze. Die Aussagen (2) und (3) besagen, dass alle Rechnungen mit Polynomen gültig bleiben, wenn wir für X Elemente aus R in P *einsetzen.*

Außerdem wollen wir darauf hinweisen, dass ein lineares Polynom $P = aX + b$ nicht notwendigerweise eine lineare Abbildung $P\colon R \to R$ liefert — das gilt nur, wenn $b = 0$.

Beweis. Zu (1) müssen wir zunächst zeigen, dass $(R[X], +)$ eine abelsche Gruppe bildet. Da die Addition von Polynomen der Addition in $R^{(\mathbb{N})}$ entspricht, folgt das aus Beispiel 2.30. Die Ringaxiome (R1)–(R4) folgen durch Nachrechnen aus den entsprechenden Axiomen für R. Wir machen das hier für das erste Distributivgesetz vor. Seien

$$P(X) = \sum_{i=1}^{k} p_i\, X^i\ , \qquad Q(X) = \sum_{i=1}^{\ell} q_i\, X^i \qquad \text{und} \qquad R(X) = \sum_{i=1}^{m} r_i\, X^i$$

Polynome über R, dann gilt

$$\begin{aligned}\big(P\cdot(Q+R)\big)(X) &= \sum_{i=0}^{k+\max(\ell,m)} \sum_{j=0}^{i} \big(p_{i-j}\,(q_j+r_j)\big)\,X^i\\ &= \sum_{i=0}^{\max(k+\ell,k+m)} \Bigg(\sum_{j=0}^{i} p_{i-j}\,q_j + \sum_{j=0}^{i} p_{i-j}\,r_j\Bigg)\,X^i\\ &= (P\cdot Q+P\cdot R)(X)\;.\end{aligned}$$

Die anderen Axiome folgen durch ähnliche Rechnungen.

Zu (2) beweisen wir (H1)–(H3). Für $\mathrm{ev}(\,\cdot\,,r)$ haben diese Axiome die folgende Form:

$$(P+Q)(r)=P(r)+Q(r)\;, \tag{H1}$$
$$(P\cdot Q)(r)=P(r)\cdot Q(r) \qquad \text{und} \tag{H2}$$
$$1_{R[X]}(r)=1\;. \tag{H3}$$

Seien also P, Q wie oben und $r\in R$, dann gilt

$$\begin{aligned}(P+Q)(r) &= \sum_{i=0}^{\max(k,\ell)} (p_i+q_i)\,r^i = \sum_{i=0}^{k} p_i\,r^i + \sum_{i=0}^{\ell} q_i\,r^i = P(r)+Q(r)\;,\\ (P\cdot Q)(r) &= \sum_{i=0}^{k+\ell}\sum_{j=0}^{i} p_{i-j}\,q_j\,r^i = \sum_{i=0}^{k+\ell}\sum_{j=0}^{i} p_{i-j}\,r^{i-j}\cdot q_j\,r^j\\ &= \Bigg(\sum_{i=0}^{k} p_i\,r^i\Bigg)\cdot\Bigg(\sum_{j=0}^{\ell} q_j\,r^j\Bigg) = P(r)\cdot Q(r)\;,\end{aligned}$$

$$\text{und}\quad 1_{R[X]}(r)=1\cdot r^0=1\;.$$

Da Addition und Multiplikation in R^R punktweise definiert sind, folgt (3) aus (2). □

Im Beweis von (H2) haben wir unter anderem benutzt, dass R kommutativ ist. Aus diesem Grund betrachten wir Polynome nur über kommutativen Ringen, denn über nichtkommutativen Ringen gäbe es keine schöne Auswertungsabbildung. Außerdem wird unsere Notation für Polynome nun klarer: wir schreiben X als Abkürzung für das Polynom

$$X=\sum_{i=0}^{\infty}\delta_{i1}\,X^i\in R[X]\;,$$

dann folgt mit vollständiger Induktion und Definition 5.9 (3), dass

$$X^k=\sum_{i=0}^{\infty}\delta_{ik}\,X^i=\underbrace{X\cdots X}_{k\text{ Faktoren}}\;,$$

und Polynome sind Linearkombinationen der Basiselemente $1 = X^0$, $X = X^1$, X^2, … Typische Polynome über $\mathbb{Z}$ sind also zum Beispiel

$$13\ , \qquad X - 7\ , \qquad \text{und} \qquad X^2 + 1\ .$$

5.12. Bemerkung. Da nur endlich viele $p_i \neq 0$ sind, ist der Grad $p_i \in \mathbb{N} \cup \{-\infty\}$ wohldefiniert, und das einzige Polynom vom Grad $-\infty$ ist *Nullpolynom*

$$0_{R[X]} = \sum_{i=0}^{\infty} 0\, X^k\ .$$

Es seien jetzt P, $Q \in R[X]$ wie oben, mit $k = \deg P$ und $\ell = \deg Q$.

(1) Das *Maximum* $\max(k, \ell)$ zweier natürlicher Zahlen ist die größere von beiden. Außerdem setzen wir

$$\max(n, -\infty) = \max(-\infty, n) = n \qquad \text{für alle } n \in \mathbb{N} \cup \{-\infty\}\ .$$

Dann gilt

$$\deg(P + Q) \leq \max(\deg P, \deg Q)\ ,$$

denn für alle $m > \max(k, \ell)$ erhalten wir $p_m + q_m = 0$ als Koeffizienten von X^m in $P + Q$ nach Definition 5.9 (2).

(2) Wenn

$$\deg(P + Q) < \max(\deg P, \deg Q)\ ,$$

dann haben beide Polynome den gleichen Grad und die Summe der Leitkoeffizienten von P und Q verschwindet. Als Beispiel betrachte $P = X - 3$ und $Q = -X + 5$, dann ist

$$P + Q = (X - 3) + (-X + 5) = 2\ .$$

(3) Für das Produkt $P \cdot Q$ gilt

$$\deg(P \cdot Q) \leq \deg P + \deg Q\ ,$$

denn nach Definition 5.9 (3) ist der Koeffizient von $X^{k+\ell}$ in $P \cdot Q$ gerade $p_k \cdot q_\ell$, und höhere Potenzen von X kommen nicht vor. Falls $P = 0$ oder $Q = 0$, gilt diese Formel immer noch, wenn wir

$$(-\infty) + n = n + (-\infty) = -\infty \qquad \text{für alle } n \in \mathbb{N}$$

setzen.

(4) Der Fall

$$\deg(P \cdot Q) < \deg P + \deg Q$$

kann nur eintreten, wenn das Produkt der Leitkoeffizienten $p_k \cdot q_\ell$ verschwindet. Nach Voraussetzung ist $p_k \neq 0 \neq q_k$, also sind p_k und q_ℓ Nullteiler, siehe Bemerkung 2.13 (2).

5.13. Satz (Polynomdivision mit Rest)**.** *Es sei R ein kommutativer Ring mit Eins, es seien P, $Q \in R[X]$ Polynome über R, und Q sei normiert. Dann existieren eindeutige Polynome S, $T \in R[X]$, so dass*

$$P = S \cdot Q + T \tag{1}$$

$$\textit{und} \qquad \deg T < \deg Q\ . \tag{2}$$

Dieser Satz ist völlig analog zur Division mit Rest in $\mathbb{N}$, siehe Abschnitt 2.2 vor Satz 2.18. Dabei entspricht der Polynomgrad hier der Größe des Restes dort.

Beweis. Die Existenz von S und T beweisen wir durch Induktion über $k = \deg P$. Im Falle $\deg P < \deg Q$ setzen wir $T = P$ und $S = 0$ und sind fertig.

Wir nehmen jetzt an, dass $k = \deg P \geq \deg Q$, und dass wir alle Polynome vom Grad $< k$ mit Rest durch Q dividieren können. Es sei $\ell = \deg Q$. Da Q normiert ist, schreiben wir Q als

$$Q(X) = X^\ell + \sum_{j=0}^{\ell-1} q_j\, X^j \;.$$

Es sei $p_k \neq 0$ der Leitkoeffizient von P, dann betrachten wir das Polynom

$$P'(X) = P(X) - p_k\, X^{k-\ell} \cdot Q(X) \in R[X] \;.$$

Dann gilt

$$p_k\, X^{k-\ell} \cdot Q(X) = \sum_{i=0}^{\ell} p_k\, q_i\, X^{k-\ell+i} = p_k\, X^k + \sum_{j=k-\ell}^{k-1} p_k\, q_{j+\ell-k}\, X^j \;,$$

also verschwindet der Koeffizient vom Grad k in P', so dass $\deg P' < k$. Nach Induktionsvoraussetzung existieren S', $T \in R[X]$ mit $\deg T < \ell = \deg Q$, so dass

$$P' = S' \cdot Q + T \;.$$

Wir setzen $S(X) = S'(X) + p_k\, X^{\ell-k}$ und erhalten

$$\begin{aligned} P(X) &= P'(X) + p_k\, X^{k-\ell} \cdot Q(X) \\ &= \big(S'(X) + p_k\, X^{k-\ell}\big) \cdot Q(X) + T(X) = S(X) \cdot Q(X) + T(X) \;, \end{aligned}$$

womit die Existenz von S und T gezeigt ist.

Um die Eindeutigkeit zu beweisen, nehmen wir an, dass

$$P = S \cdot Q + T = S' \cdot T' \in R[X] \qquad \text{mit} \qquad \deg T \;, \quad \deg T' < \deg Q \;.$$

Dann gilt

$$(*) \qquad \deg\big((S - S') \cdot Q\big) = \deg(T' - T) < \deg Q \;,$$

denn aus Bemerkung 5.12 (1) folgt $\deg(T - T') \leq \max(\deg T, \deg T') < \deg Q$. Wir nehmen an, dass $S - S' \neq 0$, dann sei $n = \deg(S - S')$, und s_n sei der Leitkoeffizient. Aus Bemerkung 5.12 (3) folgt

$$\deg\big((S - S') \cdot Q\big) \leq \deg(S - S') + \deg Q = n + \ell \;,$$

und der Koeffizient vor $X^{\ell+n}$ wird gegeben durch $s_n \cdot q_\ell = s_n \neq 0$, so dass

$$\deg\big((S - S') \cdot Q\big) = \deg(S - S') + \deg Q \geq \deg Q$$

im Widerspruch zu (*). Also gilt $S = S'$, und Eindeutigkeit folgt, da

$$T' - T = (S - S') \cdot Q = 0 \in R[X] \;. \qquad \square$$

Wenn wir über einem Körper arbeiten, können wir durch beliebige Polynome $Q \neq 0$ dividieren. Dazu dividieren wir alle Koeffizienten von Q durch den Leitkoeffizienten und erhalten so ein normiertes Polynom. Am Ende müssen wir dann das Ergebnis S noch durch den Leitkoeffizienten von Q teilen. Als Beispiel betrachte $P = X^2 - 1$ und $Q = 2X + 1$. Wir dividieren zunächst durch $\frac{1}{2}Q$ und erhalten

$$X^2 - 1 = \left(X - \frac{1}{2}\right) \cdot \left(X + \frac{1}{2}\right) - \frac{3}{4}\,,$$

also gilt

$$X^2 - 1 = \left(\frac{X}{2} - \frac{1}{4}\right) \cdot (2X + 1) - \frac{3}{4}\,.$$

Im Folgenden werden wir allerdings meistens durch normierte Polynome dividieren.

Wir schreiben $Q \mid P$, wenn die obige Division ohne Rest möglich ist, das heißt, wenn $S \in R[X]$ existiert, so dass $P = S \cdot Q$. Andernfalls schreiben wir $Q \nmid P$. Falls R ein Körper ist, können wir sogar den *größten gemeinsamen Teiler* zweier Polynome definieren und mit dem Euklidischen Algorithmus 2.18 ausrechnen (Übung). Um ein eindeutiges Ergebnis zu erhalten, verlangen wir, dass der größte gemeinsame Teiler wieder ein normiertes Polynom ist.

5.14. Definition. Ein Ring R heißt *nullteilerfrei,* wenn für alle r, $s \in R$ aus $r \cdot s = 0$ bereits folgt, dass $r = 0$ oder $s = 0$. Ein *Integritätsbereich* oder *Integritätsring* ist ein nullteilerfreier, kommutativer Ring mit Eins.

In Bemerkung 2.13 haben wir uns überlegt, dass Ringe genau dann nullteilerfrei sind, wenn für die Multiplikation Kürzungsregeln gelten. Insbesondere ist jeder Schiefkörper nullteilerfrei und somit ist jeder Körper ein Integritätsbereich. Aus Bemerkung 5.12 (4) folgt, dass ein Polynomring über einem Integritätsbereich auch wieder ein Integritätsbereich ist.

5.15. Folgerung. *Es sei R ein kommutativer Ring mit Eins, dann gilt für alle $r \in R$ und alle $P \in R[X]$, dass*

$$P(r) = 0 \qquad \Longleftrightarrow \qquad (X - r) \mid P\,. \tag{1}$$

Wenn R ein Integritätsbereich ist, gilt für alle $r \in R$ und alle Polynome P, $Q \in R[X]$, dass

$$(X - r) \mid (P \cdot Q) \qquad \Longleftrightarrow \qquad \big((X - r) \mid P \text{ oder } (X - r) \mid Q\big)\,. \tag{2}$$

Polynome vom Typ $X - r$ nennen wir auch *Linearfaktoren.*

Beweis. Wir dividieren $P \in R[X]$ durch $X - r$ mit Rest wie in Satz 5.13 und erhalten S, $T \in R[X]$ mit $\deg T < 1 = \deg(X - r)$. Also ist $T = t \in R$ ein konstantes Polynom, und es gilt

$$P = S \cdot (X - r) + t\,.$$

Nach Proposition 5.11 (2) dürfen wir $X = r$ einsetzen und erhalten (1), da

$$P(r) = S(r) \cdot (r - r) + t = t\,.$$

Es gelte $(X - r) \mid (P \cdot Q)$, dann folgt

$$(P \cdot Q)(r) = P(r) \cdot Q(r) = 0$$

aus (1) und Proposition 5.11 (2). Da R nach Voraussetzung nullteilerfrei ist, gilt $P(r) = 0$ oder $Q(r) = 0$. Mit (1) folgt $(X - r) \mid P$ oder $(X - r) \mid Q$. Also gilt „$\Longrightarrow$“ in (2), die Rückrichtung ist klar. $\square$

5.16. Bemerkung. In der obigen Folgerung brauchen wir Nullteilerfreiheit für (2). Sei etwa $R = \mathbb{Z}/6\mathbb{Z}$ und sei $P = X^2 - X \in (\mathbb{Z}/6\mathbb{Z})[X]$. Nachrechnen liefert die Tabelle

r	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$
$P(r)$	$[0]$	$[0]$	$[2]$	$[0]$	$[0]$	$[2]$

.

Also hat P vier verschiedene Nullstellen $r_1 = [0]$, $r_2 = [1]$, $r_3 = [3]$ und $r_4 = [4]$ und somit vier Teiler $Q_i = X - r_i$ für $i = 1, \dots, 4$ nach (1). Tatsächlich gilt

$$P = X \cdot (X - [1]) = (X - [3]) \cdot (X - [4]) = Q_1 \cdot Q_2 = Q_3 \cdot Q_4 .$$

Aber keines der vier Polynome Q_i teilt eines der anderen, denn wieder nach (1) reicht es zu zeigen, dass

$$0 \neq Q_i(r_j) = r_j - r_i \qquad \text{für alle } i,\, j \text{ mit } i \neq j .$$

Jetzt erhalten wir einen Widerspruch zu (2), denn beispielsweise gilt

$$Q_3 \mid (Q_1 \cdot Q_2) , \qquad \text{aber} \qquad Q_3 \nmid Q_1 \quad \text{und} \quad Q_3 \nmid Q_2 .$$

Auch Kommutativität ist wichtig: sei etwa $R = \mathbb{H}$, dann ist $\mathbb{H}$ als Schiefkörper nullteilerfrei, siehe Bemerkung 2.13 (2). Dennoch gilt

$$X^2 + 1 = (X + i) \cdot (X - i) = (X + j) \cdot (X - j) = (X + k) \cdot (X - k)$$

mit $i, j, k \in \mathbb{H}$ wie in Bemerkung 1.73, und es gibt noch unendlich viele weitere solche Zerlegungen.

Selbstverständlich kann ein Linearfaktor auch mehrfach, also in einer höheren Potenz vorkommen, beispielsweise gilt

$$X^3 - 3X - 2 = (X - 2) \cdot (X + 1)^2 .$$

5.17. Definition. Es sei $P \in R[X] \setminus \{0\}$, $r \in R$ und $k \in \mathbb{N}$. Dann heißt r eine Nullstelle der *Ordnung* k von P, wenn

$$(X - r)^k \mid P \qquad \text{und} \qquad (X - r)^{k+1} \nmid P ,$$

und wir schreiben $\operatorname{ord}_r P = k$.

Insbesondere ist r eine Nullstelle der Ordnung 0, wenn $P(r) \neq 0$. Anders gesagt ist eine „Nullstelle der Ordnung 0“ gar keine echte Nullstelle von P.

5.18. Proposition. *Es sei R ein Integritätsbereich und $P \in R[X] \setminus \{0\}$, dann gilt*

(1) $$\sum_{r \in R} \operatorname{ord}_r P \leq \deg P .$$

Insbesondere hat ein Polynom vom Grad k höchstens k verschiedene Nullstellen.

Wir nennen die linke Seite von (1) auch die *gewichtete Anzahl* der Nullstellen von P. Gemeint ist, dass Nullstellen höheren Ordnung mehrfach, also mit höherem „Gewicht“ gezählt werden.

Beweis. In der obigen Summe interessieren nur echte Nullstellen von P, also diejenigen $r \in R$ mit $\operatorname{ord}_r P \neq 0$. Es seien also zunächst endlich viele r_1, $\dots$, $r_\ell \in R$ gegeben mit $P(r_i) = 0$ für alle i und $r_i \neq r_j$ für alle $i \neq j$. Wegen Folgerung 5.15 (1) gilt wie in Bemerkung 5.16, dass $(X - r_i) \nmid (X - r_j)$ für $i \neq j$, da $r_j - r_i \neq 0$.

Durch Induktion über ℓ erhalten wir ein Polynom $S = S_\ell$ vom Grad $\deg S_\ell = \deg P - \ell$, so dass

$$P = S_\ell \cdot (X - r_1)^{\operatorname{ord}_{r_1} P} \cdots (X - r_\ell)^{\operatorname{ord}_{r_\ell} P} \;.$$

Für $\ell = 0$ setzen wir $S_0 = P$. Sei $\ell > 0$, dann existiert $S_{\ell-1}$ wie oben nach Induktion, und es folgt

$$(X - r_\ell) \mid \bigl(S_{\ell-1} \cdot (X - r_1)^{\operatorname{ord}_{r_1} P} \cdots (X - r_{\ell-1})^{\operatorname{ord}_{r_{\ell-1}} P}\bigr) \;.$$

Da $(X - r_\ell) \nmid (X - r_{\ell-1})$, gilt $(X - r_\ell) \nmid (X - r_{\ell-1})^{\operatorname{ord}_{r_{\ell-1}} P}$ und

$$(X - r_\ell) \mid \bigl(S_{\ell-1} \cdot (X - r_1) \cdots (X - r_{\ell-2})\bigr)$$

wegen Folgerung 5.15 (2). Mit dem gleichen Argument folgt schließlich, dass $(X - r_\ell) \mid S_{\ell-1}$. Falls $\operatorname{ord}_{r_\ell} P > 1$, fahren wir fort mit

$$P/(X - r_\ell) = \bigl(S_{\ell-1}/(X - r_\ell)\bigr) \cdot (X - r_1)^{\operatorname{ord}_{r_1} P} \cdots (X - r_{\ell-1})^{\operatorname{ord}_{r_{\ell-1}} P} \;.$$

Auf diese Weise erhalten wir $(X - r_\ell)^{\operatorname{ord}_{r_\ell} P} \mid S_{\ell-1}$ nach endlich vielen Schritten. Also existiert das gesuchte $S_\ell \in R[X]$ mit $S_{\ell-1} = S_\ell \cdot (X - r_\ell)^{\operatorname{ord}_{r_\ell} P}$.

Da R nullteilerfrei ist, folgt mit Bemerkung 5.12 (4), dass

$$\begin{aligned} \deg P &= \deg S_\ell + \deg(X - r_1)^{\operatorname{ord}_{r_1} P} + \dots + \deg(X - r_\ell)^{\operatorname{ord}_{r_\ell} P} \\ &= \deg S_\ell + \sum_{i=1}^{\ell} \operatorname{ord}_{r_i} P \geq \sum_{i=1}^{\ell} \operatorname{ord}_{r_i} P \;. \end{aligned}$$

Insbesondere kann es insgesamt nur endlich viele Nullstellen geben, so dass die Summe über $\operatorname{ord}_r P$ wohldefiniert ist. Die letzte Aussage ergibt sich daraus, dass jede echte Nullstelle mindestens Ordnung 1 hat. □

5.19. Bemerkung. Der obige Beweis erinnert ein wenig an die eindeutige Primfaktorzerlegung natürlicher Zahlen. Tatsächlich kann man jedes normierte Polynom über einem Körper $\Bbbk$ auf eindeutige Weise als Produkt normierter Polynome schreiben, die sich selbst nicht weiter zerlegen lassen. Welche Polynome als „Primfaktoren“ in Frage kommen, hängt von $\Bbbk$ ab. Betrachte etwa das Polynom

$$P(X) = X^3 - 2 \in \mathbb{Q}[X] \subset \mathbb{R}[X] \subset \mathbb{C}[X] \;.$$

Über $\mathbb{C}$ erhalten wir die Nullstellen

$$\sqrt[3]{2}\;, \qquad -\sqrt[3]{2}\,\frac{1+\sqrt{3}\,i}{2}\;, \qquad \text{und} \qquad -\sqrt[3]{2}\,\frac{1-\sqrt{3}\,i}{2}\;,$$

wie man durch Nachrechnen überprüft. Mehr Nullstellen kann es dank Proposition 5.18 nicht geben. Somit erhalten wir

$$X^3 - 2 = \bigl(X - \sqrt[3]{2}\bigr) \cdot \left(X + \sqrt[3]{2}\,\frac{1+\sqrt{3}\,i}{2}\right) \cdot \left(X + \sqrt[3]{2}\,\frac{1-\sqrt{3}\,i}{2}\right) \in \mathbb{C}[X]\,.$$

Über den reellen Zahlen sehen wir nur eine Nullstelle und erhalten daher die „Primfaktorzerlegung“

$$X^3 - 2 = \bigl(X - \sqrt[3]{2}\bigr) \cdot \bigl(X^2 + \sqrt[3]{2}\,X + (\sqrt[3]{2})^2\bigr) \in \mathbb{R}[X]\,.$$

Da $\sqrt[3]{2} \notin \mathbb{Q}$, ist das Polynom X^3-2 in $\mathbb{Q}$ unzerlegbar, also selbst ein Primfaktor. Mehr dazu lernen Sie in Abschnitt 6.3.

5.20. Bemerkung. Es sei R ein Integritätsbereich und P, $Q \in R[X]$ Polynome vom Grad $\le n$. Seien $x_0, \dots, x_n \in R$ paarweise verschiedene Punkte. Wenn $P(x_i) = Q(x_i)$ für $i = 0, \dots, n$, dann gilt $P = Q$, denn das Polynom $Q - P$ hat dann $n+1$ Nullstellen $x_0, \dots, x_n \in R$ und ebenfalls Grad $\le n$. Nach Proposition 5.18 geht das aber nur, wenn $Q - P = 0$.

Somit kann man ein Polynom $P \in \Bbbk[X]$ vom Grad $\le n$ eindeutig bestimmen, wenn man seine Werte an $n+1$ verschiedenen Elementen von $\Bbbk$ vorgibt. In den Übungen sehen Sie, dass ein solches Polynom auch immer existiert. Das funktioniert aber nur, wenn $\Bbbk$ überhaupt $n+1$ verschiedene Elemente enthält. Mit unendlichen Körpern wie $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ haben wir hier kein Problem. Aber wir haben in Beispiel 2.17 auch endliche Körper kennengelernt, der kleinste ist $\mathbb{Z}/2\mathbb{Z}$. Über diesem Körper kann man bereits Polynome vom Grad ≥ 2 nicht mehr anhand Ihrer Werte unterscheiden. Für das Polynom $P(X) = X^2+X \ne 0$ gilt zum Beispiel $P([0]) = P([1]) = 0$ genau wie für das Nullpolynom.

In Bemerkung 5.19 zerfällt das Polynom P über $\mathbb{C}$ in Linearfaktoren. Das liegt daran, dass jedes nicht-konstante Polynom über $\mathbb{C}$ eine Nullstelle hat.

5.21. Definition. Ein Körper $\Bbbk$ heißt *algebraisch abgeschlossen,* wenn jedes Polynom $P \in \Bbbk[X]$ mit $\deg P \ge 1$ mindestens eine Nullstelle besitzt.

Nach dem Fundamentalsatz 1.61 der Algebra ist der Körper $\mathbb{C}$ algebraisch abgeschlossen. Weitere Beispiele lernen Sie in der Vorlesung „Algebra“ kennen.

5.22. Folgerung. *Jedes normierte Polynom über einem algebraisch abgeschlossenen Körper zerfällt in Linearfaktoren. Diese Zerlegung ist bis auf die Reihenfolge der Linearfaktoren eindeutig. Darüberhinaus gilt Gleichheit in Proposition 5.18* (1).

Beweis. Es sei $P \in \Bbbk[X]$ normiert, insbesondere ist $P \ne 0$ und daher $\deg P \ge 0$. Wir beweisen die Aussage durch Induktion über $\deg P$. Für $\deg P = 0$ ist $P = 1$ das leere Produkt, und es ist nichts zu zeigen.

Sei die Aussage also für alle Polynome vom Grad $< k$ bewiesen, und sei $\deg P = k$. Nach unser Annahme hat P eine Nullstelle $x \in \Bbbk$ der Ordnung $\operatorname{ord}_x P \ge 1$. Wie im Beweis von Proposition 5.18 schreiben wir

$$P = S \cdot (X - x)^{\operatorname{ord}_x P}\,,$$

so dass $S(x) \neq 0$. Da $\deg S = \deg P - \operatorname{ord}_x P < \deg P = k$, können wir S nach Induktionsvoraussetzung als Produkt von Linearfaktoren schreiben. Wir multiplizieren mit $\operatorname{ord}_x P$ vielen Linearfaktoren $(X - x)$ und erhalten die gesuchte Zerlegung von P.

Die Zerlegung ist eindeutig, denn aus dem Beweis von Proposition 5.18 folgt auch, dass jeder Linearfaktor $(X - x)$ genau in der $(\operatorname{ord}_x P)$-ten Potenz vorkommt. Damit sind alle Aussagen bewiesen. □

Am Anfang des Abschnitts haben wir gesagt, dass wir $\lambda \mapsto \det(F - \lambda \operatorname{id}_V)$ als Polynom, und nicht als Funktion in $\lambda \in \Bbbk$ betrachten wollen, da wir dann mehr über die Nullstellen aussagen können. Als Fazit können wir festhalten, dass Nullstellen von Polynomen Ordnungen haben, und dass die — wie in Proposition 5.18 gewichtete — Anzahl der Nullstellen durch den Grad beschränkt wird. Wenn wir über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$ arbeiten, hat jedes Polynom tatsächlich auch — wieder im gewichteten Sinn — so viele Nullstellen, wie es der Grad vorgibt. All diese Aussagen sind für beliebige Funktionen in dieser Form nicht möglich. Über endlichen Körpern $\Bbbk$ enthalten Polynome P vom Grad $\deg P \geq \#\Bbbk$ mehr Informationen als die zugehörigen Funktionen $P(\,\cdot\,)\colon \Bbbk \to \Bbbk$.

5.3. Das Charakteristische Polynom und das Minimalpolynom

Es sei $\Bbbk$ ein Körper, V ein n-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$ ein Endomorphismus. Dann können wir eine Basis B von V wählen und erhalten die Abbildungsmatrix $A \in M_n(\Bbbk)$ von F bezüglich B, siehe Folgerung 2.75. Wir betrachten jetzt die Matrix

$$X \cdot E_n - A \in M_n\bigl(\Bbbk[X]\bigr) ,$$

dabei ist E_n wieder die Einheitsmatrix. Wir hatten in Kapitel 4 die Determinante quadratischer Matrizen über einem Ring definiert, siehe Definition 4.9. Also können wir von der obigen Matrix die Determinante bilden und erhalten

$$\chi_A(X) = \det\bigl(X \cdot E_n - A\bigr) \in \Bbbk[X] .$$

Wir können zeigen, dass $\chi_A(X)$ nicht von der Wahl der Basis B abhängt. Sei nämlich C eine weitere Basis und $P \in GL(n, \Bbbk)$ die Basiswechselmatrix mit $C = B \cdot P$, siehe Bemerkung 2.76. Dann hat F bezüglich der Basis C die Abbildungsmatrix $P^{-1} \cdot A \cdot P$. Wir fassen P als Matrix in $M_n(\Bbbk[X])$ auf. Da P mit der Einheitsmatrix E_n kommutiert, folgt aus Satz 4.19, dass

$$\begin{aligned}\det\bigl(X \cdot E_n - P^{-1} \cdot A \cdot P\bigr) &= \det\bigl(P^{-1} \cdot (X \cdot E_n - A) \cdot P\bigr)\\ &= \det P^{-1} \cdot \det\bigl(X \cdot E_n - A\bigr) \cdot \det P = \det\bigl(X \cdot E_n - A\bigr) .\end{aligned}$$

Somit hängt $\chi_A(X)$ nicht von der Wahl der Basis B ab, wir dürfen es also als Invariante des Endomorphismus F betrachten.

5.23. Definition. Es sei $\Bbbk$ ein Körper und $n \in \mathbb{N}$. Sei $A \in M_n(\Bbbk)$, dann heißt

$$\chi_A(X) = \det\bigl(X \cdot E_n - A\bigr) \in \Bbbk[X]$$

das *charakterische Polynom* der Matrix A. Sei V ein n-dimensionaler $\Bbbk$-Vektorraum mit Basis B und $F \in \operatorname{End}_\Bbbk V$ ein Endomorphismus mit Abbildungsmatrix A bezüglich B. Dann heißt $\chi_F(X) = \chi_A(X)$ das charakteristische Polynom von F.

Zur Berechnung des charakeristischen Polynoms empfiehlt sich zum Beispiel die Laplace-Entwicklung 4.12. Der Gauß-Algorithmus funktioniert nicht, da $\Bbbk[X]$ kein Körper ist.

5.24. Bemerkung. Es lohnt sich, das charakteristische Polynom etwas detaillierter zu betrachten.

(1) Die Einträge der Matrix $X \cdot E_n - A$ sind Polynome vom Grad ≤ 1. Aufgrund der Leibniz-Formel 4.13 lässt sich $\chi_A(X)$ als Summe von Produkten aus je n Matrixeinträgen schreiben. Also folgt aus Bemerkung 5.12, dass $\deg \chi_A(X) \leq n$. Zum Koeffizienten von X^n tragen nur diejenigen Produkte bei, bei denen alle n Faktoren den Grad 1 haben. Da dies genau die Diagonalelemente sind, trägt also nur das Produkt zur Permutation $\mathrm{id} \in S(n)$ bei. Also hat $\chi_A(X)$ den gleichen Leitkoeffizienten wie

$$(X - a_{11}) \cdots (X - a_{nn}) = X^n - (a_{11} + \cdots + a_{nn})\, X^{n-1} + \dots\ , \tag{*}$$

nämlich 1. Somit sind $\chi_A(X)$ und $\chi_F(X)$ stets normierte Polynom vom Grad

$$\deg \chi_A(X) = n \qquad \text{beziehungsweise} \qquad \deg \chi_F(X) = \dim V\ .$$

(2) Da $\chi_F(X)$ nur von F abhängt, sind die Koeffizienten von $\chi_F(X)$ interessante Invarianten des Endomorphismus F. Schreibe also

$$\chi_F(X) = X^n - \sigma_1(F)\, X^{n-1} \pm \cdots + (-1)^n\, \sigma_n(F)\, X^0\ ,$$

dann nennt man $\sigma_i(F)$ die *elementarsymmetrischen Funktionen* von F. Analog definieren wir $\sigma_i(A)$. Aus der Leibniz-Formel folgt, dass $\sigma_i(A)$ eine Summe von Produkten von je i Matrixeinträgen von A und einem Vorfaktor ist, denn jedes Produkt in der Leibniz-Formel 4.13 hat n Faktoren, die je entweder X oder ein Matrixeintrag sind.

(3) Als erstes wollen wir den konstanten Term $(-1)^n\sigma_n(F)$ des charakteristischen Polynoms bestimmen. Sei also A wieder die Abbildungsmatrix von F bezüglich einer Basis B. Dann setzen wir $X = 0$ und erhalten sofort

$$\chi_A(0) = \det(-A) = (-1)^n \det A\ ,$$

und analog für Endomorphismen F. Somit gilt $\sigma_{\dim V}(F) = \det F$.

(4) Wir schauen uns noch die Funktion $\sigma_1(A)$ an. Für jede Permutation $\rho \in S(n) \setminus \{\mathrm{id}\}$ gibt es wenigstens zwei verschiedene Indizes i,

$j \in \{1, \dots, n\}$ mit $\rho(i) \neq i$ und $\rho(j) \neq j$. Also liefert ρ einen Summand vom Grad $\leq n-2$. Der einzige Beitrag zu $\sigma_1(A)$ in der Leibniz-Formel 4.13 kommt also wieder von $\rho = \mathrm{id} \in S(n)$. Aus (*) folgt

$$\sigma_1(A) = a_{11} + \dots + a_{nn} = \mathrm{tr}(A) ,$$

siehe Definition 4.24. Insbesondere können wir die Spur eines Endomorphismus unabhängig von der Basis definieren. Für Matrizen $A \in M_n(\Bbbk)$ und $G \in GL(n, \Bbbk)$ gilt also

$$\mathrm{tr}(G^{-1} \cdot A \cdot G) = \mathrm{tr}(A) .$$

In Aufgabe 1 der Probeklausur zur linearen Algebra I haben wir noch allgemeiner gezeigt, dass für Matrizen $B \in M_{m,n}(R)$ und $C \in M_{n,m}(R)$ gilt, dass

$$\mathrm{tr}(B \cdot C) = \mathrm{tr}(C \cdot B) .$$

Indem wir $B = G^{-1}$ und $C = AG$ setzen, erhalten wir daraus die obige Gleichung.

Wir erinnern uns an die Ordnung von Nullstellen aus Definition 5.17 und bezeichnen den Eigenraum von F zum Eigenwert λ wieder mit $V_\lambda = \ker\bigl(\lambda\, \mathrm{id}_V - F\bigr)$.

5.25. Definition. Es sei V ein $\Bbbk$-Vektorraum mit $\dim V < \infty$ und $F \in \mathrm{End}_\Bbbk V$. Dann heißt $\lambda \in \Bbbk$ ein Eigenwert von F der *geometrischen Vielfachheit*

$$\dim \ker\bigl(\lambda\, \mathrm{id}_V - F\bigr) = \dim V_\lambda$$

und der *algebraischen Vielfachheit*

$$\mathrm{ord}_\lambda\, \chi_F .$$

Analog definieren wir Vielfachheiten für quadratische Matrizen über $\Bbbk$.

Insbesondere sind beide Vielfachheiten 0, wenn λ kein Eigenwert von F ist.

Als Beispiel betrachten wir die Matrix

$$A = \begin{pmatrix} 1 & -1 \\ 4 & 5 \end{pmatrix} \in M_2(\mathbb{R})$$

$$\text{mit} \qquad \chi_A(X) = \det \begin{pmatrix} X-1 & 1 \\ -4 & X-5 \end{pmatrix} = (X-1)(X-5) - 1 \cdot (-4)$$
$$= X^2 - 6X + 9 = (X-3)^2 .$$

Somit ist $\lambda = 3$ der einzige Eigenwert, und seine algebraische Vielfachheit ist $\mathrm{ord}_3\, \chi_A = 2$. Wäre die geometrische Vielfachheit ebenfalls 2, so wäre ganz $\mathbb{R}^2$ Eigenraum, und daher $A = 3\, E_2$. Da das nicht der Fall ist, der Eigenraum aber wegen $\chi_A(3) = 0$ auch nicht $\{0\}$ sein kann, ist die geometrische Vielfachheit 1. In der Tat liefert das Gauß-Verfahren

$$V_3 = \ker\bigl(3\, E_2 - A\bigr) = \left\langle \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\rangle ,$$

und es gilt $\dim V_3 = 1$.

5.26. Bemerkung. Es gilt stets

$$\dim V_\lambda \le \operatorname{ord}_\lambda \chi_F\ .$$

Denn sei V_λ der Eigenraum zu λ mit $k = \dim V_\lambda$, und sei W ein Komplement von V_λ in V, siehe Proposition 3.9. Wir wählen Basen von V_λ und W. Da $F(V_\lambda) \subset V_\lambda$ und $F|_{V_\lambda} = \lambda\ \mathrm{id}_{V_\lambda}$, wird F durch eine Blockmatrix der Gestalt

$$A = \begin{pmatrix} \lambda\, E_k & B \\ 0 & D \end{pmatrix}$$

dargestellt. Aus Folgerung 4.17 (1) folgt

$$\chi_F(X) = \det\bigl((X-\lambda)\cdot E_k\bigr)\cdot \det\bigl(X\cdot E_{n-k} - D\bigr) = (X-\lambda)^k \cdot \chi_D(X)\ .$$

Also erhalten wir

$$\operatorname{ord}_\lambda \chi_F = k + \operatorname{ord}_\lambda \chi_D \ge k = \dim V_\lambda\ .$$

Eine weitere Invariante eines Endomorphismus ist sein Minimalpolynom. Um das Minimalpolynom einzuführen, wollen wir uns überlegen, dass jeder Endomorphismus $F \in \operatorname{End}_{\Bbbk} V$ den Vektorraum V zu einem $R[X]$-Modul macht, auf dem die Variable X wie F wirkt. Mit anderen Worten dürfen wir für X nicht nur Ringelemente einsetzen. Die folgende Vorüberlegung soll das verständlicher machen. Wir erinnern uns an den Begriff des unitären Ringhomomorphismus, aus Definition 5.10.

5.27. Proposition (Universelle Eigenschaft des Polynomrings)**.** *Es sei R ein kommutativer Ring mit Eins, es sei S ein beliebiger Ring mit Eins, und es sei $h\colon R \to S$ ein unitärer Ringhomomorphismus. Dann existiert zu jedem $s \in S$ mit $s \cdot h(r) = h(r) \cdot s$ für alle $r \in R$ ein eindeutiger unitärer Ringhomomorphismus $H\colon R[X] \to S$ mit $H(r) = h(r)$ für alle $r \in R$ und $H(X) = s$.*

$$\begin{tikzcd} R[X] \arrow[dr, dashed, "H"] & \{X\} \arrow[l, hook'] \arrow[d, "s"] \\ R \arrow[u, hook] \arrow[r, "h"] & S\ . \end{tikzcd}$$

Wenn h injektiv und bekannt ist, identifizieren wir R mit $\operatorname{im} h \subset S$ und schreiben $H(\,\cdot\,) = \operatorname{ev}(\,\cdot\,, s)$ und $H(P) = P(s)$ in Analogie zu Definition 5.9 (4). Wenn wir den *Zentralisator* von R definieren als

$$Z_S(R) = \bigl\{\, s \in S \bigm| rs = sr \text{ für alle } r \in R \,\bigr\}\ ,$$

so erhalten wir eine Auswertungsabbildung

$$\operatorname{ev}\colon R[X] \times Z_S(R) \to S \qquad \text{mit} \qquad \operatorname{ev}(P, s) = P(s)\ .$$

Proposition 5.11 (2) und (3) sind Spezialfälle: in 5.11 (2) ist $S = R$, $h = \mathrm{id}$, und $s = r$ ein festes Element von R. In 5.11 (3) ist $S = R^R$, h bildet R auf die konstanten Funktionen in R^R ab, und X wird auf $s = \mathrm{id}_R \in R^R$ abgebildet. Die obige Proposition erlaubt uns, auch andere Funktionen einzusetzen, beispielsweise dürften wir für $P \in \mathbb{R}[X]$ auch $P(\sin(x)) \in C^\infty(\mathbb{R})$ betrachten. Für $P(X) = 1 - X^2$ beispielsweise ist

$$P(\sin(x)) = 1 - \sin^2(x) = \cos^2(x) \;.$$

Man sagt, „$\cos^2(x)$ ist ein Polynom in $\sin(x)$." Oder aber, wir betrachten den Raum $C^\infty(\mathbb{R})$ der unendlich oft differenzierbaren Funktionen. Die Ableitung $\frac{\partial}{\partial x}$ ist ein Endomorphismus von $C^\infty(\mathbb{R})$. Für das Polynom $P = -X^2$ erhalten wir also den Differentialoperator

$$P\biggl(\frac{\partial}{\partial x}\biggr) = -\frac{\partial^2}{\partial x^2} \in \operatorname{End} C^\infty(\mathbb{R}) \;.$$

Beweis. Wir betrachten ein Polynom

$$P(X) = \sum_{i=0}^n p_i X^i \in R[X]$$

mit $p_i \in R$ für alle i. Aus den Axiomen (H1)–(H3) folgt

$$H(P) = H\biggl(\sum_{i=0}^n p_i X^i\biggr) = \sum_{i=0}^n H\bigl(p_i X^i\bigr) = \sum_{i=0}^n H(p_i) \cdot H(X)^i = \sum_{i=0}^n h(p_i) \cdot s^i \;.$$

Das beweist die Eindeutigkeit von H.

Wir überprüfen die Axiome. Dazu sei

$$Q(X) = \sum_{i=0}^m q_j X^j$$

ein weiteres Polynom. Wir setzen $p_i = q_j = 0$ für $i > n$ und $j > m$. Dann gilt

$$\begin{aligned}
H(P+Q) &= \sum_{i=0}^\infty h(p_i + q_i)\, s^i = \sum_{i=0}^n h(p_i)\, s^i + \sum_{i=0}^m h(q_i)\, s^i = H(P) + H(Q) \;,\\
H(P \cdot Q) &= \sum_{k=0}^\infty h\biggl(\sum_{i+j=k} p_i q_j\biggr)\, s^k = \sum_{i=0}^n h(p_i)\, s^i \cdot \sum_{j=0}^m h(q_j)\, s^j = H(P) \cdot H(Q) \;,\\
H(1_{R[X]}) &= \sum_{i=0}^\infty h(\delta_{i0})\, s^i = s^0 = 1 \;.
\end{aligned}$$

Dabei haben wir ausgenutzt, dass h die Axiome (H1)–(H3) erfüllt, und für (H2) haben wir auch verwendet, dass s mit allen $h(q_j)$ kommutiert. Somit ist H tatsächlich ein Ringhomomorphismus. □

5.28. Beispiel. Es sei R ein kommutativer Ring mit Eins und M ein R-Modul. Den Endomorphismenring $\operatorname{End}_R M$ haben wir in Folgerung 2.42 betrachtet, und analog den Matrixring $M_n(R)$ in Folgerung 2.71. Wir erhalten einen injektiven Ringhomomorphismus $R \to \operatorname{End}_R M$ mit $r \mapsto r \operatorname{id}_M \in \operatorname{End}_R M$. Es gilt $r \operatorname{id}_M \circ F = F \circ r \operatorname{id}_M$ für alle $F \in \operatorname{End}_R M$ und alle $r \in R$, somit $Z_{\operatorname{End}_R M} R = \operatorname{End}_R M$, und wir erhalten eine Auswertungsabbildung

$$\operatorname{ev}\colon R[X] \times \operatorname{End}_R M \to \operatorname{End}_R M \qquad \text{mit} \qquad \operatorname{ev}(P, F) = P(F) \;,$$

und $\operatorname{ev}(\,\cdot\,, F) \colon R[X] \to \operatorname{End}_R M$ mit $P(X) \mapsto P(F)$ ist ein unitärer Ringhomomorphismus für alle $F \in \operatorname{End}_F M$.

Sei speziell $M = R^n$, so dass $\operatorname{End}_R M = M_n(R)$, dann erhalten wir für jede Matrix $A \in M_n(R)$ einen unitären Ringhomomorphismus $R[X] \to M_n(R)$ mit $P(X) \mapsto P(A)$.

5.29. Satz (Cayley-Hamilton)**.** *Es sei R ein kommutativer Ring mit Eins und $A \in M_n(R)$. Dann gilt $\chi_A(A) = 0$.*

An manchen Stellen findet sich zu diesem Satz die folgende Heuristik: „Einsetzen von A in χ_A liefert $\det(A \cdot E_n - A) = 0$.“ So einfach ist es leider nicht, denn die beiden As in der obigen Formel leben in verschiedenen Ringen. Um das zu verdeutlichen, betrachten wir stattdessen das einfachere Polynom $P_A(X) = \operatorname{tr}(X \cdot E_n - A) = nX - \operatorname{tr} A$. Es gilt $0 = P_A(A) = nA - \operatorname{tr} A \cdot E_n$ genau dann, wenn A ein Vielfaches der Einheitsmatrix ist, im allgemeinen also nicht. Die obige Heuristik würde aber immer $P_A(A) = 0$ liefern.

Beweis. Es sei zunächst $B \in M_n(R[X])$. Wir erinnern uns an die Adjunkte $\operatorname{adj} B \in M_n(R[X])$ aus Definition 4.20. Im Beweis der Cramerschen Regel 4.21 (1) haben wir gezeigt, dass

$$\operatorname{adj} B \cdot B = \det B \cdot E_n \in M_n(R[X]) \;.$$

Wir betrachten die spezielle Matrix $B = X \cdot E_n - A \in M_n(R[X])$ und erhalten

$$\operatorname{adj}\bigl(XE_n - A\bigr) \cdot \bigl(XE_n - A\bigr) = \det\bigl(XE_n - A\bigr) \cdot E_n = \chi_A(X)E_n \;.$$

Nach Definition 4.20 sind die Einträge der Adjunkten Determinanten von $(n-1)$-reihigen Untermatrizen, in diesem Fall also Polynome vom Grad $\le n-1$, da alle Matrixeinträge Grad ≤ 1 haben. Wir fassen die Koeffizienten von X^i jeweils zu einer Matrix $B_i \in M_n(R)$ zusammen und erhalten

$$\operatorname{adj}\bigl(XE_n - A\bigr) = \sum_{i=0}^{\infty} B_i \cdot (X^i E_n) \;,$$

wobei $B_i = 0$ für $i \ge n$. Außerdem seien $c_0, \dots, c_n$ die Koeffizienten von $\chi_A(X)$, und $c_i = 0$ für alle $i > n$. Dann gilt also

$$\sum_{i=0}^{\infty} B_i \cdot (X^i E_n) \cdot (XE_n - A) = \sum_{i=0}^{n} c_i X^i E_n \;.$$

Indem wir die Koeffizienten von X^i vergleichen, erhalten wir in $M_n(R)$ die Identitäten

$$B_{i-1} - B_i \cdot A = c_i E_n \qquad \text{für alle } i \ge 0 \;,$$

wobei $B_{-1} = 0$.

Wir berechnen $\chi_A(A)$ wie in Beispiel 5.28. Es folgt

$$\begin{aligned}\chi_A(A) &= \sum_{i=0}^{\infty} c_i A^i = \sum_{i=0}^{\infty} c_i E_n \cdot A^i \\ &= \sum_{i=0}^{\infty} \bigl(B_{i-1} - B_i \cdot A\bigr)\, A^i = \sum_{i=1}^{\infty} B_{i-1}\, A^i - \sum_{i=0}^{\infty} B_i\, A^{i+1} = 0 \;. \qquad \square\end{aligned}$$

Wir arbeiten wieder über einem Körper $\Bbbk$. Es sei $A \in M_n(\Bbbk)$ eine Matrix, und es sei $\operatorname{ev}(\,\cdot\,, A)\colon \Bbbk[X] \to M_n(\Bbbk)$ der zugehörige unitäre Ringhomomorphismus wie in Beispiel 5.28. Wie in Definition 2.51 definieren wir seinen *Kern* durch

$$\ker\bigl(\operatorname{ev}(\,\cdot\,, A)\bigr) = \bigl\{\, P \in \Bbbk[X] \bigm| P(A) = 0 \,\bigr\} \;.$$

Unter den Polynomen in $\ker\bigl(\operatorname{ev}(\,\cdot\,, A)\bigr) \setminus \{0\}$ gibt es ein P vom kleinstmöglichen Grad. Man überlegt sich leicht, dass mit P auch $aP \in \ker\bigl(\operatorname{ev}(\,\cdot\,, A)\bigr)$ für alle $a \in \Bbbk$, also dürfen wir P normieren.

5.30. Definition. Es sei $\Bbbk$ ein Körper, $n \in \mathbb{N}$ und $A \in M_n(R)$. Dann ist das *Minimalpolynom* $\mu_A(X) \in \Bbbk[X] \setminus \{0\}$ das normierte Polynom vom kleinstmöglichen Grad, so dass $\mu_A(A) = 0$. Entsprechend definieren wir μ_F für $F \in \operatorname{End}_\Bbbk V$, wenn V ein endlichdimensionaler $\Bbbk$-Vektorraum ist.

Das Minimalpolynom ist tatsächlich eindeutig bestimmt und damit wohldefiniert. Denn seien P, $Q \in \ker\bigl(\operatorname{ev}(\,\cdot\,, A)\bigr)$ normiert und von kleinstem Grad, dann ist $P - Q$ ein Polynom von kleinerem Grad nach Bemerkung 5.12 (2), da P und Q den gleichen Leitkoeffizienten haben. Wäre $P \neq Q$, so könnten wir $P-Q$ normieren und erhielten ein normiertes Polynom von kleinerem Grad in $\ker\bigl(\operatorname{ev}(\,\cdot\,, A)\bigr)$, was nach Wahl von P und Q aber ausgeschlossen ist. Also gilt $P = Q = \mu_A$.

5.31. Folgerung. *Es sei $\Bbbk$ ein Körper, $n \in \mathbb{N}$ und $A \in M_n(R)$. Dann gilt*

$$\mu_A \mid \chi_A \;.$$

Beweis. Wir dividieren χ_A durch μ_A mit Rest und erhalten

$$\chi_A = S \cdot \mu_A + T \;,$$

mit S, $T \in \Bbbk[X]$ und $\deg T < \deg \mu_A$. Einsetzen von A liefert nach dem Satz von Cayley-Hamilton, dass

$$T(A) = \chi_A(A) - S(A) \cdot \mu_A(A) = 0 \;.$$

Da $\deg T < \deg \mu_A$, ist das nach der obigen Definition von μ_A nur möglich, wenn $T = 0$. $\square$

5.32. Bemerkung. Sei V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$. Das Minimalpolynom μ_F lässt sich leider nicht so leicht berechnen wie das charakteristische Polynom.

(1) Wegen Folgerung 5.31 kommt für μ_F nur ein Teiler von χ_F in Frage. Wir werden später sehen, dass in $\Bbbk[X]$ der Satz von der eindeutigen Primfaktorzerlegung gilt, so dass χ_F nur endlich viele Teiler hat.
(2) Es sei $U \subset V$ ein *invarianter Unterraum,* das heißt, es gilt $F(U) \subset U$. Dann gilt nach Induktion über n auch $F^n(U) = F(F^{n-1}(U)) \subset U$. Für Polynome $P \in \Bbbk[X]$ folgt dann insbesondere

$$P(F)|_U = P(F|_U) \in \operatorname{End}_{\Bbbk} U \ .$$

Für das Minimalpolynom muss also auch $\mu_F(F|_U) = 0$ gelten.
(3) Jetzt nehmen wir an, dass V in eine direkte Summe $V = U \oplus W$ invarianter Unterräume U und W zerfällt. Nach Proposition 2.58 (2) lässt sich jeder Vektor $v \in V$ auf eindeutige Weise zerlegen als $v = u+w$ mit $u \in U$ und $w \in W$. Also gilt

$$P(F)(v) = P(F)(u + w) = P(F)(u) + P(F)(w) \ ,$$

und das verschwindet genau dann für alle $v \in V$, wenn $P(F)(u) = 0$ und $P(F)(w) = 0$ für alle $u \in U$ und alle $w \in W$. Das Minimalpolynom μ_F ist das normierte Polynom P vom kleinstmöglichen Grad, so dass $P(F|_U) = P(F|_W) = 0$, also gewissermaßen das „kleinste gemeinsame Vielfache“ von $\mu_{F|_U}$ und $\mu_{F|_W}$. Wir können es berechnen als

$$\mu_F = \operatorname{kgV}\bigl(\mu_{F|_U}, \mu_{F|_W}\bigr) = \mu_{F|_U} \cdot \mu_{F|_W} \ / \ \operatorname{ggT}\bigl(\mu_{F|_U}, \mu_{F|_W}\bigr) \ .$$

Als Beispiel betrachte die Matrix

$$A = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \ .$$

Mit Folgerung 4.17 (2) berechnet man

$$\chi_A(X) = (X - \lambda)^3 \ .$$

Sei (e_1, e_2, e_3) die Standardbasis von $\Bbbk^3$, siehe Beispiel 2.31, dann zerlegen wir

$$\Bbbk^3 = U \oplus W = \langle e_1, e_2\rangle \oplus \langle e_3\rangle = \Bbbk^2 \oplus \Bbbk \ ;$$

beide Unterräume sind A-invariant. Es gilt $\chi_{A|_U} = (X - \lambda)^2$, und $\mu_{A|_U} \mid \chi_{A|_U}$. Das einzige normierte Polynom vom Grad 0 ist 1 und wirkt wie $\mathrm{id}_U \neq 0$. Das einzige normierte Polynom vom Grad 1, das $(X - \lambda)^2$ teilt, ist $X - \lambda$ wegen Folgerung 5.15, und es gilt

$$(A|_U - \lambda) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq 0 \ .$$

Also folgt $\mu_{A|_U} = \chi_{A|_U} = (X - \lambda)^2$. Nun rechnet man noch nach, dass $(A - \lambda)^2|_W = 0$, und erhält schließlich

$$\mu_A(X) = (X - \lambda)^2 \ .$$

Weitere Beispiele sehen Sie in den Übungen.

5.33. Lemma. *Es sei V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$. Es sei $\lambda \in \Bbbk$. Dann sind die folgenden Aussagen äquivalent:*

$$\lambda \text{ ist Eigenwert von } F\ ; \tag{1}$$

$$\ker\bigl(\lambda \operatorname{id}_V - F\bigr) \neq \{0\}\ ; \tag{2}$$

$$\lambda \operatorname{id}_V - F \notin \operatorname{Aut}_\Bbbk V\ ; \tag{3}$$

$$\chi_F(\lambda) = \det\bigl(F - \lambda \operatorname{id}_V\bigr) = 0\ ; \tag{4}$$

$$\mu_F(\lambda) = 0\ . \tag{5}$$

Beweis. Bereits am Anfang von Abschnitt 5.2 haben wir die Äquivalenz von (1)–(4) überprüft.

Zu „(5) $\Longrightarrow$ (4)“ benutzen wir die Folgerungen 5.15 (1) und 5.31, und erhalten

$$\mu_F(\lambda) = 0 \qquad \Longrightarrow \qquad (X - \lambda) \mid \mu_F \mid \chi_F \qquad \Longrightarrow \qquad \chi_F(\lambda) = 0\ .$$

Zu „(1) $\Longrightarrow$ (5)“ sei $v \in V \setminus \{0\}$ ein Eigenvektor zum Eigenwert λ, dann ist der eindimensionale Unterraum $\langle \lambda \rangle \subset V$ invariant unter F, und es gilt $F|_{\langle v \rangle} = \lambda \operatorname{id}_V$. Wie in Bemerkung 5.32 (2) erhalten wir

$$0 = \mu_F(F)|_{\langle v \rangle} = \mu_F(\lambda \operatorname{id}_{\langle v \rangle}) = \mu_F(\lambda) \operatorname{id}_{\langle v \rangle} \qquad \Longrightarrow \qquad \mu_F(\lambda) = 0\ . \qquad \square$$

Zur Motivation für die folgenden Abschnitte fügen wir noch einen interessanten Satz an, den wir aber erst später vollständig beweisen können. Wir erinnern uns an den Begriff der Diagonalisierbarkeit aus Definition 5.3.

5.34. Satz (Diagonalisierbarkeit)**.** *Es sei $\Bbbk$ ein Körper und V ein endlichdimensionaler $\Bbbk$-Vektorraum. Dann sind für einen Endomorphismus $F \in \operatorname{End}_\Bbbk V$ die folgenden Aussagen äquivalent.*

(1) *F ist diagonalisierbar;*
(2) *V besitzt eine Basis aus Eigenvektoren von F;*
(3) *V zerfällt in eine direkte Summe von Eigenräumen von F;*
(4) *Das charakteristische Polynom zerfällt in Linearfaktoren, und für jeden Eigenwert $\lambda \in \Bbbk$ stimmen algebraische und geometrische Vielfachheit überein, das heißt, es gilt*

$$\operatorname{ord}_\lambda \chi_A = \dim \ker\bigl(\lambda \operatorname{id}_V - F\bigr)\ ;$$

(5) *Das Minimalpolynom zerfällt in paarweise verschiedene Linearfaktoren.*

Die Diagonaleinträge sind gerade die Eigenwerte von F, und kommen entsprechend ihrer Vielfachheit oft vor. Insbesondere ist die Diagonalmatrix bis auf Reihenfolge der Diagonaleinträge eindeutig durch F bestimmt.

Eine hinreichende, aber nicht notwendige Bedingung für Diagonalisierbarkeit hatten wir in Folgerung 5.6 (2) bereits kennengelernt.

Einen ähnlichen Satz über Trigonalisierbarkeit formulieren wir später. Dann zeigen wir auch, dass ein ganz bestimmter einfacherer Typ von Dreiecksmatrizen völlig ausreicht.

Beweis. Wir verschieben den Beweis, dass (5) eine der drei anderen Aussagen impliziert, auf später. Die Äquivalenz von (1)–(3) haben wir in Proposition 5.4 und Folgerung 5.6 (4) gezeigt, und daraus ergibt sich auch die Eindeutigkeit der Diagonalmatrix bis auf Reihenfolge der Einträge.

Zu „(3) $\Longrightarrow$ (4)“ sei V_λ der Eigenraum zum Eigenwert λ, dann gilt

$$\chi_{F|_{V_\lambda}} = \det\bigl(X \operatorname{id}_{V_\lambda} - F|_{V_\lambda}\bigr) = (X - \lambda)^{\dim V_\lambda} .$$

Aus Folgerung 4.17 (1) folgt induktiv über die Anzahl der verschiedenen Eigenwerte, dass

$$\chi_F = \prod_{\lambda \in \Bbbk} (X - \lambda)^{\dim V_\lambda} ,$$

also stimmen algebraische und geometrische Vielfachheit für alle λ überein, und χ_F zerfällt in Linearfaktoren.

Zu „(4) $\Longrightarrow$ (2)“ sei

$$U = \sum_{\lambda \in \Bbbk} \ker\bigl(\lambda \operatorname{id}_V - F\bigr) \subset V$$

die Summe der Eigenräume. Nach Folgerung 5.6 (3) ist diese Summe direkt. Indem wir Basen der einzelnen Eigenräume zu einer Basis von U zusammensetzen, sehen wir wegen (4), dass

$$\dim U = \sum_{\lambda \in \Bbbk} \dim \ker\bigl(\lambda \operatorname{id}_V - F\bigr) = \sum_{\lambda \in \Bbbk} \operatorname{ord}_\lambda \chi_F = \dim V .$$

es folgt $U = V$, und wir erhalten eine Basis von V aus Eigenvektoren. $\square$

KAPITEL 6

Endomorphismen und Normalformen

In diesem Kapitel wollen wir Normalformen für Endomorphismen endlichdimensionaler Vektorräume finden. Dabei betrachten wir zwei Endomorphismen $F \in \operatorname{End}_{\Bbbk} V$ und $G \in \operatorname{End}_{\Bbbk} W$ als isomorph, wenn das Diagramm in Bemerkung 5.8 kommutiert. Als vollständige Invariante erhalten wir eine Reihe von Polynomen, die eng mit dem Minimalpolynom aus Abschnitt 5.3 verwandt sind. Bei der Wahl einer geeigneten Normalform in $M_n(\Bbbk)$ werden wir eine gewisse Auswahl haben. Wir können folgende Kriterien anlegen.

(1) Die Normalform sollte möglichst eindeutig sein, genauer, eindeutig bis auf die Reihenfolge gewisser elementarer Bausteine. Das ist etwas schwächer als in Bemerkung 3.17 und entspricht in etwa der Formulierung in Satz 5.34.
(2) Die Normalform sollte so einfach wie möglich sein. Wenn F also diagonalisierbar oder trigonalisierbar ist, dann sollte die Normalform eine Diagonal- beziehungsweise Dreiecksmatrix sein.
(3) Die Normalform sollte die Struktur von F möglichst gut wiedergeben. Wenn V invariante Unterräume enthält oder gar in eine direkte Summe invarianter Unterräume zerfällt, möchten wir das der Matrix in Normalform ansehen. Sie sollte also eine entsprechende Blockgestalt haben.

Wir werden uns diesem Ideal im Laufe des Kapitels immer mehr annähern. Am Ende konstruieren wir die sogenannte „rationale Normalform“, die alle diese Eigenschaften hat. Auf dem Weg dahin lernen wir einiges über Ringe und Ideale, unter anderem den Satz von der eindeutigen Primfaktorzerlegung und den chinesischen Restsatz. Außerdem betrachten wir die Smith-Normalform für Homomorphismen zwischen freien R-Moduln.

6.1. Euklidische Ringe und Hauptidealringe

In diesem Kapitel führen wir einige Begriffe aus der Ringtheorie ein. Ziel ist es, im nächsten Abschnitt die Struktur von endlich erzeugten Moduln über dem Ring $\mathbb{Z}$ der ganzen Zahlen und dem Polynomring $\Bbbk[X]$ über einem Körper besser zu verstehen. Ein Modul über $\mathbb{Z}$ ist eine abelsche Gruppe nach Aufgabe 2 von Blatt 7 zur Linearen Algebra I. Sei V ein $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$, dann wird V zu einem $\Bbbk[X]$-Modul mit

$$P \cdot v = P(F) \cdot v .$$

Im letzten Abschnitt dieses Kapitels nutzen wir das aus, um Endomorphismen besser zu verstehen und Normalformen dafür herleiten.

Für die folgende Definition wollen wir einen kommutativen Ring R als Modul über sich selbst auffassen wir in Beispiel 2.21 (1). In Definition 2.24 haben wir den von einer Teilmenge $E \subset M$ erzeugten Untermodul $\langle E\rangle \subset M$ eingeführt. In Definition 2.44 haben wir Axiome für Untermoduln aufgestellt.

6.1. Definition. Es sei R ein kommutativer Ring mit Eins. Ein *Ideal* von R ist ein R-Untermodul I von R. Sei $E \subset R$ eine Teilmenge, dann ist das *von E erzeugte Ideal* (E) gerade der von E erzeugte Untermodul $\langle E\rangle \subset R$. Ein Ideal, das von einem einzigen Element $a \in R$ erzeugt wird, ist ein *Hauptideal.*

Wenn die Menge $E = \{a_1, \dots, a_k\} \subset R$ endlich ist, schreiben wir einfach $(a_1, \dots, a_k)$ für $\langle E\rangle$. Für das von a erzeugte Hauptideal schreiben wir also (a).

6.2. Beispiel.

(1) In jedem kommutativem Ring R mit Eins gibt es zwei triviale Ideale, nämlich
$$(0) = \{0\} \qquad \text{und} \qquad (1) = R\;.$$
Denn $\{0\}$ ist offensichtlich ein Ideal, und wenn I ein Ideal mit $1 \in I$ ist, folgt für alle $r \in R$, dass
$$r = 1 \cdot r \in I\;.$$
(2) In einem Körper $\Bbbk$ gibt es nur die trivialen Ideale. Denn sei $I \neq \{0\}$ ein Ideal, dann gibt es ein Element $0 \neq k \in I$, somit liegt $1 = k \cdot k^{-1} \in I$, also gilt $I = \Bbbk$.

(3) Das Ideal $I = (2, X) \in \mathbb{Z}[X]$ ist kein Hauptideal (Übung).

6.3. Bemerkung. Ideale treten zum Beispiel im Zusammenhang mit Ringhomomorphismen und Quotienten auf.

(1) Es sei $S \subset R$ ein Unterring. Nach Übung 2(b) von Blatt 8 zur Linearen Algebra I induziert die Multiplikation auf R genau dann eine Multiplikation auf dem Quotienten R/S, wenn S ein Ideal ist. In diesem Fall ist die Quotientenabbildung $R \to R/S$ ein Ringhomomorphismus. In Beispiel 2.9 haben wir den Ring $\mathbb{Z}/n\mathbb{Z}$ konstruiert, dabei ist $n\mathbb{Z} \subset \mathbb{Z}$ gerade das von n erzeugte Hauptideal (n).

(2) Es sei $f\colon R \to S$ ein Ringhomomorphismus, dann ist $\ker f \subset R$ ein Ideal. Wir überprüfen die Untermodulaxiome, dazu seien a, $b \in \ker f$ und $r \in R$:
$$\begin{array}{rcrcrl} & & f(0) = 0 & \Longrightarrow & 0 \in \ker f\;, & \text{(U1)}\\ \text{(H1)} & \Longrightarrow & f(a+b) = f(a) + f(b) = 0 & \Longrightarrow & a + b \in \ker f\;, & \text{(U2)}\\ \text{(H2)} & \Longrightarrow & f(ar) = f(a) \cdot f(r) = 0 & \Longrightarrow & a \cdot r \in \ker f\;. & \text{(U3)} \end{array}$$
Wegen (1) ist jedes Ideal $I \subset R$ Kern eines Ringhomomorphismus, nämlich der Quotientenabbildung $R \to R/I$.

(3) In Analogie zu Folgerung 2.54 gilt auch für Ringe ein Homomorphiesatz. Dabei zerlegen wir einen Ringhomomorphismus $f\colon R \to S$ wie folgt.

$$R \twoheadrightarrow R/\ker f \xrightarrow{\cong} \operatorname{im} f \hookrightarrow S\ .$$

(4) Auf der anderen Seite sind Bilder von Ringhomomorphismen oftmals keine Ideale. Nach Proposition 5.27 dürfen wir X^2 in Polynome einsetzen und erhalten einen Homorphismus $F\colon R[X] \to R[X]$ mit

$$\operatorname{im} F = \left\{ P = \sum_{i=0}^{\deg P} a_i X^i \in R[X] \;\middle|\; a_i = 0 \text{ für alle ungeraden } i \in \mathbb{N} \right\} .$$

Dann gilt $1 \in \operatorname{im} F$, aber $X \notin \operatorname{im} F$ im Widerspruch zu Beispiel 6.2 (1).

6.4. Definition. Es sei R ein kommutativer Ring mit Eins, und es seien r, $s \in R$. Dann ist r ein *Teiler* von s, kurz $r \mid s$, wenn ein $t \in R$ mit $rt = s$ existiert. Andernfalls schreiben wir $r \nmid s$.

Die Elemente r und s heißen *assoziiert,* wenn sowohl $r \mid s$ als auch $s \mid r$ gilt. Ein Element $r \in R$ heißt *Einheit,* wenn $r \mid 1$. Die Menge aller Einheiten heißt die *Einheitengruppe* $R^\times$ von R.

Die Einheitengruppe $R^\times$ haben wir im Zusammenhang mit der ersten Cramerschen Regel in Folgerung 4.21 eingeführt. In den Übungen sehen Sie, dass $R[X]^\times = R^\times$, wenn R Integritätsbereich ist.

Jede ganze Zahl $n \in \mathbb{Z}$ ist assoziiert zu $|n| \in \mathbb{N}$. Sei $\Bbbk$ ein Körper, dann ist jedes Polynom $P \in \Bbbk[X] \setminus \{0\}$ assoziiert zu genau einem normierten Polynom.

6.5. Bemerkung. Teilbarkeit lässt sich mit Hilfe von Idealen ausdrücken: es gilt

$$r \mid s \qquad\Longleftrightarrow\qquad s \in (r) \qquad\Longleftrightarrow\qquad (s) \subset (r)\ . \tag{1}$$

Denn $r \mid s$ bedeutet, dass $s = rt$ für ein $t \in R$, also $s \in (r)$. Für die Umkehrung benutzen wir, dass

$$(r) = \left\{\, r \cdot t \;\middle|\; t \in R \,\right\} ,$$

siehe Definition 2.24. Mit s liegen auch alle Vielfachen von s in (r), also $(s) \subset (r)$, und aus $s \in (s) \subset (r)$ folgt $s \in (r)$.

Sei jetzt R ein Integritätsbereich. Für r, $s \in R$ folgt

$$(r) = (s) \quad\Longleftrightarrow\quad r \mid s \text{ und } s \mid r \quad\Longleftrightarrow\quad s = ru \text{ für ein } u \in R^\times\ . \tag{2}$$

Die erste Äquivalenz folgt aus (1). Zur zweiten nehmen wir als erstes an, dass u, $v \in R$ mit $s = ru$ und $r = sv$ existieren. Dann gilt $r = ruv$, also $0 = r(1 - uv)$. Falls $r = 0$, folgt $s = 0 = r \cdot 1$. Ansonsten gilt $uv = 1$ wegen Nullteilerfreiheit, so dass $u \in R^\times$. Die Rückrichtung folgt, da $r = su^{-1}$, wenn $u \in R^\times$.

6.6. Definition. Es sei R ein kommutativer Ring mit Eins. Eine Funktion $d\colon R \setminus \{0\} \to \mathbb{N}$ heißt *Gradfunktion,* wenn für alle p, $q \in R$ mit $q \neq 0$ Elemente s, $t \in R$ existieren, so dass $p = qs + t$ und entweder $t = 0$ oder $d(t) < d(q)$.

Ein *Euklidischer Ring* ist ein Integritätsbereich, für den eine Gradfunktion existiert. Ein *Hauptidealring* ist ein Integritätsbereich, in dem jedes Ideal ein Hauptideal ist.

Ein Euklidischer Ring ist also ein Integritätsbereich, in dem Division mit Rest möglich ist — Eindeutigkeit ist nicht gefordert. Auch die Gradfunktion ist nicht eindeutig bestimmt. Wenn wir eine Gradfunktion d gefunden haben, nennen wir $d(r)$ einfach den *Grad* von r.

6.7. Beispiel. Die wichtigsten Euklidischen Ringe kennen wir bereits.

(1) Jeder Körper ist Euklidischer Ring. Da die Division immer ohne Rest aufgeht, können wir die Gradfunktion beliebig wählen.
(2) Die ganzen Zahlen $\mathbb{Z}$ bilden einen Euklidischen Ring mit Gradfunktion $d(n) = |n|$. Division mit Rest ist fast eindeutig:
$$5 = 3 \cdot 1 + 2 = 3 \cdot 2 + (-1) , \qquad \text{und} \qquad |2| , |-1| < |3| \ .$$
(3) Sei $\Bbbk$ ein Körper, dann ist der Polynomring $\Bbbk[X]$ ein Euklidischer Ring mit Gradfunktion $d(P) = \deg P$. Denn wegen Satz 5.13 können wir durch normierte Polynome mit Rest dividieren, und jedes Polynom $P \neq 0$ lässt sich normieren, indem man alle Koeffizienten durch den Leitkoeffizienten teilt.
(4) Wenn R kein Körper ist, ist $R[X]$ kein Euklidischer Ring. Das folgt aus Beispiel 6.2 (3) und Proposition 6.8 unten.

In den Beispielen (2) und (3) existieren Algorithmen, um die Division mit Rest durchzuführen. Dadurch lassen sich manche der Berechnungen, die wir im Folgenden durchführen werden, auch von einem Computeralgebrasystem erledigen.

6.8. Proposition. *Jeder Euklidische Ring ist ein Hauptidealring.*

Der Beweis verläuft ähnlich wie der Beweis der Eindeutigkeit des Minimalpolynoms nach Definition 5.30. In der Tat ist ja $\Bbbk[X]$ ein Euklidischer Ring, und wegen Bemerkung 6.3 (2) ist $\ker\bigl(\mathrm{ev}(\,\cdot\,, A)\bigr) \subset \Bbbk[X]$ ein Ideal, und zwar das Hauptideal (μ_A).

Beweis. Es sei R ein Euklidischer Ring mit Gradfunktion d und es sei $I \subset R$ ein Ideal. Falls $I = \{0\}$, ist $I = (0)$ ein Hauptideal. Andernfalls hat die Menge $\bigl\{\, d(a) \bigm| 0 \neq a \in I \,\bigr\}$ als Teilmenge von $\mathbb{N}$ ein kleinstes Element, wir finden also ein $a \in I \setminus \{0\}$ von kleinstem Grad. Sei $b \in I$, dann bestimme s, $t \in R$, so dass
$$b = as + t \qquad \text{und} \qquad d(t) < d(a) \ .$$
Aus den Untermodulaxiomen folgt $t \in I$, wegen $d(t) < d(a)$ also $t = 0$. Also gilt $b \in (a)$, und da das für alle $b \in I$ gilt, auch $I \subset (a)$. Wegen $a \in I$ gilt umgekehrt auch $(a) \subset I$, also $I = (a)$. □

6.9. Bemerkung. In jedem Hauptidealring gibt es größte gemeinsame Teiler. Seien etwa a, $b \in R$ in R, dann existiert nach Definition ein $c \in R$ mit $(a, b) = (c)$. Da a, $b \in (c)$ ist c ein Teiler von a und b nach Bemerkung 6.5. Sei $d \in$

R ein weiterer Teiler von a und b, dann folgt a, $b \in (d)$, also auch $(c) = (a,b) \subset (d)$. Somit ist d auch ein Teiler von c, und wir dürfen c als größten gemeinsamen Teiler auffassen. Aus diesem Grund schreiben manche Autoren kurz (a,b) für $\operatorname{ggT}(a,b)$. Man beachte, dass der größte gemeinsame Teiler nur bis auf Multiplikation mit einer Einheit eindeutig ist.

Aus $(a,b) = (c)$ folgt insbesondere, dass es r, $s \in R$ gibt mit

$$ar + bs = c\ .$$

Solche Elemente hatten wir mit dem Euklidischen Algorithmus in Satz 2.18 explizit konstruiert. Indem wir durch c teilen, sehen wir, dass

$$\frac{a}{c} \cdot r + \frac{b}{c} \cdot s = 1\ ,$$

insbesondere sind r und s teilerfremd, da $(r,s) = (1)$.

6.2. Die Smith-Normalform und invariante Faktoren

In diesem Kapitel wollen wir zunächst den Rangsatz (3.13) für Abbildungen zwischen freien Moduln über Euklidischen Ringen verallgemeinern. Anschließend beschreiben wir endlich erzeugte Moduln über solchen Ringen. Zu den Beispielen gehören endlich erzeugte abelsche Gruppen sowie Vektorräume mit einem fest vorgegebenen Endomorphismus. Auf diese Weise erhalten wir eine erste Normalform für Endomorphismen von Vektorräumen.

6.10. Satz (Smith-Normalform)**.** *Es sei R ein Hauptidealring und $A \in M_{m,n}(R)$. Dann existieren invertierbare Matrizen $S \in GL(m,R)$, $T \in GL(n,R)$ und $\operatorname{rg} A \in \mathbb{N}$, $a_1, \dots, a_{\operatorname{rg} A} \in R \setminus \{0\}$ mit $a_i \mid a_{i+1}$ für $i = 1, \dots, \operatorname{rg} A - 1$, so dass*

$$S \cdot A \cdot T = \begin{pmatrix} a_1 & 0 & & \cdots & & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ & \ddots & a_{\operatorname{rg} A} & 0 & \cdots & 0 \\ \vdots & & 0 & 0 & \cdots & 0 \\ & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}\ . \tag{1}$$

Der Rang $\operatorname{rg} A$ und die Elemente $a_1, \dots, a_{\operatorname{rg} A}$ sind bis auf Multiplikation mit einer Einheit eindeutig, die Matrizen S und T jedoch nicht.

Die Elemente $a_1, \dots, a_{\operatorname{rg} A}$ heißen die *Elementarteiler* oder *invarianten Faktoren* von A, und $\operatorname{rg} A$ der *Rang* von A. Nach Bemerkung 6.5 (2) können wir die Eindeutigkeitsaussage umformulieren: Die Ideale $(a_1), \dots, (a_k)$ sind eindeutig. Im Fall eines Körpers ist jedes Element $a \neq 0$ zu 1 assoziiert, und wir erhalten den Rangsatz 3.13.

Beweis. Wir beginnen mit der Existenz in dem Fall, dass R Euklidischer Ring ist. Die restlichen Teile des Beweises verschieben wir in den nächsten Abschnitt. Dazu geben wir ein Verfahren an, bei dem wir mit elementaren Zeilen- und Spaltenumformungen der Typen (1) und (3) die gegebene Matrix A in die gewünschte Form bringen. Wir erinnern uns, das elementare Zeilenumformungen der Multiplikation mit einer Elementarmatrix von links ensprechen, siehe Bemerkung 3.23. Analog beschreibt die Multiplikation mit einer Elementarmatrix von rechts eine elementare Spaltenumformung. Am Ende ist S das Produkt der Elementarmatrizen, die wir für die Zeilenumformungen gebraucht haben, und T ist entsprechend das Produkt der Elementarmatrizen für die Spaltenumformungen.

Es sei $d\colon R \to \mathbb{N}$ die Gradfunktion. Wir konstruieren S und T durch Induktion über $\min(m,n)$. Für $\min(m,n) = 0$ ist nichts zu tun, also nehmen wir an, dass $\min(m,n) \geq 1$.

(1) Wenn $A = 0$ gilt, sind wir fertig. Andernfalls bestimmen wir ein Element $a_{ij} \neq 0$, so dass $d(a_{ij}) \leq d(a_{k\ell})$ für alle $(k,\ell) \in \{1,\dots,n\}^2$. Falls $i \neq 1$, vertauschen wir die i-te Zeile mit der ersten durch Linksmultiplikation mit P_{1i}. Falls $j \neq 1$, vertauschen wir die j-te Spalte mit der ersten durch Rechtsmultiplikation mit P_{1j}. Jetzt hat das Element a_{11} unter allen nichtverschwindenden Matrixeinträgen den kleinsten Grad.

(2) Wenn a_{11} alle anderen Elemente in der ersten Spalte und in der ersten Zeile teilt, machen wir weiter mit Schritt (3). Falls $a_{11} \nmid a_{i1}$, bestimmen wir r, s mit $d(s) < d(a_{11})$, so dass
$$a_{i1} = r\, a_{11} + s\ .$$
Dann ziehen wir das r-fache der ersten Zeile von der i-ten ab durch Linksmultiplikation mit $E_{i1}(-r)$. Danach ist $d(a_{i1}) = d(s) < d(a_{11})$, und wir machen weiter mit Schritt (1). Falls $a_{11} \nmid a_{1j}$, verfahren wir entsprechend.

(3) Jetzt teilt a_{11} alle Einträge der ersten Zeile und der ersten Spalte. Falls a_{11} auch alle anderen Matrixeinträge teilt, machen wir weiter mit Schritt (4). Andernfalls existieren $i > 1$, $j > 1$ mit $a_{11} \nmid a_{ij} \neq 0$. Nach Annahme gilt $a_{11} \mid a_{i1}$. Wir subtrahieren das $\frac{a_{i1}}{a_{11}}$-fache der ersten Zeile von der i-ten, so dass an der Stelle $(i,1)$ jetzt eine Null steht. Dann addieren wir die i-te Zeile zur ersten. Dabei bleibt a_{11} unverändert. An der Stelle $(1,j)$ steht jetzt
$$a_{1j} + \left(a_{ij} - \frac{a_{i1}}{a_{11}}\, a_{1j}\right) = a_{ij} + \left(1 - \frac{a_{i1}}{a_{11}}\right) a_{1j}\ ,$$
das nun nicht mehr von a_{11} geteilt wird, und wir fahren fort mit (2).

(4) Jetzt teilt a_{11} alle anderen Matrixeinträge. Für jedes $i > 1$ ziehen wir das $\frac{a_{i1}}{a_{11}}$-fache der ersten Zeile von der i-ten ab. Anschließend beseitigen wir analog alle Einträge a_{1j} für $j > 1$ und erhalten eine Blockmatrix

der Form

$$S \cdot A \cdot T = \begin{pmatrix} a_{11} & 0 \\ 0 & A' \end{pmatrix} ,$$

und a_{11} teilt nach wie vor alle Einträge von A'. Nach Induktionsvoraussetzung existieren $S' \in GL(m-1, R)$ und $T' \in GL(n-1, R)$, so dass $S' \cdot A' \cdot T'$ die gewünschte Form hat. Die Einträge dieser neuen Matrix sind Linearkombinationen der Einträge von A' nach Definition 2.64 des Matrixproduktes, und werden daher nach wie vor von a_{11} geteilt. Dann hat insgesamt

$$\begin{pmatrix} 1 & 0 \\ 0 & S' \end{pmatrix} \cdot S \cdot A \cdot T \cdot \begin{pmatrix} 1 & 0 \\ 0 & T' \end{pmatrix}$$

die gesuchte Form.

Wir wollen uns noch überlegen, dass dieser Algorithmus tatsächlich in endlich vielen Schritten eine Lösung liefert. Solange wir uns in der Schleife der Schritte (1)–(3) bewegen, erzeugen wir in jedem Schleifendurchgang ein Element a_{ij} mit $d(a_{ij}) < d(a_{11})$, das wir anschließend auf den Platz $(1,1)$ bringen. Also nimmt die Zahl

$$d(A) = \min\bigl\{\, d(a_{ij}) \bigm| 1 \le i \le m, 1 \le j \le n, a_{ij} \neq 0 \,\bigr\}$$

in jedem Schritt um mindestens 1 ab. Wenn wir also mit einer Matrix $A \neq 0$ starten, gelangen wir nach höchstens $d(A)$ Schleifendurchläufen zu Schritt (4). Dann wiederholen wir den Induktionsschritt maximal $\min(m, n)$ mal und sind daher nach endlich vielen Schritten fertig. Damit ist die Existenzaussage bewiesen. $\square$

Genaues Hinsehen zeigt, dass der obige Algorithmus bei einer 1×2-Matrix gerade dem Euklidischen Algorithmus entspricht.

6.11. Beispiel. Wir führen das Verfahren anhand einer 3×2-Matrix über $\mathbb{Z}$ vor.

$$\begin{aligned}
\begin{pmatrix} 6 & 16 & 14 \\ \mathbf{4} & 4 & 6 \end{pmatrix} &\overset{(1)}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{4} & 4 & 6 \\ 6 & 16 & 14 \end{pmatrix} \\
&\overset{(2)}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 & 4 & 6 \\ \mathbf{2} & 12 & 8 \end{pmatrix} \\
&\overset{(1)}{=} \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}} \cdot \begin{pmatrix} \mathbf{2} & 12 & 8 \\ 4 & 4 & 6 \end{pmatrix} \\
&\overset{(4)}{=} \underbrace{\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & -20 & \mathbf{-10} \end{pmatrix} \cdot \begin{pmatrix} 1 & 6 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&\overset{(1)}{=} \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & \mathbf{10} & -20 \end{pmatrix} \cdot \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 6 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}
\end{aligned}$$

$$\overset{(4)}{=} \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & 10 & 0 \end{pmatrix} \cdot \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 6 & 4 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}}$$
$$= \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & 10 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 6 & 4 \\ 0 & -2 & -1 \\ 0 & 1 & 0 \end{pmatrix} .$$

Indem wir mit den Inversen der beiden quadratischen Matrizen multiplizieren, erhalten wir so wie im Satz

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 10 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 6 & 16 & 14 \\ 4 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 & 2 \\ 0 & 0 & 1 \\ 0 & -1 & -2 \end{pmatrix} .$$

Die invarianten Faktoren dieser Matrix sind also 2 und 10. Im letzten Schritt vom Typ (1) haben wir zusätzlich „normiert“, das heißt, das Vorzeichen des zweiten invarianten Faktor positiv gemacht. Für größere Matrizen oder für Matrizen über $\Bbbk[X]$ funktioniert das Verfahren genauso, ist allerdings entsprechend rechenaufwendiger.

Wir betrachten den Quotientenmodul $\mathbb{Z}^2/\operatorname{im} A$. Indem wir $\left(\binom{3}{2}, \binom{1}{1}\right)$ als Basis von $\mathbb{Z}^2$ wählen, erhalten wir einen Isomorphismus

$$\mathbb{Z}^2/\operatorname{im} A \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/10\mathbb{Z} .$$

Das ist ein Beispiel zum Satz 6.13 über invariante Faktoren.

6.12. Beispiel. Als nächstes betrachten wir Moduln über Hauptidealringen.

(1) Der Ring $\mathbb{Z}$ der ganzen Zahlen ist ein Euklidischer Ring nach Beispiel 6.7 (2), also erst recht ein Hauptidealring nach Proposition 6.8. Moduln über $\mathbb{Z}$ sind genau die abelschen Gruppen (Übung 2, Blatt 7 zur linearen Algebra I). Dabei setzt man in einer abelschen Gruppe $(A, +)$ für $a \in A$ und $n \in \mathbb{Z}$ gerade

$$a \,.\, n = \underbrace{a + \cdots + a}_{n \text{ Summanden}} .$$

(2) Der Polynomring $\Bbbk[X]$ über einem Körper ist ein Euklidischer Ring nach Beispiel 6.7 (3). Es sei V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$. Nach Beispiel 5.28 ist die Auswertungsabbildung

$$\operatorname{ev}(\,\cdot\,, F)\colon \Bbbk[X] \to \operatorname{End}_\Bbbk V \qquad \text{mit} \qquad P \mapsto P(F) \in \operatorname{End}_\Bbbk V$$

ein unitärer Ringhomomorphismus. Dadurch wird V zu einem unitären $\Bbbk[X]$-Modul (Übung) mit

$$P \cdot v = P(F)(v) .$$

6.13. Satz (Invariante Faktoren)**.** *Es sei R ein Hauptidealring und M ein endlich erzeugter unitärer R-Modul. Dann existieren $\operatorname{rg} M$, $k \in \mathbb{N}$ und $a_1, \dots, a_k \in R \setminus (\{0\} \cup R^\times)$ mit $a_i \mid a_{i+1}$ für $i = 1, \dots, k-1$ und ein Isomorphismus*

$$F\colon R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^k R/(a_i) \longrightarrow M .$$

Dabei sind $\operatorname{rg} M$ *und* k *eindeutig durch* M *bestimmt, und* $a_1, \dots, a_k$ *sind eindeutig bis auf Multiplikation mit einer Einheit. Der Untermodul*

$$\operatorname{Tor} M = F\biggl(\bigoplus_{i=1}^{k} R/(a_i)\biggr) \subset M$$

ist ebenfalls eindeutig bestimmt, nicht jedoch der Isomorphismus F*.*

Nach Bemerkung 6.5 (2) können wir die Eindeutigkeit der Elemente a_1, …, a_k auch so formulieren: Die Ideale (a_1), …, (a_k) sind eindeutig. Da $R/(a_i)$ keine Untermoduln eines fest vorgegebenen Moduls sind, müssten wir auf der rechten Seite eigentlich „$\coprod$“ anstelle von „$\bigoplus$“ schreiben; die Schreibweise „$\bigoplus$“ ist jedoch gebräuchlicher.

Der Untermodul $\operatorname{Tor} M \subset M$ heißt der *Torsionsuntermodul* von M, und die Zahl $\operatorname{rg} M$ heißt der *Rang* von M. Im Verlauf des Beweises sehen wir, dass

$$\operatorname{Tor} M = \bigl\{\, m \in M \bigm| \text{es gibt ein } r \in R \setminus \{0\} \text{ mit } m \,.\, r = 0 \,\bigr\} .$$

Sollte $M = \operatorname{Tor} M$ gelten, ist $\operatorname{rg} M = 0$, und M heißt ein *Torsionsmodul.* Aus dem Satz folgt, dass ein endlich erzeugter Modul M mit $\operatorname{Tor} M = \{0\}$ bereits frei ist.

Für endlich erzeugte freie R-Moduln hatten wir den Rang bereits in Bemerkung 4.23 betrachtet und seine Eindeutigkeit bewiesen. Für den Rang von Moduln über Hauptidealringen und linearen Abbildungen zwischen ihnen lassen sich ähnliche Formeln beweisen wie die Dimensionsformel für Vektorräume und lineare Abbildungen, zum Beispiel im Rangsatz 3.13.

Falls $R = \Bbbk$ ein Körper ist, gibt es nach Beispiel 6.2 (2) nur triviale Ideale, somit gilt $\operatorname{Tor} M = \{0\}$. Das bestätigt die Folgerung 3.5 aus den Steinitz-Sätzen, wonach jeder endlich erzeugte $\Bbbk$-Modul frei ist.

Beweis. Wir zeigen zunächst nur die Existenz der Zerlegung sowie die Eindeutigkeit des Torsionsuntermoduls und des Ranges. Den Rest verschieben wir auf später.

Als erstes bestimmen wir Abbildungen

$$(*) \qquad R^n \xrightarrow{\;A\;} R^m \xrightarrow{\;p\;} \!\!\!\!\to M$$

mit $\ker p = \operatorname{im} A$. Anschließend bringen wir A in Smith-Normalform und leiten daraus die behauptete Darstellung von M ab.

Es sei $(e_1, \dots, e_m)$ ein Erzeugendensystem von M. Dann existiert eine surjektive Abbildung $p\colon R^m \to M$ mit

$$p(r_1, \dots, r_m) = \sum_{i=1}^{m} e_i \,.\, r_i \,.$$

Als erstes zeigen wir durch Induktion über j, dass $\ker p|_{\langle e_1,\dots,e_j\rangle}$ ein freier R-Modul von endlichem Rang ist. Im Fall $j = 0$ ist $\ker p|_{\{0\}} = \{0\}$ frei, und wir sind fertig. Wir nehmen jetzt an, dass $\ker(p|_{\langle e_1,\dots,e_{j-1}\rangle})$ ein freier Modul mit Basis $(f_1, \dots, f_\ell)$ ist und betrachten die Menge

$$I = \left\{ r_j \in R \;\middle|\; \text{es existieren } r_1, \dots, r_{j-1} \in R, \text{ so dass } \sum_{i=1}^{j} e_i \,.\, r_i \in \ker p \right\} .$$

Dann ist $I \subset R$ ein Ideal, beispielsweise folgt aus r_i, $s_i \in I$ mit

$$\sum_{i=1}^{j} e_i \,.\, r_i \in \ker p \qquad \text{und} \qquad \sum_{i=1}^{j} e_i \,.\, s_i \in \ker p$$

bereits, dass

$$\sum_{i=1}^{j} e_i \,.\, (r_i + s_i) \in \ker p ,$$

also $r_j + s_j \in I$. Analog folgt $r_j \cdot t \in I$ für alle $t \in R$. Falls $I = \{0\}$, gilt $\ker p = \langle f_1, \dots, f_\ell \rangle$, und wir sind fertig.

Andernfalls existiert ein Erzeuger $a \in R \setminus \{0\}$ mit $I = (a)$, da R ein Hauptidealring ist. Wir bestimmen $a_1, \dots, a_{j-1} \in R$ und $a_j = a$ so, dass

$$f_{\ell+1} = \sum_{i=1}^{j} e_i \,.\, a_i \in \ker p .$$

Sei jetzt

$$v = \sum_{i=1}^{j} e_i \,.\, r_i \in \ker p ,$$

dann folgt $r_j \in I$, somit $r_j = a_j r$ für ein $r \in R$. Aber dann ist

$$v - f_{\ell+1} \,.\, r \in \ker p|_{\langle e_1,\dots,e_{j-1}\rangle} = \langle f_1, \dots, f_\ell \rangle ,$$

und es folgt $v \in \langle f_1, \dots, f_{\ell+1} \rangle$, also ist $(f_1, \dots, f_{\ell+1})$ ein Erzeugendensystem von $\ker p$. Sei auf der anderen Seite

$$\sum_{i=1}^{\ell+1} f_i \,.\, s_i = 0 .$$

Da $f_1, \dots, f_\ell \in \langle e_1, \dots, e_{j-1} \rangle$ und $f_{\ell+1} \notin \langle e_1, \dots, e_{j-1} \rangle$, folgt $s_{\ell+1} = 0$. Da f_1, $\dots$, f_ℓ linear unabhängig sind, folgt auch $s_1 = \dots = s_\ell = 0$, also sind $f_1, \dots, f_{\ell+1}$ linear unabhängig und bilden somit eine Basis von $\ker p|_{\langle e_1,\dots,e_j\rangle}$.

Wir konstruieren die Abbildung A in (*), indem wir die Standardbasisvektoren $e_1, \dots, e_n$ auf die oben konstruierten Basisvektoren $f_1, \dots, f_n$ von $\ker p$ abbilden. Dann gilt insbesondere

$$\operatorname{im} A = \ker p .$$

Indem wir Satz 6.10 auf die Matrix A anwenden, erhalten wir $S \in GL(m, R)$ und $T \in GL(n, R)$, so dass die Matrix $A' = S \cdot A \cdot T$ in Smith-Normalform ist:

$$\begin{tikzcd} R^n \arrow[r, "A"] & R^m \arrow[r, two heads, "p"] \arrow[d, "\cong"', "S"] & M \arrow[d, equal] \\ R^n \arrow[u, "T", "\cong"'] \arrow[r, "A'"] & R^m \arrow[r, two heads, "p'"] & M\ . \end{tikzcd}$$

Es sei $p' = p \circ S^{-1}$, so dass $\operatorname{im} A' = S(\operatorname{im} A) = S(\ker p) = \ker p'$. Aus dem Homomorphiesatz 2.54 folgt

$$M = \operatorname{im} p' \cong R^m / \ker p' = R^m / \operatorname{im} A' \ .$$

Da A' die Gestalt aus Satz 6.10 hat, folgt die obige Behauptung, falls keines der a_i eine Einheit ist. Insbesondere wird der Isomorphismus F im Satz von p induziert.

Falls unter den a_i Einheiten sind, folgt aus $a_i \mid a_{i+1}$ für $1 \leq i \leq k-1$, dass ein j existiert, so dass genau $a_1, \ldots, a_j \in R^\times$. In diesem Fall sind die ersten Summanden von $R^m / \operatorname{im} A'$ gerade $R/(a_1), \ldots, R/(a_i) = \{0\}$. Wir lassen in R^n und R^m jeweils die ersten j Basisvektoren weg und erhalten eine analoge Darstellung mit $A' \in M_{m-j,n-j}(R)$, was die obige Behauptung liefert.

Wir beweisen als nächstes die Eindeutigkeit des Torsionsuntermoduls $\operatorname{Tor} M$ und des Ranges $\operatorname{rg} M$. Dazu betrachten wir die eindeutig bestimmte Teilmenge

$$N = \bigl\{\, m \in M \bigm| \text{es gibt } r \in R \setminus \{0\} \text{ mit } m\,.\,r = 0 \,\bigr\} \ .$$

Es seien $v_1, \ldots, v_{\operatorname{rg} M}$ und $w_1, \ldots, w_k \in R$, dann betrachten wir

$$\bigl(v_1, \ldots, v_{\operatorname{rg} M}, [w_1], \ldots, [w_k]\bigr) \in R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^k R/(a_i) \ ,$$

sowie das Bild $m \in M$ unter dem Isomorphismus F. Es sei $r \neq 0$, dann gilt $m \cdot r = 0$ genau dann, wenn

$$\bigl(v_1, \ldots, v_{\operatorname{rg} M}, [w_1], \ldots, [w_k]\bigr)\,.\,r = \bigl(v_1 \cdot r, \ldots, v_{\operatorname{rg} M} \cdot r, [w_1 \cdot r], \ldots, [w_k \cdot r]\bigr) = 0 \ .$$

Da jeder Hauptidealring nach Definition 6.6 ein Integritätsbereich ist, gilt $v_i \cdot r = 0$ genau dann, wenn $v_i = 0$, und es folgt $N \subset \operatorname{Tor} M$. Auf der anderen Seite sei $v_1 = \cdots = v_{\operatorname{rg} M} = 0$ und $r = a_k$, dann gilt $a_i \mid a_k \mid w_i \cdot a_k$ für alle $i = 1, \ldots, k$. Es folgt

$$\bigl(0, \ldots, 0, [w_1], \ldots, [w_k]\bigr)\,.\,a_k = \bigl(0, \ldots, 0, [w_1 \cdot a_k], \ldots, [w_k \cdot a_k]\bigr) = 0 \ ,$$

und somit $N \supset \operatorname{Tor} M$, also $N = \operatorname{Tor} M$.

Da $\operatorname{Tor} M$ eindeutig ist, ist auch der Quotientenmodul $M / \operatorname{Tor} M$ eindeutig durch M bestimmt. In der obigen Darstellung ist

$$M \cong R^{\operatorname{rg} M} \oplus \operatorname{Tor} M \ ,$$

also ist $R^{\operatorname{rg} M}$ ein Komplement von $\operatorname{Tor} M$ in M. Aus Proposition 2.58 (3) folgt

$$R^{\operatorname{rg} M} \cong M / \operatorname{Tor} M \ .$$

Insbesondere ist $M / \operatorname{Tor} M$ ein endlich erzeugter freier R-Modul, und $\operatorname{rg} M = \operatorname{rg}(M / \operatorname{Tor} M)$ ist eindeutig bestimmt nach Bemerkung 4.23. □

Wir erinnern uns an die *zyklische Normalform* für einen Endomorphismus $F \in \operatorname{End}_{\Bbbk} V$ eines endlich-dimensionalen $\Bbbk$-Vektorraums V aus den Übungen. Wenn es einen Vektor $v \in V$ gibt, so dass

$$V = \langle v, F(v), F^2(v), \dots \rangle ,$$

dann bildet $\bigl(v, F(v), \dots, F^{\dim V - 1}(v)\bigr)$ eine Basis von V. In diesem Fall nennen wir $F \in \operatorname{End}_{\Bbbk} V$ einen *zyklischen Endomorphismus* und v einen *zyklischen Erzeuger* von (V, F).

6.14. Definition. Es sei $P(X) = X^n + c_1 X^{n-1} + \ldots + c_n \in \Bbbk[X]$ ein normiertes Polynom vom Grad $n \geq 1$, dann wird seine *Begleitmatrix* $M(P) \in M_n(\Bbbk)$ gegeben durch

$$M(P) = \begin{pmatrix} 0 & \cdots & 0 & 0 & -c_n \\ 1 & \ddots & \vdots & \vdots & \vdots \\ 0 & \ddots & 0 & 0 & -c_3 \\ \vdots & \ddots & 1 & 0 & -c_2 \\ 0 & \cdots & 0 & 1 & -c_1 \end{pmatrix} .$$

Bezüglich der obigen Basis wird der obige Endomorphismus F durch eben diese Begleitmatrix dargestellt, siehe Übung 4 von Blatt 2. Hierbei ist $P(X) = \chi_F(X) = \mu_F(X)$ sowohl das Minimalpolynom als auch das charakteristische Polynom von F, siehe Übung 4 von Blatt 3. Hieraus haben wir gefolgert, dass die Gestalt der Matrix nicht von der Wahl des zyklischen Erzeugers v abhängt, und dass zwei zyklische Endomorphismen $F \in \operatorname{End}_{\Bbbk} V$ und $G \in \operatorname{End}_{\Bbbk} W$ genau dann isomorph sind, wenn sie das gleiche charakteristische Polynom besitzen. Außerdem sieht man, dass es zu jedem normierten Polynom P einen zyklischen Endomorphismus F mit $\chi_F = P$ gibt.

6.15. Satz (Frobenius-Normalform)**.** *Es sei V ein endlich-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$. Dann existieren eindeutig bestimmte normierte Polynome $P_1, \dots, P_k \in \Bbbk[X]$ mit $P_i \mid P_{i+1}$ für $1 \leq i \leq k-1$, so dass F bezüglich einer geeigneten Basis als Block-Diagonalmatrix*

$$\begin{pmatrix} M(P_1) & 0 & \cdots & 0 \\ 0 & M(P_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M(P_k) \end{pmatrix} \tag{1}$$

aus den Begleitmatrizen der P_i dargestellt wird. Ausserdem gilt

$$\mu_F = P_k \qquad \textit{und} \qquad \chi_F = \prod_{i=1}^{k} P_i .$$

Beweis. Wir fassen V wie in Beispiel 6.12 (2) als unitären $\Bbbk[X]$-Modul auf, auf dem X wie F wirkt. Eine Basis von V als $\Bbbk$-Vektorraum ist ein Erzeugendensystem von V als $\Bbbk[X]$-Modul, da jede $\Bbbk$-Linearkombination auch eine $\Bbbk[X]$-Linearkombination ist. Also ist V endlich erzeugt.

Da $\Bbbk[X]$ ein Hauptidealring ist, ist V als $\Bbbk[X]$-Modul nach Satz 6.13 isomorph zu einem Modul der Form

$$(*) \qquad \Bbbk[X]^{\operatorname{rg} V} \oplus \bigoplus_{i=1}^{k} \Bbbk[X]/(P_i)\;,$$

wobei $P_i \mid P_{i+1}$ für $1 \le i \le k-1$. Wenn wir die Polynome $P_1, \dots, P_k$ normieren, sind sie durch die Modulstruktur eindeutig bestimmt.

Es gilt $\operatorname{rg} M = 0$, denn V ist endlich-dimensional, aber $\Bbbk[X] \cong \Bbbk^{(\mathbb{N})}$ ist als $\Bbbk$-Vektorraum unendlich-dimensional. Also ist V ein $\Bbbk[X]$-Torsionsmodul.

Wir betrachten jetzt einen Modul der Form $\Bbbk[X]/(Q)$ für ein normiertes Polynom

$$Q = X^n + c_1 X^{n-1} + \dots + c_n \in \Bbbk[X]$$

vom Grad n. Es sei $[P] \in \Bbbk[X]/(P)$ mit $P \in \Bbbk[X]$. Polynomdivision von P durch Q liefert einen eindeutigen Rest T vom Grad $\deg T < \deg Q$, siehe Satz 5.13. Es folgt $[P] = [T]$, so dass $\Bbbk[X]/(Q)$ als $\Bbbk$-Vektorraum isomorph ist zum Vektorraum der Polynome vom Grad kleiner als $\deg Q$. Eine $\Bbbk$-Basis von $\Bbbk[X]/(Q)$ wird gegeben durch $\big([1], [X], \dots, [X^{n-1}]\big)$. Multiplikation mit X wird bezüglich dieser Basis dargestellt durch die Begleitmatrix $M(Q)$, denn $X \cdot [X^i] = [X^{i+1}]$ für $i < n-1$, und

$$X \cdot [X^{n-1}] = [X^n] = [X^n - Q] = -c_1\,[X^{n-1}] - \dots - c_n\,[1]\;.$$

Zusammen mit (*) folgt, dass F insgesamt durch eine Matrix vom angegebenen Typ dargestellt werden kann.

Da V in F-invariante Unterräume $V_i = \Bbbk[X]/(P_i)$ zerfällt, und da nach den Vorüberlegungen und Übungen gerade

$$\mu_{F|_{V_i}} = \chi_{F|_{V_i}} = P_i$$

gilt, folgt zunächst mit Folgerung 4.17 (1), dass $\chi_F = P_1 \cdots P_k$ gerade das charakteristische Polynom ist. Aus $P_i \mid P_k$ für alle i folgt außerdem mit Bemerkung 5.32 (3), dass $\mu_F = P_k$. □

6.16. Bemerkung. Wir wollen überprüfen, inwieweit die Frobenius-Normalform unseren Anforderungen am Anfang des Kapitels genügt. Sobald wir die Eindeutigkeit bewiesen haben, ist Forderung (1) erfüllt.

Betrachten wir jedoch eine Diagonalmatrix mit paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_n \in \Bbbk$ und zugehörigen Eigenvektoren $e_1, \dots, e_n \in \Bbbk^n$, so können wir überprüfen, dass $v = e_1 + \dots + e_n$ ein zyklischer Erzeuger ist. Somit liefert die Frobenius-Normalform die Begleitmatrix

$$M(\chi_A) = M\big((X - \lambda_1) \cdots (X - \lambda_n)\big)\;,$$

und das ist keine Diagonalmatrix, falls $n \ge 2$. Somit ist Forderung (2) nicht erfüllt.

Wir bleiben beim obigen Beispiel. Da $\Bbbk^n$ in eine direkte Summe invarianter Unterräume $\langle e_i \rangle$ zerfällt, sollte die Frobenius-Normalform in n diagonale Blöcke der Größe 1×1 zerfallen. Da sie keine Diagonalmatrix ist, ist auch Forderung (3) nicht erfüllt.

Auf der anderen Seite liefert der Algorithmus zur Smith-Normalform, angewandt auf die Matrix $X\,E - n - A \in M_n(\Bbbk)$, gerade die invarianten Faktoren von A. Man kann die Frobenius-Normalform also effektiv berechnen (Übung).

6.3. Primfaktorzerlegung in Hauptidealringen

In diesem Abschnitt beweisen wir, dass in allen Hauptidealringen eine Primfaktorzerlegung ähnlich wie in den natürlichen Zahlen möglich ist. Als Anwendung definieren die Länge eines Moduls über einem Hauptidealring. Dieser Begriff verallgemeinert die Dimension eines Vektorraums über einem Körper auf eine andere Weise als der Begriff des Ranges. Durch den Vergleich von Längen von Moduln beweisen wir anschließend die Eindeutigkeit der Konstruktionen aus dem letzten Abschnitt.

6.17. Definition. Es sei R ein kommutativer Ring mit Eins. Ein Element $0 \neq a \in R \setminus R^\times$ heißt

(1) *irreduzibel* in R falls aus $a = r \cdot s$ folgt, dass $r \in R^\times$ oder $s \in R^\times$, und
(2) *prim* in R, falls aus $a \mid r \cdot s$ folgt, dass $a \mid r$ oder $a \mid s$.

Wir nennen einen Teiler von $a \in R$ *echt,* wenn er weder eine Einheit noch zu a assoziiert ist. Eine Element von R ist also genau dann irreduzibel, wenn es weder 0 noch eine Einheit ist und keine echten Teiler besitzt. Somit sind Primzahlen, wie Sie sie aus der Schule kennen, nach dieser Terminologie irreduzible Elemente von $\mathbb{Z}$. Die Zahl 12 ist nicht prim, denn $12 \mid 3 \cdot 4$, aber $12 \nmid 3$ und $12 \nmid 4$.

Wenn ein Element $a \in R$ prim oder irreduzibel ist, gilt das gleiche nach Definition 6.17 automatisch auch für alle assoziierten Elemente von R. Wegen Bemerkung 6.5 (2) könnten wir, falls R Integritätsbereich ist, also auch sagen, dass (a) prim beziehungsweise irreduzibel ist.

Sei R ein Integritätsbereich, dann sind Linearfaktoren $X - r$ prim in $R[X]$ nach Folgerung 5.15 (2). In $(\mathbb{Z}/6\mathbb{Z})[X]$ sind Linearfaktoren zwar irreduzibel, aber wegen Bemerkung 5.16 nicht prim.

6.18. Proposition. *Es sei R ein Integritätsbereich.*

(1) *Jedes Primelement von R ist irreduzibel in R.*
(2) *Sei R ein Hauptidealring, dann ist jedes irreduzible Element von R prim in R.*

Beweis. Zu (1) sei zunächst R ein Integritätsbereich und a prim. Dann ist a irreduzibel, denn

$$a = r \cdot s \quad \Longrightarrow \quad a \mid r \cdot s \quad \Longrightarrow a \mid r \text{ oder } a \mid s\ .$$

Gelte etwa $a \mid r$, dann sind a und r assoziiert, denn nach Voraussetzung gilt ja auch $r \mid a$. Nach Bemerkung 6.5 (2) ist s dann eine Einheit. Der Fall $a \mid s$ geht genauso. Also ist entweder $r \in R^\times$ oder $s \in R^\times$, also ist a irreduzibel.

Zu (2) sei R jetzt Hauptidealring und a irreduzibel. Es gelte $a \mid r \cdot s$, dann hat das Ideal (a, r) einen Erzeuger c, insbesondere gilt $a = cd$ für ein $d \in R$. Da a irreduzibel ist, ist entweder c oder d eine Einheit. Wenn d eine Einheit ist, folgt $a \mid ad^{-1} = c \mid r$. Wenn c eine Einheit ist, suchen wir u, $v \in R$, so dass $c = au + rv$ wie in Bemerkung 6.9. Nach Multiplikation mit sc^{-1} folgt $s = ac^{-1}su + c^{-1}rsv$. Da $a \mid rs$, erhalten wir $a \mid ac^{-1}su + c^{-1}rsv = s$. Also gilt entweder $a \mid r$ oder $a \mid s$, und a ist prim. $\square$

Im Folgenden sei $\mathcal{P}(R) \subset R$ eine Menge von Primelementen von R, die zu jedem Primelement genau ein assoziiertes Element enthält. Diese Menge ist im Allgemeinen nicht eindeutig, sondern muss für jeden Ring neu gewählt werden. Wir einigen uns darauf, dass $\mathcal{P}(\mathbb{Z})$ genau aus den positiven Primzahlen besteht, und dass $\mathcal{P}(\Bbbk[X])$ für jeden Körper $\Bbbk$ nur normierte Polynome enthält. Dadurch sind $\mathcal{P}(\mathbb{Z})$ und $\mathcal{P}(\Bbbk[X])$ eindeutig festgelegt.

6.19. Satz (Primfaktorzerlegung)**.** *Sei R ein Hauptidealring, dann lässt sich jedes Element $r \in R \setminus \{0\}$ schreiben als*

$$r = e \cdot p_1 \cdots p_k \,, \tag{1}$$

wobei $p_1, \dots, p_k \in \mathcal{P}(R)$ Primelemente sind, und $e \in R^\times$ eine Einheit. In dieser Zerlegung sind die Faktoren bis auf Reihenfolge eindeutig.

Typische Beispiele sind

$$\begin{aligned} -60 &= (-1) \cdot 2 \cdot 2 \cdot 3 \cdot 5 \\ \text{und} \qquad 2X^2 + 4X + 2 &= 2 \cdot (X+1) \cdot (X+1) \,. \end{aligned}$$

Beweis. Wir beweisen zunächst die Existenz durch Widerspruch. Sei also $r \in R$ ein Element, das sich nicht wie in (1) schreiben lässt. Dann lässt sich r auch nicht als endliches Produkt aus einer Einheit e und beliebigen Primelementen $a_1, \dots, a_k$ schreiben, denn dann könnten wir jedes a_i durch Multiplikation mit einer Einheit zu einem Element von $\mathcal{P}(R)$ abändern, und e entsprechend korrigieren.

Das Element $r_1 = r$ ist weder eine Einheit noch irreduzibel, also existieren s_1, $t_1 \in R \setminus R^\times$ mit $r_1 = s \cdot t$. Mindestens einer der beiden Faktoren lässt sich ebenfalls nicht wie in (1) schreiben, andernfalls könnten wir die beiden Zerlegungen multiplizieren und erhielten (indem wir die beiden Einheiten zusammenfassen) auch eine entsprechende Zerlegung von r_1. Sei etwa s das Element, für das keine solche Zerlegung existiert, dann machen wir mit $r_2 = s$ weiter.

So erhalten wir eine Folge $(r_i)_{i \in \mathbb{N}}$ von Elementen von R mit $r_{i+1} \mid r_i$ und $r_i \nmid r_{i+1}$ für alle $i \in \mathbb{N}$. Für die zugehörigen Ideale gilt nach Bemerkung 6.5 also

$$(r) = (r_1) \subsetneqq (r_2) \subsetneqq \dots \,.$$

Die Vereinigung dieser Ideale ist wieder ein Ideal (Übung), also existiert ein Erzeuger c, so dass

$$\bigcup_{i=1}^{\infty}(r_i) = (c)\ .$$

Nach Definition der Vereinigung existiert ein $i_0 \in \mathbb{N}$ mit $c \in (r_{i_0})$. Wir erhalten einen Widerspruch, denn

$$(c) \subset (r_{i_0}) \subsetneqq (r_{i_0+1}) \subsetneqq \dots \subset \bigcup_{i=1}^{\infty}(r_i) = (c)\ .$$

Mithin war unsere Annahme am Anfang falsch, und jedes Element $r \in R$ lässt eine Zerlegung wie in (1) zu.

Der Beweis der Eindeutigkeit verläuft analog zum Beweis von Proposition 5.18. Es sei

$$e \cdot p_1 \cdots p_k = f \cdot q_1 \cdots q_\ell\ , \tag{*}$$

mit $p_1, \dots, p_k, q_1, \dots, q_\ell \in \mathcal{P}(R)$ prim, also auch irreduzibel, und e und f seien Einheiten. Da p_1 prim ist und das rechte Produkt teilt, kann man nach Induktion schließen, dass $p_1 \mid q_j$ für ein $j \in \{1, \dots, \ell\}$ (es gilt $p_1 \nmid f$, da $f \mid 1$, aber $p_1 \nmid 1$, da $p_1 \notin R^\times$ nach Definition 6.17). Nun ist aber q_j irreduzibel und p_1 keine Einheit, also folgt $q_j = p_1$ nach Wahl von $\mathcal{P}$. Da Hauptideale nullteilerfrei sind, dürfen wir p_1 kürzen. Wir erhalten also eine Gleichung wie (*) mit je einem Faktor weniger auf beiden Seiten, wobei wir f durch fu ersetzen. Nach endlich vielen Schritten steht auf einer der beiden Seiten eine Einheit. Dann ist auch die andere Seite eine Einheit, und die Eindeutigkeit ist bewiesen. □

6.20. Beispiel. Es gibt keine effizienten Algorithmen zur Primfaktorzerlegung.

(1) Im Falle $R = \mathbb{Z}$ reicht es, sukzessive durch alle Primzahlen zu teilen, die nicht größer sind als die Quadratwurzel der verbleibenden Zahl. Beispielsweise gilt

$$999 = 3 \cdot 333 = 3 \cdot 3 \cdot 111 = 3 \cdot 3 \cdot 3 \cdot 37\ ,$$

denn $2 \nmid 999$, und $3,\ 5 \nmid 37$. Die Zahl 7 kann dann kein Teiler von 37 mehr sein, denn dann wäre $|37/7| < 7$, und wir hätten $37/7$ oder einen Teiler davon bereits gefunden. Dieser Algorithmus ist für sehr große Zahlen sehr rechenaufwändig; darauf beruhen kryptographische Verfahren wie der RSA-Code.

(2) Jedes Polynom $P \in \mathbb{C}[X]$ zerfällt nach Folgerung 5.22 in Linearfaktoren, und diese sind nach Folgerung 5.15 (2) prim. Es gibt aber keinen Algorithmus, der diese Linearfaktoren findet, falls $\deg P \geq 5$.

Beweis von Satz 6.10 (Fortsetzung). Wenn R nur ein Hauptidealring ist, ersetzen wir die Schritte (1) und (2) wie folgt.

(1') Falls $A = 0$ ist, sind wir fertig. Anderfalls stellen wir durch Vertauschen von Zeilen und Spalten sicher, dass $a_{11} \neq 0$. Diesen Schritt müssen wir nur einmal durchführen.

(2') Falls a_{11} alle anderen Elemente der ersten Zeile und der ersten Spalte teilt, machen wir mit (3) weiter, und kehren danach gegebenenfalls zu (2') zurück. Ansonsten gelte etwa $a_{11} \nmid a_{i1}$ für ein $i \geq 2$. Nach Bemerkung 6.9 existieren teilerfremde Elemente r, $s \in R$, mit

$$\operatorname{ggT}(a_{11}, a_{i1}) = a_{11}r + a_{i1}s \;.$$

Wir kürzen den größten gemeinsamen Teiler und erhalten

$$1 = ar + bs \qquad \text{mit } a,\, b \in R \;.$$

Also hat die Matrix

$$\begin{array}{c c}
 & \begin{array}{cccccccc} 1 & & & & i & & & \end{array} \\
\begin{array}{c} 1 \\ \\ \\ \\ i \\ \\ \\ \\ \end{array} &
\begin{pmatrix}
r & 0 & \cdots & 0 & s & 0 & \cdots & 0 \\
0 & 1 & & & & & & \\
\vdots & & \ddots & & & & & \vdots \\
0 & & & 1 & & & & \\
-b & & & & a & & & \\
0 & & & & & 1 & & \\
\vdots & & & & & & \ddots & 0 \\
0 & & & \cdots & & & 0 & 1
\end{pmatrix}
\end{array}$$

Determinante 1 und ist daher invertierbar nach der Cramerschen Regel 4.21. Durch Linksmultiplikation mit dieser Matrix erhalten wir an der Stelle $(1, 1)$ den größten gemeinsamen Teiler $ra_{11} + sa_{i1}$ von a_{11} und a_{i1}. Wir machen weiter mit (2').

Bei diesem Verfahren ersetzen wir in jedem Schritt (2') das Element a_{11} durch einen echten Teiler. Wie im Existenzbeweis zu Satz 6.19 folgt, dass das Verfahren nach endlich vielen Durchgängen abbricht, so dass wir dann zu Schritt (4) gelangen. Mit diesem Verfahren folgt die Existenzaussage, falls R nur ein Hauptidealring ist. □

6.21. Bemerkung. Wir betrachten die Primfaktorzerlegung etwas genauer.

(1) Wenn $p \in \mathcal{P}(R)$ insgesamt ℓ Mal in der Zerlegung (1) von $r \neq 0$ auftaucht, sagen wir, dass p in r die *Multiplizität* ℓ hat, und schreiben $\mu_p(r) = \ell$. Die Eindeutigkeitsaussage ist äquivalent dazu, dass $\mu_p(r)$ nicht von der Zerlegung (1) abhängt. Es gilt stets $\mu_p(r) \geq 0$. In Definition 5.17 hatten wir die Ordnung einer Nullstelle r eines Polynoms $P \in R[X]$ eingeführt. Falls $R = \Bbbk$ ein Körper ist, gilt

$$\operatorname{ord}_r P = \mu_{X-r}(P) \;.$$

(2) Für s, $t \in R \setminus \{0\}$ multiplizieren wir die Primfaktorzerlegungen und erhalten eine Primfaktorzerlegung von $s \cdot t$. Aus der obigen Eindeutigkeitsaussage folgt

$$\mu_p(st) = \mu_p(s) + \mu_p(t) \qquad \text{für alle } p \in \mathcal{P}(R) \;.$$

Für r, $s \in R \setminus \{0\}$ gilt insbesondere $r \mid s$ genau dann, wenn $\mu_p(r) \leq \mu_p(s)$ für alle $p \in \mathcal{P}(R)$. Zu „$\Longleftarrow$" bilden wir den Quotienten, indem wir wie im Beweis der Eindeutigkeitsaussage sukzessive alle Primfaktoren von r aus der Primfaktorzerlegung von s herausdividieren. Anschließend bleibt von r nur eine Einheit übrig, durch die wir ebenfalls teilen können.

Um die Eindeutigkeit in den Sätzen 6.10, 6.13 und 6.15 zu zeigen, benötigen wir einen weiteren Begriff.

6.22. Definition. Es sei R ein Hauptidealring und M ein R-Modul. Eine Kette von Untermoduln

$$0 = M_0 \subsetneqq M_1 \subsetneqq \cdots \subsetneqq M_\ell = M$$

heißt *maximal,* wenn es kein $i \in \{1, \dots, \ell\}$ und kein Untermodul $N \subset M$ gibt, so dass

$$M_{i-1} \subsetneqq N \subsetneqq M_i \;.$$

Wenn ein Modul eine maximale Kette von Untermoduln der Länge ℓ besitzt, heißt $\ell = \ell(M)$ die *Länge* von M. Wenn es keine maximale Kette gibt, hat M unendliche Länge, kurz $\ell(M) = \infty$.

Zur Wohldefiniertheit von $\ell(M)$ ist einiges zu beweisen. Alternativ kann man $\ell(M)$ als das Minimum der Länge einer maximalen Kette definieren, dann ist die Wohldefiniertheit klar. Andererseits sollte man dann im Folgenden auch noch zeigen, dass alle maximalen Ketten gleiche Länge haben, wenn $\ell(M) < \infty$ gilt, damit man in der Praxis nicht alle möglichen maximalen Ketten bestimmen und vergleichen muss.

6.23. Proposition. *Es sei R ein Hauptidealring und M ein R-Modul. Wenn M eine maximale Kette endlicher Länge besitzt, dann haben alle maximalen Ketten die gleiche Länge. Außerdem gilt*

$$\ell(M) = 0 \qquad \Longleftrightarrow \qquad M = 0\;; \tag{1}$$

$$\ell(M) < \infty \qquad \Longrightarrow \qquad M \text{ ist endlich erzeugt.} \tag{2}$$

$$\ell(R) = \begin{cases} 1 & \text{wenn } R \text{ ein Körper ist, und} \\ \infty & \text{sonst;} \end{cases} \tag{3}$$

$$\ell\big(R/(a)\big) = \sum_{p \in \mathcal{P}(R)} \mu_p(a)\;; \tag{4}$$

$$\ell(M) = \ell(N) + \ell(M/N) \qquad \text{für jeden Untermodul } N \subset M\;; \tag{5}$$

Ein Spezialfall von (5) liegt vor, wenn $M = N \oplus L$ eine direkte Summe ist, denn dann gilt $L = M/N$ nach Proposition 2.58 (3). Die Länge verhält sich daher ähnlich wie die Dimension von Vektorräumen, und stimmt mit ihr (und dem Rang) überein, wenn R ein Körper ist.

Wenn R kein Körper ist, sind Länge und Rang zwei unterschiedliche Verallgemeinerungen der Dimension. Wenn wir einen endlich erzeugten R-Modul wie in Satz 6.13 darstellen, folgt aus (3)–(5), dass

$$\ell\biggl(R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^k R/(a_i)\biggr) = \begin{cases} \sum_{i=1}^k \sum_{p \in \mathcal{P}(R)} \mu_p(a_i) & \text{falls } \operatorname{rg} M = 0 \text{, und} \\ \infty & \text{falls } \operatorname{rg} M > 0\;. \end{cases}$$

Insbesondere ist die Länge nur für Torsionsmodule interessant, während der Rang nur den freien Anteil $M/\operatorname{Tor} M$ des Moduls sieht. Wegen (2) können wir alle Längen zumindest theoretisch bestimmen.

Im Beispiel 6.11 hatten wir am Ende den Modul $\mathbb{Z}^2/\operatorname{im} A$ betrachtet. Er hat Länge 3, da er zu $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/10\mathbb{Z}$ isomorph ist.

Beweis. Aus der obigen Vorüberlegung folgt bereits die Wohldefiniertheit von $\ell(M)$, sobald wir (2)–(5) gezeigt haben.

Sei M ein R-Modul. Die einzige Kette der Länge Null ist $\{0\} = M_0 = M$, also folgt (1).

Sei M ein Modul endlicher Länge, und sei

$$0 = M_0 \subsetneqq M_1 \subsetneqq \cdots \subsetneqq M_m = M$$

eine maximale Kette. Wähle $m_i \in M_i \setminus M_{i-1}$ für alle $i = 1, \dots, m$. Wir zeigen durch Induktion über i, dass $M_i = \langle m_1, \dots, m_i \rangle$. Für $i = 0$ ist nichts zu zeigen. Sei also $M_{i-1} = \langle m_1, \dots, m_{i-1} \rangle$, dann folgt

$$M_{i-1} \subsetneqq \langle m_1, \dots, m_i \rangle \subset M_i \; .$$

Da die Kette maximal ist, gilt $\langle m_1, \dots, m_i \rangle = M_i$. Somit ist auch $M = M_m$ endlich erzeugt, und (2) ist ebenfalls bewiesen.

Sei $R = \Bbbk$ ein Körper. Nach Beispiel 6.2 (2) sind $\{0\}$ und $\Bbbk$ die einzigen Untermoduln von $\Bbbk$, also ist die einzige maximale Kette gerade

$$\{0\} = M_0 \subsetneqq M_1 = \Bbbk \; .$$

Sei R kein Körper, dann existiert $0 \neq r \in R \setminus R^\times$, und r hat mindestens einen Primfaktor. Da alle Untermoduln von R Ideale sind, hat jede Kette die Gestalt

$$0 \subsetneqq (a_1) \subsetneqq \cdots \; .$$

Nach Bemerkung 6.21 (2) ist a_1 ein echter Teiler von $a_1 r$. Aus Bemerkung 6.5 folgt

$$0 \subsetneqq (ra_1) \subsetneqq (a_1) \; ,$$

also ist die obige Kette nicht maximal. Da es demnach keine endlichen maximalen Ketten gibt, folgt (3).

Es sei jetzt $M = R/(a)$ und $N \subset M$ ein Untermodul. Wir betrachten die Verkettung der Quotientenabbildungen

$$F \colon R \longrightarrow R/(a) = M \longrightarrow M/N \; .$$

Es sei b ein Erzeuger von $\ker F$. Dann gilt

$$N = \left\{ [r] \in R/(a) \;\middle|\; r \in (b) \right\} = (b)/(a) \; ,$$

denn es gilt $[r] \in N$ genau dann, wenn $r \in \ker F$. Insbesondere ist $(a) \subset (b)$. Sei umgekehrt $(b) \subset (a)$, dann ist $(b)/(a)$ ein Untermodul von M.

Also entsprechen Untermoduln von M genau Idealen von R, die (a) enthalten, wegen Bemerkung 6.5 also gerade den Teilern von a, jeweils bis auf Multiplikation mit einer Einheit. Jede Kette von Untermoduln von M ist also von der Form

$$0 = (a_n)/(a) \subsetneqq (a_{n-1})/(a) \subsetneqq \cdots \subsetneqq (a_0)/(a) = M \; ,$$

mit $a_0 = 1$, $a_n = a$, sowie $a_{i-1} \mid a_i$ und $a_i \nmid a_{i-1}$ für alle $i \in \{1, \dots, n\}$. Die Kette ist nach Bemerkung 6.21 (2) genau dann maximal, wenn die Primfaktorzerlegung von a_i jeweils genau einen Primfaktor mehr enthält als die von a_{i-1}. Insbesondere ist n dann die gewichtete Anzahl der Primfaktoren von a, und es folgt (4).

Es seien

$$\begin{aligned} &0 = N_0 \subsetneqq N_1 \subsetneqq \cdots \subsetneqq N_n = N \\ \text{und} \qquad &0 = L_0 \subsetneqq L_1 \subsetneqq \cdots \subsetneqq L_\ell = M/N \end{aligned}$$

maximale Ketten, und es sei $p\colon M \to M/N$ die Quotientenabbildung, dann erhalten wir eine Kette von Untermoduln

$$0 = N_0 \subsetneqq \cdots \subsetneqq N_n = N = p^{-1}(L_0) \subsetneqq \cdots \subsetneqq p^{-1}(L_\ell) = M ,$$

also gilt $\ell(M) \geq n + m = \ell(N) + \ell(M/N)$.

Andererseits sei

$$0 = M_0 \subsetneqq M_1 \subsetneqq \cdots \subsetneqq M_m = M$$

eine maximale Kette und $1 \leq i \leq m$. Wir behaupten, dass entweder

$$M_{i-1} \cap N \subsetneqq M_i \cap N \subset N \qquad \text{oder} \qquad p(M_{i-1}) \subsetneqq p(M_i) \subset M/N .$$

In diesem Fall erhalten wir Ketten von Unterräumen von N und von M/N, und für jedes i „wächst“ mindestens eine von beiden, so dass umgekehrt $\ell(M) \leq \ell(N) + \ell(M/N)$.

Um die Behauptung zu beweisen, nehmen wir an, dass $p(M_{i-1}) = p(M_i)$ für ein i. Sei $m \in M_i \setminus M_{i-1}$, dann existiert $m' \in M_{i-1}$ mit $p(m) = p(m') \in M/N$. Daraus folgt $m - m' \in \ker p = N$ und $m - m' \in M_i \setminus M_{i-1}$, denn $m - m' \in M_{i-1}$ würde $m = (m - m') + m' \in M_{i-1}$ bedeuten. Es gilt also

$$m - m' \in (M_i \setminus M_{i-1}) \cap N = (M_i \cap N) \setminus (M_{i-1} \cap N) ,$$

so dass $M_{i-1} \cap N \subsetneqq M_i \cap N$. Das beweist (5). □

Beweis von Satz 6.10 (Schluss). Es sei also $A \in M_{m,n}(R)$, und es seien $S \in GL(m, R)$ und $T \in GL(n, R)$ gegeben, so dass SAT die Gestalt (1) hat. Dabei seien $a_1, \dots, a_{\operatorname{rg} A} \in R \setminus \{0\}$ die Diagonaleinträge mit $a_1 \mid \cdots \mid a_k$. Zu zeigen ist noch die Eindeutigkeit des Ranges $\operatorname{rg} A$ und der a_i, jeweils bis auf Multiplikation mit einer Einheit.

Der Kern der Matrix SAT ist ein freier R-Modul mit Basis $(e_{\operatorname{rg} A+1}, \dots, e_n)$. Da S und T invertierbar sind, ist $T\colon \ker(SAT) \to \ker(A)$ ein Isomorphismus, denn

$$(SAT) \cdot v = 0 \qquad \Longrightarrow \qquad A \cdot (Tv) = 0 .$$

Also ist $\ker A$ ebenfalls ein freier R-Modul. Da der Rang $\operatorname{rg} \ker A$ nach Bemerkung 4.23 invariant unter Isomorphismen ist, folgt

$$\operatorname{rg} A = n - \operatorname{rg} \ker(SAT) = n - \operatorname{rg} \ker A ,$$

also ist $\operatorname{rg} A$ eindeutig.

Unter den a_i können Einheiten von R sein. Da $a_i|a_{i+1}$ für alle i, gibt es ein $i_0 \in \mathbb{N}$, so dass $a_i \in R^\times$ genau dann, wenn $i < i_0$. Wir betrachten jetzt den R-Modul $M = R^m/\operatorname{im}(SAT)$. Wie im Beweis von Satz 6.13 gilt

$$M \cong R^{m-\operatorname{rg} A} \oplus \bigoplus_{i=i_0+1}^{\operatorname{rg} A} R/(a_i) ,$$

denn für $i \leq i_0$ gilt $R/(a_i) = R/(1) = \{0\}$. Da S und T invertierbar sind, induziert S einen Isomorphismus $M/\operatorname{im} A \cong M/\operatorname{im}(SAT)$. Aufgrund der soeben bewiesenen Eindeutigkeitsaussage in Satz 6.13 sind $a_{i_0+1}, \dots, a_{\operatorname{rg} A}$ eindeutig bis auf Multiplikation mit einer Einheit durch A bestimmt, und $i_0 \in \mathbb{N}$ ist ebenfalls eindeutig. Da $a_1, \dots, a_{i_0}$ Einheiten sind, sind sie ebenfalls eindeutig bis auf Multiplikation mit einer Einheit durch A bestimmt. Damit ist Satz 6.10 jetzt vollständig bewiesen. $\square$

Beweis von Satz 6.13 (Schluss). Fall R ein Hauptidealring ist, folgt die Existenzaussage im Satz jetzt aus dem vollständigen Existenzbeweis für die Smith-Normalform aus Satz 6.10. Zu zeigen ist noch, dass in

$$\operatorname{Tor} M \cong \bigoplus_{i=1}^{k} R/(a_i)$$

die Elemente $a_1, \dots, a_k$ mit $a_i \mid a_{i+1}$ eindeutig sind bis auf Multiplikation mit Einheiten.

Als Vorüberlegung betrachte die Abbildung $b\,\mathrm{id}_{R/(a)} \colon R/(a) \to R/(a)$. Für $r, t \in R$ gilt $[r] = [t] \,.\, b$ genau dann, wenn ein $s \in R$ mit $r = as + bt$ existiert. Also gilt $[r] \in \operatorname{im}\bigl(b\,\mathrm{id}_{R/(a)}\bigr)$ genau dann, wenn $r = as + tb$ für $s, t \in R$, so dass

$$\operatorname{im}\bigl(b\ \mathrm{id}_{R/(a)}\bigr) = (a, b)/(a) \subset R/(a) .$$

Nach Bemerkung 6.9 ist $(a, b) = (c)$, wobei c ein größter gemeinsamer Teiler von a und b ist. Insbesondere gilt $\ell\bigl(\operatorname{im}(b\,\mathrm{id}_{R/(a)})\bigr) \geq 0$, mit

$$\ell\bigl(\operatorname{im}(b\,\mathrm{id}_{R/(a)})\bigr) = \ell\bigl((a,b)/(a)\bigr) = 0 \quad\Longleftrightarrow\quad (a, b) = (a) \quad\Longleftrightarrow\quad a \mid b .$$

Es gelte jetzt

$$\operatorname{Tor} M \cong \bigoplus_{i=1}^{k} R/(a_i) \cong \bigoplus_{j=1}^{\ell} R/(b_\ell)$$

mit $a_i, b_j \in R \setminus (\{0\} \cup R^\times)$ und $a_1 \mid \cdots \mid a_k$ und $b_1 \mid \cdots \mid b_\ell$. Wir zeigen durch Induktion über j, dass $(a_{k-j}) = (b_{\ell-j})$.

Es sei also $j \geq 0$, und es gelte $(a_{k-i}) = (b_{\ell-i})$ für alle $i < j$. Im Falle $j = 0$ ist diese Voraussetzung trivialerweise erfüllt. Der Untermodul

$$\operatorname{im}\bigl(a_{k-j}\,\mathrm{id}_{\operatorname{Tor} M}\bigr) \subset \operatorname{Tor} M$$

hängt nicht von der gewählten Darstellung ab. Da die direkten Summanden unter jedem Vielfachen der Identität invariant sind, können wir das obige Bild

Summand für Summand bestimmen. Mit Proposition 6.23 (5) erhalten wir also

$$\begin{aligned}\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{\operatorname{Tor}M})\bigr)&=\ell\biggl(\bigoplus_{i=1}^k\operatorname{im}\bigl(a_{k-j}\operatorname{id}_{R/(a_i)}\bigr)\biggr)\\&=\sum_{i=1}^k\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(a_i)})\bigr)=\sum_{i=k-j+1}^k\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(a_i)})\bigr)\;,\end{aligned}$$

da $a_i\mid a_{k-j}$ für alle $i\le k-j$ nach Voraussetzung. Indem wir analog vorgehen und die Induktionsvoraussetzung ausnutzen, erhalten wir

$$\begin{aligned}\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{\operatorname{Tor}M})\bigr)&=\sum_{i=1}^\ell\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(b_i)})\bigr)\\&=\sum_{i=1}^{\ell-j}\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(b_i)})\bigr)+\sum_{i=k-j+1}^k\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(a_i)})\bigr)\;.\end{aligned}$$

Aus dem Vergleich der beiden Formeln erhalten wir

$$0=\sum_{i=1}^{\ell-j}\ell\bigl(\operatorname{im}(a_{k-j}\operatorname{id}_{R/(b_i)})\bigr)\;,$$

und aufgrund unser Vorüberlegung folgt $b_i\mid a_{k-j}$ für alle $i\le\ell-j$, insbesondere gilt also $b_{\ell-j}\mid a_{k-j}$. Indem wir die Rollen der beiden Darstellungen vertauschen, erhalten wir auch $a_{k-j}\mid b_{\ell-j}$, und wegen Bemerkung 6.5 gilt $(a_{k-j})=(b_{\ell-j})$.

Sei ohne Einschränkung $k\le\ell$. Sobald die obige Behauptung für $j=k$ gezeigt ist, folgt

$$0=\ell(\operatorname{Tor}M)-\ell(\operatorname{Tor}M)=\sum_{i=1}^\ell\ell\bigl(R/(b_i)\bigr)-\sum_{i=1}^k\ell\bigl(R/(a_i)\bigr)=\sum_{i=1}^{\ell-k}\ell\bigl(R/(b_i)\bigr)\;,$$

so dass $R/(b_i)=\{0\}$ für alle $i\le\ell-k$. Da aber $b_i\notin R^\times$ nach Voraussetzung, folgt $k=\ell$. Damit ist Satz 6.13 jetzt vollständig bewiesen. $\square$

Beweis von Satz 6.15 (Schluss). Es sei V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F\in\operatorname{End}_\Bbbk V$. Wir nehmen an, dass eine Basis von V existiert, bezüglich der F durch eine Matrix in Frobenius-Normalform (1) dargestellt wird. Insbesondere zerfällt V in F-invariante direkte Summanden $V_1,\dots,V_K$, auf V_i wirkt F durch die Begleitmatrix $M(P_i)$, und es gilt $P_1\mid\dots\mid P_k$. Zu zeigen ist nur noch die Eindeutigkeit der normierten Polynome $P_1,\dots,P_k$.

Wir betrachten den Untervektorraum V_i als $\Bbbk[X]$-Modul wie in Beispiel 6.12 (2). Da $F|_{V_i}$ ein zyklischer Endomorphismus ist, ist V als $\Bbbk[X]$-Modul isomorph zu $\Bbbk[X]/(\chi_{F|_{V_i}})\cong\Bbbk[X]/(P_i)$, und da $\mu_{F|_{V_i}}=\chi_{F|_{V_i}}=P_i$. Insgesamt ist also V als $\Bbbk[X]$-Modul isomorph zu

$$\bigoplus_{i=1}^k\Bbbk[X]\;/\;(P_i)\;.$$

Aus der oben bewiesenen Eindeutigkeitsaussage im Satz 6.13 über invariante Faktoren folgt jetzt auch die Eindeutigkeit der Polynome $P_1, \dots, P_k$. □

Mit der Frobenius-Normalform können wir unsere Folgerung 5.31 aus dem Satz 5.29 von Cayley-Hamilton noch etwas verbessern.

6.24. Folgerung (aus Satz 6.15)**.** *Es sei V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$. Dann gilt $\mu_F \mid \chi_F$, und jedes prime Polynom $P \in \Bbbk[X]$, das χ_F teilt, teilt auch μ_F.*

Beweis. Es seien $P_1, \dots, P_k$ die eindeutigen normierten Polynome zu F aus Satz 6.15. Die Aussage $\mu_F \mid \chi_F$ war der Inhalt von Folgerung 5.31. Wir erhalten einen alternativen Beweis, denn

$$\mu_F = P_k \,\Big|\, \prod_{i=1}^k P_i = \chi_F \,.$$

Sei jetzt $P \in \Bbbk[X]$ prim mit $P \mid \chi_F$. Da P prim ist, teilt P einen der Faktoren P_i von χ_F. Wegen $P_i \mid P_k$ teilt P also auch $P_k = \mu_F$. □

Auch der Beweis von Satz 5.34 über die Diagonalisierbarkeit von Vektorraum-Endomorphismen war noch nicht vollständig.

Beweis von Satz 5.34 (Schluss). Es sei wieder V ein endlichdimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$. Zu zeigen ist die Äquivalenz der Aussage (5) mit den anderen Aussagen im Satz. Dazu seien wieder $P_1, \dots, P_k$ die eindeutigen normierten Polynome aus Satz 6.15.

Zu (5) $\Longrightarrow$ (1) nehmen wir an, dass das Minimalpolynom $P_k = \mu_F$ vollständig in paarweise verschiedene Linearfaktoren zerfällt. Für alle i ist P_i ein Teiler von P_k. Nach Bemerkung 6.21 (2) ist P_i dann ebenfalls ein Produkt paarweise verschiedener Linearfaktoren. Wegen Folgerung 5.6 (2) ist F auf dem zu P_i gehörigen direkten Summanden von V diagonalisierbar. Als direkte Summe diagonalisierbarer Endomorphismen ist dann auch F diagonalisierbar.

Zu „(3) $\Longrightarrow$ (5)“ schließlich überlegen wir uns, dass für einen Eigenraum V_λ zum Eigenwert λ aus $F|_{V_\lambda} = \lambda \operatorname{id}_{V_\lambda}$ folgt, dass

$$\mu_{F|_{V_\lambda}}(X) = \mu_{\lambda \operatorname{id}_{V_\lambda}}(X) = X - \lambda \,.$$

Aus Bemerkung 5.32 (3) folgt, dass μ_F vollständig in paarweise verschiedene Linearfaktoren zerfällt. □

Damit haben wir alle noch offenen Beweise beendet. Wir sehen, dass die Frobenius-Normalform zwar nicht alle Wünsche vom Anfang des Kapitels erfüllt, es aber ermöglicht, einige interessante Resultate zu beweisen.

6.4. Der chinesische Restsatz und der Elementarteilersatz

In diesem Abschnitt sei R stets ein Hauptidealring, und $\mathcal{P}(R) \subset R$ sei wieder eine Menge von Primelementen, die zu jedem Primelement genau ein assoziiertes enthält.

6.25. Definition. Es sei R ein kommutativer Ring mit Eins, dann heißen zwei Elemente a_1 und a_2 *teilerfremd,* wenn es b_1, $b_2 \in R$ mit $a_1b_1 + a_2b_2 = 1$ gibt.

6.26. Bemerkung. Die obige Definition ist äquivalent dazu, dass $(a_1, a_2) = R$. Auch in allgemeinen Ringen ist ein gemeinsamer Teiler r von a_1 und a_2 auch ein Teiler von $a_1b_1 + a_2b_2$. Insbesondere ist r nach Definition 6.4 eine Einheit, wenn es b_1, b_2 wie in der obigen Definition gibt. Wenn R kein Hauptidealring ist, kann es aber sein, dass a_1 und a_2 zwar keinen gemeinsamen Teiler besitzen, der keine Einheit ist, aber trotzdem $(a_1, a_2) \neq R$ gilt. Daher ist die Bezeichnung „teilerfremd“ etwas unglücklich.

Falls R ein Hauptidealring ist, existiert $c \in R$ mit $(a_1, a_2) = (c)$. Nach Bemerkung 6.9 ist c ein größter gemeinsamer Teiler von a_1 und a_2. In diesem Fall haben a_1 und a_2 also genau dann einen nichttrivialen gemeinsamen Teiler, wenn c keine Einheit ist. In diesem Fall entspricht der Begriff „teilerfremd“ also unserer Vorstellung.

Wenn R ein Euklidischer Ring ist, können wir b_1 und b_2 mit dem Euklidischen Algorithmus aus Satz 2.18 berechnen.

Das folgende Resultat geht zurück auf ein chinesisches Manuscript, vermutlich aus dem dritten Jahrhundert. Die Formulierung dort lautet sinngemäß: „Es seien $a_1, \dots, a_k$ paarweise teilerfremde natürliche Zahlen, dann existiert für jedes Tupel ganzer Zahlen $n_1, \dots, n_k$ eine ganze Zahl n, die die folgende simultane Kongruenz erfüllt:

$$n \equiv n_i \mod a_i \qquad \text{für } i = 1, \dots, k\ .$$

Alle Lösungen dieser Kongruenz sind kongruent modulo $a_1 \cdots a_k$.“ Diese Aussage ist für $R = \mathbb{Z}$ äquivalent zur Existenz einer Abbildung G wie im folgenden Beweis mit der Eigenschaft, dass $F \circ G = \mathrm{id}$.

6.27. Satz (Chinesischer Restsatz)**.** *Es sei R ein kommutativer Ring mit Eins, und $a_1, \dots, a_k \in R$ seien paarweise teilerfremd. Dann induziert die Quotientenabbildung $R \to R/(a_i)$ eine R-lineare Abbildung*

$$\pi_i\colon R/(a_1 \cdots a_k) \to R/(a_i)\ , \tag{1}$$

und wir erhalten einen R-Modul-Isomorphismus

$$\begin{gathered} F\colon R/(a_1 \cdots a_k) \longrightarrow \bigoplus_{i=1}^k R/(a_i) \\ \text{mit} \qquad F([a]) = \bigl(\pi_1([a]), \dots, \pi_k([a])\bigr)\ . \end{gathered} \tag{2}$$

Da es sich auf der rechten Seite nicht um eine Summe von Untermoduln handelt, müssen wir eigentlich $\coprod$ anstelle von $\bigoplus$ schreiben. Wegen Proposition 2.58 (2) sind diese beiden Schreibweisen äquivalent, und die obige ist etwas gebräuchlicher.

In Aufgabe 1 von Blatt 7 zur Linearen Algebra I haben wir uns den Fall $R = \mathbb{Z}$ und $k = 2$ bereits angeschaut. Wenn man die direkte Summe von Ringen auf der rechten Seite von (2) wieder als Ring mit summandenweiser Multiplikation auffasst, ist F sogar ein Ringisomorphismus. Der Satz gilt noch etwas allgemeiner, wenn man $(a_1), \dots, (a_k)$ durch beliebige Ideale ersetzt, dazu muss man allerdings erst das Produkt von Idealen definieren.

Beweis. Die Abbildungen π_i sind wohldefiniert, denn $(a_1 \cdots a_k) \subset (a_i)$ für alle i nach Bemerkung 6.5 (1), so dass für alle r, $s \in R$ gilt

$$\begin{aligned}
[r] = [s] \in R/(a_1 \cdots a_k) \qquad \Longrightarrow \qquad & r - s \in (a_1 \cdots a_k) \subset (a_i) \\
\Longrightarrow \qquad & [r] = [s] \in R/(a_i) \; .
\end{aligned}$$

Also gilt (1). Wir zeigen (2) und

$$\big(a_1, a_2 \cdots a_k\big) = R \qquad \text{für alle } k \geq 2 \tag{3}$$

durch Induktion über k. Für $k = 1$ ist nichts zu zeigen.

Für $k = 2$ ist (3) klar nach Voraussetzung. Nach Definition 6.25 existieren zwei Elemente $b_1, b_2 \in R$, so dass $a_1 b_1 + a_2 b_2 = 1$. Wir definieren eine Abbildung

$$G\colon R/(a_1) \oplus R/(a_2) \to R/(a_1 a_2) \qquad \text{mit} \qquad G\big([r_1], [r_2]\big) = [a_2 b_2 r_1 + a_1 b_1 r_2]$$

für alle $[r_1] \in R/(a_1)$, $[r_2] \in R/(a_2)$, wobei r_1, $r_2 \in R$. Diese Abbildung ist wohldefiniert, denn seien s_1, $s_2 \in R$, dann gilt

$$\begin{aligned}
\big[a_2 b_2 (r_1 + a_1 s_1) + a_1 b_1 (r_2 + a_2 s_2)\big] &= \big[a_2 b_2 r_1 + a_1 b_1 r_2 + a_1 a_2 (b_2 s_1 + b_1 s_2)\big] \\
&= [a_2 b_2 r_1 + a_1 b_1 r_2] \; ,
\end{aligned}$$

so dass das Ergebnis modulo $a_1 a_2$ nicht von der Wahl der Repräsentanten r_1 und r_2 abhängt.

Es gilt $F \circ G = \mathrm{id}_{R/(a_1) \oplus R/(a_2)}$, denn seien $[r_1] \in R/(a_1)$, $[r_2] \in R/(a_2)$, dann folgt

$$\begin{aligned}
(F \circ G)\big([r_1], [r_2]\big) &= \big(\pi_1([a_2 b_2 r_1 + a_1 b_1 r_2]), \pi_2([a_2 b_2 r_1 + a_1 b_1 r_2])\big) \\
&= \big(\pi_1([(1 - a_1 b_1) r_1 + a_1 b_1 r_2]), \pi_2([a_2 b_2 r_1 + (1 - a_2 b_2) r_2])\big) \\
&= \big([r_1], [r_2]\big) \; .
\end{aligned}$$

Umgekehrt ist auch $G \circ F = \mathrm{id}_{R/(a_1 a_2)}$, denn für $r \in R$ gilt

$$(G \circ F)[r] = [a_2 b_2 r + a_1 b_1 r] = [r] \; .$$

Somit ist F invertierbar mit Umkehrfunktion G, und (2) ist bewiesen. Zu (3) ist für $k = 2$ nichts zu zeigen.

Sei nun $k \geq 3$, und (2) und (3) seien bewiesen für alle kleineren Werte von k. Dann existieren c, d, e, $f \in R$, so dass

$$1 = a_1 c + a_3 \cdots a_k d = a_1 e + a_2 f \; .$$

Es folgt

$$1 = (a_1 c + a_3 \cdots a_k d) \cdot (a_1 e + a_2 f) = a_1 (e + a_2 c f) + a_2 \cdots a_k (df) \;,$$

insbesondere folgt (3). Wir zeigen (2), indem wir F als Verkettung

$$F \colon R/(a_1 \cdots a_k) \longrightarrow R/(a_1) \oplus R/(a_2 \cdots a_k) \longrightarrow R/(a_1) \oplus \bigoplus_{i=2}^{k} R/(a_i)$$

schreiben. Die erste Abbildung ist ein Isomorphismus wegen unseres Arguments zu (2). Die Abbildung $R/(a_2 \cdots a_k) \to R/(a_2) \oplus \cdots \oplus R/(a_k)$ ist ein Isomorphismus nach Induktionsvoraussetzung. Nach Definition 2.61 ist die direkte Summe als Menge das kartesische Produkt der beiden Summanden. Jetzt sieht man leicht, dass die zweite Abbildung oben bijektiv, und somit ein Isomorphismus ist. □

6.28. Beispiel (Lagrange-Interpolation)**.** Es sei $P \in \Bbbk[X]$ ein Polynom und $x \in \Bbbk$. Wie im Beweis von Folgerung 5.15 dividieren wir P mit Rest durch $X - x$ und erhalten

$$P = S \cdot (X - x) + y \;,$$

es folgt

$$P(x) = S(x) \cdot (x - x) + y = y \;.$$

Mit anderen Worten gilt

$$P \equiv y \mod (X - x) \qquad \Longleftrightarrow \qquad P(x) = y \;.$$

Es seien jetzt $x_0, \dots, x_n \in \Bbbk$ paarweise verschieden, dann sind die linearen Polyome $X - x_0, \dots, X - x_n$ paarweise teilerfremd, denn

$$\frac{1}{x_j - x_i}\,(X - x_i) - \frac{1}{x_j - x_i}\,(X - x_j) = 1 \;.$$

Zu jeder beliebigen Wahl von $y_0, \dots, y_n$ existiert dann nach dem chinesischen Restsatz 6.27 ein Polynom P, so dass

$$P \equiv y_i \mod (X - x_i) \;, \qquad \text{also} \qquad P(x_i) = y_i \qquad \text{für alle } i = 0, \dots, n \;.$$

Da P nur bis auf Vielfache von $Q = (X - x_0) \cdots (X - x_n)$ eindeutig bestimmt ist, können wir P durch seinen Rest modulo Q ersetzen. Mit anderen Worten dürfen wir annehmen, dass $\deg P \leq n = \deg Q - 1$, und wegen der Eindeutigkeit der Division mit Rest nach Satz 5.13 ist P dann sogar eindeutig. Man nennt P auch das *Lagrange-Polynom* durch die Punkte $(x_0, y_0), \dots, (x_n, y_n)$. Ein direktes Verfahren zur Bestimmung des Lagrange-Polynoms haben wir in Aufgabe 1 von Blatt 3 kennengelernt.

6.29. Bemerkung. Es sei jetzt R wieder ein Hauptidealring und $\mathcal{P}(R) \subset R$ wieder eine Menge von Primelementen von R, die zu jedem Primelement $p \in R$ genau ein assoziiertes Element enthält. Es sei

$$r = e \cdot \prod_{p \in \mathcal{P}(R)} p^{\mu_p(r)} = e \cdot p_1^{\ell_1} \cdots p_k^{\ell_k}$$

die Primfaktorzerlegung von r gemäß Satz 6.19, dabei sei $e \in R^\times$, und $p_1, \dots, p_k \in \mathcal{P}(R)$ seien paarweise verschieden. Die Multiplizitäten $\mu_p(r) \in \mathbb{N}$ sind wie in Bemerkung 6.21 (1) definiert. In der zweiten Schreibweise haben wir nur diejenigen Primelemente $p \in \mathcal{P}(R)$ augeführt, für die $\mu_p(r) > 0$. Insbesondere sind die Terme $p_1^{\ell_1}, \dots, p_k^{\ell_k}$ paarweise teilerfremd. Der folgende Isomorphismus ist eine typische Anwendung des chinesischen Restsatzes:

$$R/(r) \cong R/\bigl(p_1^{\ell_1}\bigr) \oplus \dots \oplus R/\bigl(p_k^{\ell_k}\bigr) \cong \bigoplus_{p \in \mathcal{P}(R)} R/\bigl(p^{\mu_p(r)}\bigr)\;.$$

Wir erinnern uns an den Rang $\operatorname{rg} M$ eines endlich erzeugten R-Moduls aus Satz 6.13.

6.30. Satz (Elementarteilersatz)**.** *Es sei R ein Hauptidealring und M ein endlich erzeugter unitärer R-Modul. Dann existieren für jedes Primelement $p \in \mathcal{P}(R)$ Zahlen $k_p \in \mathbb{N}$ und $1 \le \ell_{p,1} \le \dots \le \ell_{p,k_p}$, sowie ein Isomorphismus*

$$F\colon R^{\operatorname{rg} M} \oplus \bigoplus_{p \in \mathcal{P}(R)} \bigoplus_{i=1}^{k_p} R/\bigl(p^{\ell_{p,i}}\bigr) \longrightarrow M\;.$$

Dabei die Zahlen k_p und $\ell_{p,1}, \dots, \ell_{p,k_p}$ eindeutig durch M bestimmt, und für fast alle $p \in \mathcal{P}(R)$ gilt $k_p = 0$. Für jedes $p \in \mathcal{P}(R)$ ist der Untermodul

$$\operatorname{Tor}_p M = F\Biggl(\bigoplus_{i=1}^{k_p} R/\bigl(p^{\ell_{p,i}}\bigr)\Biggr) \subset M$$

ebenfalls eindeutig bestimmt, nicht jedoch der Isomorphismus F. Außerdem gilt

$$\operatorname{Tor} M = \bigoplus_{p \in \mathcal{P}(R)} \operatorname{Tor}_p M\;.$$

Ähnlich wie bei Satz 6.13 sehen wir im Laufe des Beweises, dass

$$\operatorname{Tor}_p M = \bigl\{\, m \in M \bigm| \text{es gibt } k \in \mathbb{N} \text{ mit } m \,.\, p^k = 0 \,\bigr\}\;.$$

Wir nennen $\operatorname{Tor}_p M$ den *p-Torsionsuntermodul* von M. Wenn $M = \operatorname{Tor}_p M$ heißt M ein *p-Torsionsmodul.* Die Zahlen $p^{\ell_{p,i}}$ heißen die *Elementarteiler* von M.

Beweis. Zur Existenz bestimmen wir mit Satz 6.13 zunächst die invarianten Faktoren $a_1, \dots, a_k \in R$ mit $a_1 \mid \dots \mid a_k$ und einen Isomorphismus

$$G\colon M \cong R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^{k} R/(a_i)\;.$$

Für jedes i liefert der chinesische Restsatz 6.27 wie in Bemerkung 6.29 einen Isomorphismus

$$F_i\colon R/(a_i) \cong \bigoplus_{p \in \mathcal{P}(R)} R/\bigl(p^{\mu_p(a_i)}\bigr)\;.$$

Für $p \in \mathcal{P}(R)$ sei k_p die Anzahl der Indizes i mit $\mu_p(a_i) \neq 0$, also die Anzahl der i mit $p \mid a_i$. Aus Satz 6.19 folgt, dass $k_p = 0$ für fast alle p, da insgesamt nur endlich viele Primfaktoren in den Zerlegungen von $a_1, \dots, a_k$ vorkommen.

Aus $a_1 \mid \cdots \mid a_k$ folgt mit Bemerkung 6.21 (2), dass $\mu_p(a_1) \le \cdots \le \mu_p(a_k)$. Also seien

$$1 \le \ell_{p,1} = \mu_p\bigl(a_{1+k-k_p}\bigr) \le \cdots \le \ell_{p,k_p} = \mu_p\bigl(a_k\bigr)$$

genau die Multiplizitäten $\mu_p(a_j)$, die nicht verschwinden. Wir erhalten den Isomorphismus F als Verkettung

$$\begin{aligned}(1)\quad M \xrightarrow{G} R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^k R/(a_i) \xrightarrow{\operatorname{id}\oplus F_1 \oplus \cdots \oplus F_k} R^{\operatorname{rg} M} \oplus \bigoplus_{i=1}^k \bigoplus_{p\in\mathcal{P}(R)} R\bigl(p^{\mu_p(a_i)}\bigr)&\\ \longrightarrow R^{\operatorname{rg} M} \oplus \bigoplus_{p\in\mathcal{P}(R)} \bigoplus_{i=1}^{k_p} R/\bigl(p^{\ell_{p,i}}\bigr)&\;.\end{aligned}$$

Der letzte Isomorphismus sortiert nur die Summanden um, und lässt alle Summanden der Form $R/(e) = \{0\}$ weg. Damit ist die Existenzaussage bewiesen.

Zur Eindeutigkeit beginnen wir mit einer Vorüberlegung. Für alle $p, q \in \mathcal{P}(R)$ mit $p \neq q$ und alle $\ell \ge 1$ ist $[q] \in R/(p^\ell)$ eine Einheit, denn p^ℓ und q sind teilerfremd, also existieren $b, c \in R$ mit $ap^\ell + bq = 1$, und es folgt $[b][q] = [1] \in R/(p^\ell)$. Insbesondere gilt $[r] \,.\, q^k \neq 0$ für alle $[r] \in R/(p^\ell) \setminus \{0\}$ und alle $k \in \mathbb{N}$.

Wir betrachten für jedes $q \in \mathcal{P}(R)$ die Teilmenge

$$N_q = \bigl\{\, m \in M \bigm| \text{es gibt } k \in \mathbb{N} \text{ mit } m \,.\, q^k = 0 \,\bigr\} \;.$$

Wie N im Beweis von Satz 6.13 ist auch N_q eindeutig durch M und q bestimmt. Es sei $n = \sum_{p\in\mathcal{P}(R)} k_p \in \mathbb{N}$ die Anzahl der Summanden vom Typ $R/(p^{\ell_{p,i}})$, dann schreiben wir

$$\bigoplus_{p\in\mathcal{P}(R)} \bigoplus_{i=1}^{k_p} R/\bigl(p^{\ell_{p,i}}\bigr) = \bigoplus_{i=1}^{n} R/\bigl(p_i^{\ell_i}\bigr) \;.$$

Seien jetzt $v_1, \dots, v_{\operatorname{rg} M} \in R$ und $[s_i] \in R/(p_i^{\ell_i})$ für $i = 1, \dots, n$. Dann gilt

$$\begin{aligned}F\bigl((v_1,\dots,v_{\operatorname{rg} M}),[s_1],\dots,[s_n]\bigr) \,.\, q^k&\\ = F\bigl((v_1\,q^k,\dots,v_{rgM}\,q^k),[s_1\,q^k],\dots,[s_n\,q^k]\bigr)&\;.\end{aligned}$$

Da $\Bbbk$ ein Integritätsbereich ist, gilt $v_i\,q^k = 0$ genau dann, wenn $v_i = 0$. Falls $p_i \neq q$, gilt nach Vorüberlegung $[s_i\,q^k] = [0] \in R/(p_i^{\ell_i})$ genau dann, wenn $[s_i] = [0]$. Wenn allerdings $p_i = q$ gilt, dann ist $[s_i\,q^k] = [0]$ für alle $k \ge \ell_i$. Insgesamt folgt daraus, dass

$$\begin{aligned}&F\bigl((v_1,\dots,v_{\operatorname{rg} M}),[s_1],\dots,[s_n]\bigr) \,.\, q^k = 0\\ &\qquad\Longleftrightarrow\qquad \bigl((v_1,\dots,v_{\operatorname{rg} M}),[s_1],\dots,[s_n]\bigr) \in \bigoplus_{i=1}^{k_q} R/\bigl(q^{\ell_{q,i}}\bigr)\end{aligned}$$

für alle hinreichend großen k gilt. Mithin gilt

$$\operatorname{Tor}_q M = N_q \;,$$

insbesondere ist $\operatorname{Tor}_p M \subset M$ eindeutig bestimmt.

Die Eindeutigkeit der Zahlen k_p und $\ell_{p,1} \le \cdots \le \ell_{p,k_p}$ folgt, indem wir die Eindeutigkeitsaussage aus Satz 6.13 auf den p-Torsionsuntermodul $\operatorname{Tor}_p M$ anwenden. Schließlich folgt aus der Konstruktion in (1), dass

$$\operatorname{Tor} M = \bigoplus_{p\in\mathcal{P}(R)} \operatorname{Tor}_p M \;.$$

Insbesondere ist auch der Rang $\operatorname{rg} M$ eindeutig wegen Satz 6.13, denn

$$\operatorname{rg} M = \operatorname{rg}\bigl(M/\operatorname{Tor} M\bigr) = \operatorname{rg}\biggl(M \Bigm/ \bigoplus_{p\in\mathcal{P}(R)} \operatorname{Tor}_p M\biggr) \;. \qquad \square$$

Nach Beispiel 6.12 (1) sind abelsche Gruppen gerade $\mathbb{Z}$-Moduln, und G ist als abelsche Gruppe genau dann endlich erzeugt, wenn G als $\mathbb{Z}$-Modul endlich erzeugt ist. Wir können also die Sätze 6.13 und 6.30 auf endlich erzeugte abelsche Gruppen anwenden. Dabei schreiben wir G additiv, das heißt, das neutrale Element ist 0 und

$$g \,.\, n = \underbrace{g + \cdots + g}_{n \text{ Summanden}} \;.$$

6.31. Satz (Hauptsatz über endlich erzeugte abelsche Gruppen)**.** *Es sei G eine endlich erzeugte abelsche Gruppe.*

(1) Invariante Faktoren. *Dann existieren eindeutig bestimmte Zahlen* $\operatorname{rg} G$, $k \in \mathbb{N}$ *und* $a_1, \dots, a_k \in \mathbb{N}$ *mit* $a_1 \ge 2$ *und* $a_i \mid a_{i+1}$ *für alle* $i = 1, \dots, k-1$, *und ein Isomorphismus*

$$F\colon \mathbb{Z}^{\operatorname{rg} G} \oplus \bigoplus_{i=1}^{k} \mathbb{Z}/a_i\mathbb{Z} \longrightarrow G \;.$$

Die Torsionsuntergruppe

$$\operatorname{Tor} G = F\biggl(\bigoplus_{i=1}^{k} \mathbb{Z}/a_i\mathbb{Z}\biggr) = \bigl\{\, g \in G \bigm| \textit{es gibt } n > 0 \textit{ mit } g \,.\, n = 0 \,\bigr\}$$

ist ebenfalls eindeutig bestimmt.

(2) Elementarteiler. *Es existiert eine eindeutig bestimmte Zahl* $\operatorname{rg} G$ *und für jede positive Primzahl* $p \in \mathcal{P}(\mathbb{Z})$ *existieren eindeutige Zahlen* $k_p \in \mathbb{N}$ *und* $1 \le \ell_{p,1} \le \cdots \le \ell_{p,k_p}$, *und ein Isomorphismus*

$$F\colon \mathbb{Z}^{\operatorname{rg} G} \oplus \bigoplus_{p\in\mathcal{P}(\mathbb{Z})} \bigoplus_{i=1}^{k_p} \mathbb{Z}/p^{\ell_{p,i}} \longrightarrow G \;.$$

Für jedes $p \in \mathcal{P}(\mathbb{Z})$ *ist die* p-Torsionsuntergruppe

$$\operatorname{Tor}_p M = F\biggl(\bigoplus_{i=1}^{k_p} \mathbb{Z}/p^{\ell_{p,i}}\biggr) = \bigl\{\, g \in G \bigm| \textit{es gibt } k \ge 0 \textit{ mit } g \,.\, p^k = 0 \,\bigr\}$$

eindeutig bestimmt, und es gilt $k_p = 0$ *und* $\operatorname{Tor}_p M = \{0\}$ *für fast alle* $p \in \mathcal{P}(\mathbb{Z})$.

Insbesondere ist der Rang $\operatorname{rg} G$ *von* G *eindeutig bestimmt, außerdem gilt*

$$\operatorname{Tor} G = \bigoplus_{p \in \mathcal{P}(R)} \operatorname{Tor}_p G\,.$$

Beweis. Der Satz folgt unmittelbar aus den Sätzen 6.13 und 6.30. □

6.32. Beispiel. Einige Beispiele endlich erzeugter abelscher Gruppen lernen Sie in den Übungen kennen. Wenn G nicht endlich erzeugt ist, könnte G beispielsweise eine unendliche direkte Summe von Gruppen der Form $\mathbb{Z}$ oder $\mathbb{Z}/p^\ell\mathbb{Z}$ sein. Es gibt aber auch andere Möglichkeiten, beispielsweise ist die additive Gruppe der rationalen Zahlen $\mathbb{Q}$ nicht endlich erzeugt. Jede endlich erzeugte Untergruppe von $\mathbb{Q}$ ist entweder $\{0\}$ oder aber isomorph zu $\mathbb{Z}$, insbesondere lässt sich $\mathbb{Q}$ nicht als direkte Summe zweier echter Untergruppen schreiben.

Ähnliche Beispiele nicht endlich erzeugter R-Moduln lassen sich über jedem Hauptidealring R konstruieren, wenn R kein Körper ist, beispielsweise auch über $\Bbbk[X]$.

6.33. Bemerkung. Mit dem Satz 6.13 und dem Elementarteilersatz 6.30 haben wir zwei Normalformen für endlich erzeugte Moduln M über Hauptidealringen kennengelernt. Das widerspricht nicht unser Forderung in Bemerkung 3.17, wonach Normalformen immer eindeutig sein sollen, denn sobald wir uns auf eine Normalform geeinigt haben, ist diese für jeden Modul M eindeutig (eventuell bis auf die Reihenfolge gewisser direkter Summanden).

Die zugehörigen vollständigen Invarianten sind im ersten Fall der Rang $\operatorname{rg} M$ sowie k und die invarianten Faktoren $a_1 \mid \cdots \mid a_k \in R \setminus (R^\times \cup \{0\})$, im zweiten Fall wieder der Rang $\operatorname{rg} M$ und für jedes Primelement $p \in \mathcal{P}(R)$ die Zahlen k_p und $1 \le \ell_{p,1} \le \cdots \le \ell_{p,k_p}$. Strenggenommen hätten wir noch zeigen müssen, dass es zu jeder Wahl der obigen Zahlen und Elemente jeweils auch einen entsprechenden endlich erzeugten R-Modul gibt, aber das ist klar, da wir diesen Modul ja einfach hinschreiben können.

Man kann zeigen, dass der Elementarteilersatz den Modul M in die größtmögliche Zahl von Untermoduln der Form R oder $R/(r)$ zerlegt, während der Satz über invariante Faktoren M in die kleinstmögliche Anzahl solcher Untermoduln zerlegt. Das bedeutet auf der anderen Seite, dass die Darstellung durch invariante Faktoren in der Regel weniger Ringelemente a_i braucht, um $\operatorname{Tor} M$ vollständig zu beschreiben, als die Elementarteilerdarstellung Exponenten.

Da die Zerlegung im Elementarteilersatz feiner als im Satz über invariante Faktoren ist, sagt sie mehr über M aus. Auf der anderen Seite benötigen wir zur Bestimmung der Exponenten $\ell_{p,i}$ im Elementarteilersatz aber die Primfaktorzerlegungen der invarianten Faktoren $a_1, \dots, a_k$, und wir haben in Beispiel 6.20 gesehen, dass es mitunter nicht möglich ist, die Primfaktoren durch einen Algorithmus zu bestimmen. Die invarianten Faktoren sind dagegen leichter zu berechnen.

6.5. Allgemeine und Jordan-Normalform

Wie in Satz 6.15 können wir den Elementarteilersatz 6.30 benutzen, um Normalformen für Endomorphismen F endlich-dimensionaler Vektorräume V hinzuschreiben. Die Frage ist dabei, durch welche Matrix wir F auf einem invarianten Unterraum $U \subset V$ darstellen, der als $\Bbbk[X]$-Untermodul isomorph ist zu $\Bbbk[X]/(P^k)$, für ein primes normiertes Polynom $P \in \mathcal{P}(\Bbbk[X])$ und $k \geq 1$.

6.34. Beispiel. Es sei $P = X^2 - X + 1 \in \mathbb{Q}[X]$. Für alle $q \in \mathbb{Q}$ gilt

$$q^1 - q + 1 = \left(q - \frac{1}{2}\right)^2 + \frac{3}{4} > 0\ ,$$

also hat P keine Nullstellen und spaltet daher nach Folgerung 5.15 keine Linearfaktoren ab. Da jeder echte Teiler von P ein Polynom von kleinerem Grad, also ein Linearfaktor, sein muss, ist P somit irreduzibel.

Wir betrachten den zyklischen $\mathbb{Q}[X]$-Torsionsmodul $V = \mathbb{Q}[X]/(P^2)$ als $\mathbb{Q}$-Vektorraum mit einem Endomorphismus $F = X$ und erinnern uns an unsere Überlegungen zur zyklischen Normalform aus den Übungen, die wir vor dem Satz 6.15 von der Frobenius-Normalform zusammengefasst hatten. Es sei v ein zyklischer Erzeuger. Da $P^2 = X^4 - 2X^3 + 3X^2 - 2X + 1$, wird F bezüglich der $\mathbb{Q}$-Basis $(v, F(v), F^2(v), F^3(v))$ dargestellt durch die Begleitmatrix

$$M(P^2) = \begin{pmatrix} 0 & & & -1 \\ 1 & 0 & & 2 \\ & 1 & 0 & -3 \\ & & 1 & 2 \end{pmatrix} . \tag{1}$$

In Übung 4 von Blatt 5 haben wir uns die Basis $(1, X, P(X), X \cdot P(X))$ angeschaut. Indem wir F einsetzen und die resultierenden Abbildungen auf v anwenden, erhalten wir eine neue Basis

$$(b_{11}, b_{12}, b_{21}, b_{22}) = \big(v, F(v), F^2(v) - F(v) + v, F^3(v) - F^2(v) + F(v)\big)$$

von V. Bezüglich dieser Basis wird F durch die Blockmatrix

$$\begin{pmatrix} M(P) & 0 \\ Z & M(P) \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad \text{mit} \qquad Z = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{2}$$

dargestellt. Anhand der Blockgestalt kann man erkennen, dass der Unterraum $U = \langle b_{21}, b_{22} \rangle$ invariant unter F ist. Dazu benutzen wir das Schema zu

Definition 2.64:

$$\begin{pmatrix} 0\\0\\ *\\ * \end{pmatrix}$$

$$\begin{pmatrix} * & * & 0 & 0\\ * & * & 0 & 0\\ * & * & * & *\\ * & * & * & * \end{pmatrix} \begin{pmatrix} 0\\0\\ *\\ * \end{pmatrix}$$

Man kann überprüfen, dass in diesem Beispiel $U = \operatorname{im} P(F) = \ker P(F)$ gilt. Da man den invarianten Unterraum anhand der Matrix $M(P^2)$ aus (1) nicht unmittelbar erkennen kann, beschreibt die Darstellung (2) den Endomorphismus F etwas besser.

6.35. Definition. Es sei $P \in \Bbbk[X]$ ein normiertes Polynom vom Grad $n \geq 1$, und es sei $\ell \geq 1$. Dann definieren wir den *verallgemeinerten Jordan-Block* $M_\ell(P) \in M_{\ell n}(\Bbbk)$ der Größe ℓ zum Polynom P durch

$$M_\ell(P) = \begin{pmatrix} M(P) & & & 0 \\ Z_n & M(P) & & \\ & \ddots & \ddots & \\ 0 & & Z_n & M(P) \end{pmatrix} \quad \text{mit} \quad Z_n = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & & 0 \\ & & & \vdots \\ 0 & \cdots & & 0 \end{pmatrix}.$$

Für $\lambda \in \Bbbk$ heißt $J(\lambda, \ell) = M_\ell(X - \lambda)$ der *Jordan-Block* der Größe ℓ zum Eigenwert λ.

Die Begleitmatrix zum Polynom $X - \lambda$ ist die Matrix $M(X - \lambda) = (\lambda) \in M_1(\Bbbk)$. Ein typischer „echter" Jordan-Block hat also die konkrete Gestalt

$$M_\ell(\lambda) = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 1 & \lambda & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & & 1 & \lambda \end{pmatrix}$$

In der Literatur findet man oft auch die „an der Diagonalen gespiegelte" Matrix, bei der also die Einsen oberhalb der Diagonalen stehen. Wenn man die Reihenfolge der Basisvektoren des zugrundeliegenden ℓ-dimensionalen Vektorraums umdreht, geht die eine in die andere Form über.

6.36. Satz (Allgemeine Normalform)**.** *Es sei V ein endlich-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$. Dann existieren $k \in \mathbb{N}$, irreduzible normierte Polynome $P_1, \dots, P_k \in \mathcal{P}(\Bbbk[X])$ und $\ell_1, \dots, \ell_k \in \mathbb{N} \setminus \{0\}$ und eine Basis B von V, so dass F bezüglich B durch die Block-Diagonalmatrix*

$$\begin{pmatrix} M_{\ell_1}(P_1) & 0 & \cdots & 0 \\ 0 & M_{\ell_2}(P_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M_{\ell_k}(P_k) \end{pmatrix}$$

dargestellt wird. Dabei sind die Paare (P_1, ℓ_1), …, (P_k, ℓ_k) *bis auf die Reihenfolge eindeutig bestimmt, die Basis* B *jedoch nicht. Es gilt*

$$\dim V = \sum_{i=1}^{k} \ell_i \deg P_i \;, \quad \chi_F = \prod_{i=1}^{k} P_i^{\ell_i} \quad \text{und} \quad \mu_F = \mathrm{kgV}\bigl(P_1^{\ell_1}, \dots, P_k^{\ell_k}\bigr) \;.$$

Sei $A \in M_n(\Bbbk)$ *eine Matrix, dann existiert eine invertierbare Matrix* $B \in GL(n, \Bbbk)$*, so dass* $B\,A\,B^{-1}$ *die oben angegebene Gestalt besitzt. Dabei ist die resultierende Matrix wieder bis auf die Reihenfolge der Blöcke eindeutig, die Matrix* B *jedoch nicht.*

Andere Namen für diese Normalform sind „rationale Normalform“ oder „verallgemeinerte Jordan-Normalform“. Eine Matrix vom obigen Typ heißt manchmal auch verallgemeinerte Jordan-Matrix. Diese Matrizen sind zunächst einmal nur eindeutig bis auf die Reihenfolge der Blöcke. Man kann sie eindeutig machen, indem man willkürlich eine Anordnung auf der Menge $\mathcal{P}(\Bbbk[X])$ der normierten irreduziblen Polynome festlegt, und, sollte ein Polynom mehrfach vorkommen, die zugehörigen Blöcke der Größe nach sortiert.

Beweis. Wir beginnen mit der Fassung für $F \in \mathrm{End}_\Bbbk V$. Wie im Beweis von Satz 6.15 zur Frobenius-Normalform fassen wir V als $\Bbbk[X]$-Modul auf, wobei X durch F wirkt. Wir wissen bereits, dass V ein $\Bbbk[X]$-Torsionsmodul ist, da $\dim_\Bbbk V < \infty$. Anschließend zerlegen wir V mit dem Elementarteilersatz in eine direkte Summe von Moduln der Form $V_i = \Bbbk[X]/\bigl(P_i^{\ell_i}\bigr)$, wobei die $P_i \in \mathcal{P}(\Bbbk[X])$ normierte irreduzible Polynome seien.

Nach Übung 4 von Blatt 5 besitzt $\Bbbk[X]/\bigl(P_i^{\ell_i}\bigr)$ eine $\Bbbk$-Basis B_i, bezüglich der F durch den verallgemeinerten Jordan-Block $M_{\ell_i}(P_i)$ dargestellt wird. Wir setzen die Basen B_i zu einer $\Bbbk$-Basis B von V zusammen. Dann wird F bezüglich B durch die angegebene Matrix dargestellt. Damit ist die Existenzaussage bewiesen.

Zur Eindeutigkeit überlegen wir uns, dass ein invarianter direkter Summand V_i von V mit einer Basis B_i, bezüglich der $F|_{V_i}$ durch einen Jordanblock $M_{\ell_i}(P_i)$ mit $P_i \in \mathcal{P}(\Bbbk[X])$ dargestellt wird, als $\Bbbk[X]$-Modul zu $\Bbbk[X]/\bigl(P_i^{\ell_i}\bigr)$ isomorph ist. Jetzt folgt die Eindeutigkeitsaussage aus der Eindeutigkeitsaussage im Elementarteilersatz.

Wir haben im Beweis von Satz 6.15 gesehen, dass $\dim\bigl(\Bbbk[X]/(P)\bigr) = \deg P$ für jedes Polynom P gilt. Also hat der Summand V_i die Dimension

$$\dim V_i = \deg P_i^{\ell_i} = \ell_i \deg P_i \;,$$

und es folgt die Dimensionsformel. Die Formeln für das charakteristische Polynom und das Minimalpolynom ergeben sich aus Folgerung 4.17 (1) und aus Bemerkung 5.32 (3).

Sei jetzt $A \in M_n(\Bbbk)$ gegeben, dann fassen wir A als Endomorphismus von $\Bbbk^n$ auf. Dabei fassen wir die Basis B als Basisabbildung $B\colon \Bbbk^n \to \Bbbk^n$ auf, siehe Bemerkung 2.74, und erhalten so die gesuchte Matrix $B \in GL(n, \Bbbk)$. $\square$

6.37. Bemerkung. Wir vergleichen die Frobenius-Normalform mit der allgemeinen Normalform.

Die allgemeine Normalform ist in zweierlei Hinsicht „feiner“ als die Frobenius-Normalform: Zum einen benutzt sie den Elementarteilersatz, der V nach Bemerkung 6.33 in eine direkte Summe von möglichst vielen zyklischen $\Bbbk[X]$-Untermoduln zerlegt, während die Frobenius-Normalform den Satz über invariante Faktoren benutzt, der eine entsprechende Summe mit möglichst wenigen Summanden liefert. Insbesondere sind die einzelnen Blöcke in der allgemeinen Normalform eher kleiner, und damit „einfacher“, als in der Frobenius-Normalform. Damit kommen wir unser Forderung (2) vom Anfang des Kapitels also wesentlich näher.

Zum anderen benutzt die allgemeine Normalform für die einzelnen Summanden $V_i \cong \Bbbk[X]/(P^\ell)$ den verallgemeinerten Jordanblock $M_\ell(P)$, der eine untere Block-Dreiecksgestalt hat. Wie in Übung 5 von Blatt 4 und Beispiel 6.34 ersichtlich, kann man anhand dieser Gestalt die Existenz einer Folge F-invarianter Unterräume ablesen, die gegeben werden durch

$$\begin{aligned}\{0\} = \ker\bigl(P(F|_{V_i})^0\bigr) = \operatorname{im}\bigl(P(F|_{V_i})^\ell\bigr) \subsetneqq \dots \subsetneqq&\\ \subsetneqq \ker\bigl(P(F|_{V_i})^\ell\bigr) = \operatorname{im}\bigl(P(F|_{V_i})^0\bigr) = V_i\;.&\end{aligned}$$

Wir lernen also weitaus mehr über die Struktur von V und F, und erfüllen damit auch Forderung (3) weitaus besser.

Auf der anderen Seite lässt sich die Frobenius-Normalform mit einem effektiven Algorithmus bestimmen. Zur Berechnung der allgemeinen Normalform müssen wir die invarianten Faktoren aus der Frobenius-Normalform in Primfaktoren zerlegen, was im Allgemeinen nicht durch einen effektiven Algorithmus erreicht werden kann.

Wir können eine Matrix $A \in M_n(\Bbbk)$ auch als Matrix über jedem anderen Körper auffassen, der $\Bbbk$ als Teilkörper enthält. Beispielsweise ist eine Matrix mit rationalen Koeffizienten auch eine reelle oder gar komplexe Matrix. Aufgrund der Eindeutigkeitsaussage im Satz 6.15 ändert sich die Frobenius-Normalform nicht, wenn man zu einem größeren Körper übergeht. Das bedeutet umgekehrt, dass die Frobenius-Normalform automatisch eine Darstellung einer gegebenen Matrix mit Koeffizienten aus einem möglichst kleinen Teilkörper von $\Bbbk$ sucht. Auf der anderen Seite kann A über einem größeren Körper eine feinere allgemeine Normalform haben, siehe Übung. Man kann also zwei Matrizen in allgemeiner Normalform über verschiedenen Körpern nicht ohne weiteres ansehen, ob sie von derselben Matrix über einem gemeinsamen Grundkörper stammen.

6.38. Bemerkung. Wir betrachten jetzt den Spezialfall, dass das charakteristische Polynom χ_F komplett in Linearfaktoren zerfällt. Wegen Folgerung 5.22 passiert das auf jeden Fall dann, wenn der Grundkörper $\Bbbk$ algebraisch abgeschlossen ist, wegen des Fundamentalsatzes 1.61 also insbesondere für $\Bbbk = \mathbb{C}$.

(1) Nach dem Elementarteilersatz 6.30 zerfällt V als $\Bbbk[X]$-Torsionsmodul in die $(X-\lambda)$-Torsionsuntermoduln $\operatorname{Tor}_{X-\lambda} V$. Wir nennen $H_\lambda V =$

$\operatorname{Tor}_{X-\lambda} V$ den *Hauptraum* zum Eigenwert λ. Es folgt

$$\begin{aligned} H_\lambda V = \operatorname{Tor}_{X-\lambda} V &= \bigl\{\, v \in V \bigm| \text{es gibt } k \in \mathbb{N} \text{ mit } v \,.\, (X-\lambda)^k = 0 \,\bigr\} \\ &= \bigl\{\, v \in V \bigm| \text{es gibt } k \in \mathbb{N} \text{ mit } (F - \lambda \operatorname{id}_V)^k(v) = 0 \,\bigr\} \,. \end{aligned}$$

Insbesondere ist $F - \lambda \operatorname{id}_{H_\lambda V}|_{H_\lambda V}$ nicht invertierbar, falls $H_\lambda V \neq \{0\}$, also ist λ nach Lemma 5.33 (3) ein Eigenwert von F.

(2) Nach dem Satz 6.36 finden wir eine Basis vom Hauptraum U_λ zum Eigenwert λ, bezüglich der $F_{H_\lambda V}$ durch eine Block-Diagonalmatrix aus Jordan-Blöcken zum Eigenwert λ besteht. Da jeder Jordan-Block $J(\lambda, \ell)$ eine untere Dreiecksmatrix ist, ist diese Matrix ebenfalls eine Block-Diagonalmatrix.

(3) Mit (1) und (2) finden wir insgesamt eine Basis von V, bezüglich der F durch eine untere Dreiecksmatrix der Form

$$\begin{pmatrix} \begin{smallmatrix} \lambda_1 & 0 & .. & 0 \\ 1 & \lambda_1 & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & & 1 & \lambda_1 \end{smallmatrix} & & 0 \\ & \ddots & \\ 0 & & \begin{smallmatrix} \lambda_k & 0 & .. & 0 \\ 1 & \lambda_k & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & & 1 & \lambda_k \end{smallmatrix} \end{pmatrix}$$

dargestellt wird. Eine solche Matrix heißt „in *Jordan-Normalform,*“ oder kurz *Jordan-Matrix*.

Wir erinnern uns an den Begriff der „Trigonalisierbarkeit“ von Endomorphismen und Matrizen aus Definition 5.3. Der folgende Satz ist analog zum Satz 5.34 über Diagonalisierbarkeit.

6.39. Satz (Trigonalisierbarkeit, Jordan-Normalform)**.** *Es sei V ein endlich-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_{\Bbbk} V$. Dann sind die folgenden Aussagen äquivalent.*

(1) *Der Endomorphismus F ist trigonalisierbar.*
(2) *Das charakteristische Polynom χ_F zerfällt vollständig in Linearfaktoren.*
(3) *Das Minimalpolynom μ_F zerfällt vollständig in Linearfaktoren.*
(4) *V zerfällt in eine direkte Summe von Haupträumen zu Eigenwerten von F.*
(5) *Der Endomorphismus F lässt sich bezüglich einer geeigneten Basis als Jordan-Matrix schreiben.*

Dabei ist die Jordan-Matrix in (5) *bis auf die Reihenfolge der einzelnen Jordan-Blöcke eindeutig, die zugehörige Basis jedoch nicht.*

Analoge Aussagen gelten für quadratische Matrizen über $\Bbbk$.

Beweis. Wir zeigen hier nur die Aussagen über Endomorphismen.

Zu „(1) $\Longrightarrow$ (2)“ sei F dargestellt durch eine Dreiecksmatrix $A \in M_n(\Bbbk)$. Wir wenden Folgerung 4.17 (2) auf die Dreiecksmatrix $X \cdot E_n - A$ an und erhalten die Zerlegung

$$\chi_F(X) = \det(X \cdot E_n - A) = \prod_{i=1}^{n}(X - a_{ii}) \;.$$

Zu „(2) $\Longrightarrow$ (3)“ benutzen wir Folgerung 5.31 zum Satz von Cayley-Hamilton, wonach $\mu_F \mid \chi_F$, und Bemerkung 6.21 (2), woraus folgt, dass mit χ_F auch μ_F vollständig in Linearfaktoren zerfällt.

Zu „(3) $\Longrightarrow$ (4)“ schreiben wir wieder V als $\Bbbk[X]$-Torsionsmodul, so dass X wie F wirkt. Nach dem Elementarteilersatz 6.30 zerfällt V in eine direkte Summe von Moduln vom Typ $\Bbbk[X]/(P^\ell)$ mit $P \in \mathcal{P}(\Bbbk[X])$ und $P \mid \mu_F$. Wenn μ_F in Linearfaktoren zerfällt, ist ein lineares Polynom, also gilt $P = X - \lambda$ für ein geeignetes $\lambda \in \Bbbk$. Damit können wir wie in Bemerkung 6.38 (1) zeigen, dass V in eine direkte Summe von Haupträumen zerfällt.

Den Schritt „(4) $\Longrightarrow$ (5)“ haben wir in Bemerkung 6.38 (2), (3) gezeigt, und „(5) $\Longrightarrow$ (1)“ folgt, da die Jordan-Matrix aus Bemerkung 6.38 (3) eine untere Dreiecksmatrix ist. □

Wir listen einige Folgerungen aus diesem Satz und dem Satz 6.36 über die allgemeine Normalform auf.

6.40. Folgerung. *Es sei $\Bbbk$ ein Körper, V ein endlich-dimensionaler $\Bbbk$-Vektorraum und $F \in \operatorname{End}_\Bbbk V$.*

(1) *Es sei $\Bbbk$ algebraisch abgeschlossen, zum Beispiel $\Bbbk = \mathbb{C}$, dann lässt sich F durch eine Matrix in Jordan-Normalform darstellen.*
(2) *Wenn F trigonalisierbar ist, dann ist die allgemeine Normalform identisch mit der Jordan-Normalform, und daher eine Dreiecksmatrix.*
(3) *Wenn F diagonalisierbar ist, ist die allgemeine Normalform identisch mit der Jordan-Normalform, und auch wieder eine Diagonalmatrix.*

Die Punkte (2) und (3) zeigen, dass die allgemeine Normalform tatsächlich für diagonalisierbare oder trigonalisierbare Matrizen im Sinne unser Forderung (2) am Anfang des Kapitels die einfachste Darstellung findet.

Beweis. Punkt (1) folgt aus Satz 6.39, da χ_F nach Folgerung 5.22 in Linearfaktoren zerfällt.

Nach Satz 6.39 lässt sich ein trigonalisierbarer Endomorphismus F in Jordan-Normalform A bringen. Da die Jordan-Normalform ein Spezialfall der allgemeinen Normalform ist, liefert die allgemeine Normalform für F aufgrund der Eindeutigkeitsaussage in Satz 6.36 automatisch die Jordan-Matrix A, die ja selbst auch eine Dreiecksmatrix ist, also gilt (2).

Wenn F sogar diagonalisierbar ist, folgt analog aus der Eindeutigkeitsaussage in Satz 6.36, dass die allgemeine Normalform Diagonalgestalt hat, denn jede Diagonalmatrix ist in Jordan-Normalform, also erst recht in allgemeiner Normalform. Also gilt (3). □

Zum Schluss des Kapitels skizzieren wir zwei Anwendungen der Jordan-Normalform im Zusammenhang mit Analysis.

Es sei $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$, und es sei $(p_n)_{n\in\mathbb{N}}$ eine Folge in $\Bbbk$. Eine *Potenzreihe* in einer Variablen X ist ein Ausdruck der Form

$$P(X) = \sum_{n=0}^{\infty} p_n\, X^n\ ,$$

vergleiche Definition 5.9. Im Unterschied zu einem Polynom ist es erlaubt, dass beliebig viele p_i von 0 verschieden sind.

In der Analysis definiert man den Begriff der *Konvergenz* einer Potenzreihe an einer Stelle $x \in \Bbbk$; das Gegenteil davon ist *Divergenz.* Man zeigt dann, dass es einen Konvergenzradius $\rho \in [0,\infty]$ in Abhängigkeit von den Koeffizienten $(p_n)_n$ gibt, so dass

$$\begin{aligned} |x| < \rho \qquad &\Longrightarrow \qquad P(x) \text{ konvergiert, und} \\ |x| > \rho \qquad &\Longrightarrow \qquad P(x) \text{ divergiert.} \end{aligned}$$

Im Fall $|x| = \rho$ ist sowohl Konvergenz als auch Divergenz möglich.

Ähnlich wie in Proposition 5.27 ist es möglich, Matrizen $A \in M_n(\Bbbk)$ in Potenzreihen einzusetzen. Es sei $C = B^{-1} \cdot A \cdot B \in M_n(\mathbb{C})$ die Jordan-Normalform über $\mathbb{C}$ von A. Die Rechnung

$$\begin{aligned} P(A) &= \sum_{i=0}^{\infty} p_n \big(B \cdot C \cdot B^{-1}\big)^n = \sum_{i=0}^{\infty} p_n\, B \cdot C^n \cdot B^{-1} \\ &= B \cdot \left(\sum_{i=0}^{\infty} p_n\, C^n\right) \cdot B^{-1} = B \cdot P(C) \cdot B^{-1} \end{aligned}$$

zeigt, dass $P(A)$ genau dann konvergiert, wenn $P(C)$ konvergiert.

Als nächstes überlegt man sich, dass man jeden Jordanblock einzeln behandeln kann, da

$$\begin{pmatrix} J(\lambda_1,\ell_1) & & 0 \\ & \ddots & \\ 0 & & J(\lambda_k,\ell_k) \end{pmatrix}^n = \begin{pmatrix} J(\lambda_1,\ell_1)^n & & 0 \\ & \ddots & \\ 0 & & J(\lambda_k,\ell_k)^n \end{pmatrix}\ .$$

In Übung 3 von Blatt 6 haben Sie gezeigt, dass

$$J(\lambda,\ell)^n = \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}^n = \begin{pmatrix} \binom{n}{0}\lambda^n & & & 0 \\ \binom{n}{1}\lambda^{n-1} & \binom{n}{0}\lambda^n & & \\ \vdots & \ddots & \ddots & \\ \binom{n}{\ell-1}\lambda^{n+1-\ell} & \dots & \binom{n}{1}\lambda^{n-1} & \binom{n}{0}\lambda^n \end{pmatrix}\ .$$

Da die Binomialkoeffizienten langsamer wachsen als die Potenzen r^n für alle $r > 1$, kann man zeigen, dass $P\bigl(J(\lambda,\ell)\bigr)$ konvergiert, wenn $|\lambda| < \rho$, und divergiert, wenn $|\lambda| > \rho$. Im ersten Fall gilt sogar

$$P\bigl(J(\lambda,\ell)\bigr) = \begin{pmatrix} \frac{1}{0!}\,P(\lambda) & & & 0 \\ \frac{1}{1!}\,P'(\lambda) & \frac{1}{0!}\,P(\lambda) & & \\ \vdots & \ddots & \ddots & \\ \frac{1}{(\ell-1)!}\,P^{(\ell-1)}(\lambda) & \cdots & \frac{1}{1!}\,P'(\lambda) & \frac{1}{0!}\,P(\lambda) \end{pmatrix} .$$

Wir fassen zusammen.

6.41. Proposition. *Es sei P eine Potenzreihe über $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$ mit Konvergenzradius $\rho > 0$ und $A \in M_n(\Bbbk)$. Wenn alle Eigenwerte von A vom Betrag kleiner als ρ sind, dann konvergiert die Reihe $P(A)$. Hat ein Eigenwert größeren Betrag als ρ, dann divergiert sie.* □

Es ist also nicht einmal nötig, die Eigenwerte exakt zu bestimmen. Es reicht, den Betrag der Nullstellen des charakteristischen Polynoms χ_F gegen ρ abzuschätzen. Das kann in Spezialfällen deutlich leichter sein. Auch die Jordan-Normalform selbst taucht in der Formulierung der Proposition nicht auf.

6.42. Beispiel. Der *Arcustangens* ist die Umkehrfunktion der Funktion

$$\tan = \frac{\sin}{\cos} \colon \mathbb{C} \setminus \Bigl\{(2n+1)\,\frac{\pi}{2} \Bigm| n \in \mathbb{Z}\Bigr\} \longrightarrow \mathbb{C}$$

und wird dargestellt durch die Reihe

$$\arctan(X) = X - \frac{1}{3}\,X^3 + \frac{1}{5}\,X^5 - \frac{1}{7}\,X^7 + \dots$$

mit Konvergenzradius 1. Das heißt, für alle $z \in \mathbb{C}$ mit $|z| < 1$ konvergiert

$$z - \frac{1}{3}\,z^3 + \frac{1}{5}\,z^5 - \frac{1}{7}\,z^7 + \dots$$

gegen den Wert $\arctan z$.

Wir betrachten speziell die Matrix

$$A = \begin{pmatrix} \frac{37}{\sqrt{3}} & -16 \\ 27 & -\frac{35}{\sqrt{3}} \end{pmatrix} .$$

Ihre Einträge sind so groß, dass man erst einmal nicht glaubt, dass die Reihe $\arctan(A)$ konvergiert. Wir bestimmen das charakteristische Polynom von A und erhalten

$$\chi_F(X) = \det \begin{pmatrix} X - \frac{37}{\sqrt{3}} & 16 \\ 27 & X + \frac{35}{\sqrt{3}} \end{pmatrix} = X^2 - \frac{2}{\sqrt{3}}X + \frac{1}{3} = \Bigl(X - \frac{1}{\sqrt{3}}\Bigr)^2 .$$

Da der einzige Eigenwert von A gleich $\frac{1}{\sqrt{3}} < 1$ ist, konvergiert die Reihe $\arctan(A)$. In der Tat ist

$$A = \begin{pmatrix} \sqrt{3} & 4 \\ 2 & 3\sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 1 & \frac{1}{\sqrt{3}} \end{pmatrix} \cdot \begin{pmatrix} 3\sqrt{3} & -4 \\ -2 & \sqrt{3} \end{pmatrix}$$

und

$$\begin{aligned} \arctan(A) &= \begin{pmatrix} \sqrt{3} & 4 \\ 2 & 3\sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} \arctan \frac{1}{\sqrt{3}} & 0 \\ \arctan' \frac{1}{\sqrt{3}} & \arctan \frac{1}{\sqrt{3}} \end{pmatrix} \cdot \begin{pmatrix} 3\sqrt{3} & -4 \\ -2 & \sqrt{3} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{3} & 4 \\ 2 & 3\sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} \frac{\pi}{6} & 0 \\ \frac{3}{4} & \frac{\pi}{6} \end{pmatrix} \cdot \begin{pmatrix} 3\sqrt{3} & -4 \\ -2 & \sqrt{3} \end{pmatrix} \\ &= \begin{pmatrix} 9\sqrt{3} - \frac{3\pi}{2} & \frac{\sqrt{3}\pi}{2} - 12 \\ \frac{81}{4} & \frac{\pi}{6} - 9\sqrt{3} \end{pmatrix} . \end{aligned}$$

Dabei haben wir benutzt, dass

$$\tan \frac{\pi}{6} = \sin \frac{\pi}{6} \Big/ \cos \frac{\pi}{6} = \frac{1}{2} \Big/ \frac{\sqrt{3}}{2} = \frac{1}{\sqrt{3}} ,$$

somit $\arctan \frac{1}{\sqrt{3}} = \frac{\pi}{6}$, und dass $\arctan'(z) = \frac{1}{1+z^2}$, somit $\arctan' \frac{1}{\sqrt{3}} = \frac{3}{4}$.

Wir kommen zu einer zweiten Anwendung der Jordan-Normalform. Diesmal geht es um die Lösung von *gewöhnlichen linearen Differentialgleichungssystemen* mit konstanten Koeffizienten. Dabei sei eine Matrix $A \in M_n(\mathbb{R})$ gegeben. Gesucht sind Funktionen $f_1, \ldots, f_n \colon \mathbb{R} \to \mathbb{R}$, so dass gilt

$$(*) \qquad \begin{pmatrix} f_1' \\ \vdots \\ f_n' \end{pmatrix} = A \cdot \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} .$$

Nach dem Satz von Picard-Lindelöff gibt es zu jedem Anfangsvektor $v \in \mathbb{R}^n$ und zu jedes Startzeit $t_0 \in \mathbb{R}$ eine eindeutige Lösung

$$f = \left(\begin{smallmatrix} f_1 \\ \vdots \\ f_n \end{smallmatrix}\right) \colon \mathbb{R} \longrightarrow \mathbb{R}^n$$

von (*) mit $f(t_0) = v$. Diese Lösung ist auf ganz $\mathbb{R}$ definiert und beliebig oft differenzierbar. Außerdem ist das Differentialgleichungssystem zeitunabhängig, das heißt, für alle $s \in \mathbb{R}$ erhalten wir weitere Lösungen

$$f(\,\cdot\, + s) \in L \qquad \text{mit} \qquad t \longmapsto f(t+s) \in \mathbb{R}^n .$$

Unter einer *Fundamentallösung* versteht man eine Abbildung $F \colon t \to M_n(\mathbb{R})$ mit $F(0) = E_n$, so dass für alle $v \in \mathbb{R}^n$ die Abbildung

$$t \longmapsto F(t) \cdot v \in \mathbb{R}^n$$

eine Lösung von (*) mit Anfangswert v bei $t = 0$ ist. Die Fundamentallösung erfüllt die Gleichung

$$F'(t) = A \cdot F(t)$$

mit $F(0) = E_n$. Wegen des Satzes von Picard-Lindelöff existiert sie und ist eindeutig bestimmt. Dann ist für alle t_0 die Abbildung

$$t \longmapsto F(t - t_0) \cdot v \in \mathbb{R}^n$$

die eindeutig bestimmte Lösung von (*), die zur Zeit t_0 den Wert v annimmt.

Als Ansatz wählen wir $F(t) = \exp(tA)$. Die Exponentialreihe hat Konvergenzradius $\rho = \infty$, wegen Proposition 6.41 konvergiert $\exp(tA)$ also für alle $t \in \mathbb{R}$. Für $t = 0$ gilt

$$\exp(0\,A) = A^0 = E_n \;.$$

Außerdem erwarten wir, dass

$$\frac{d}{dt} \exp(tA) = A \cdot \exp(tA) \;.$$

Im folgenden gehen wir zu Funktionen $f\colon \mathbb{R} \to \mathbb{C}$ über, damit wir mit komplexen Matrizen und ihren Jordan-Normalformen rechnen können. Es sei $B \in GL(n, \mathbb{C})$ invertierbar, und es sei $f \in L$ eine Lösung von (*), dann ist $g = B \cdot f\colon \mathbb{R} \to \mathbb{C}^n$ eine Lösung des Differentialgleichungssystems

$$g' = B \cdot f' = B \cdot A \cdot f = (B \cdot A \cdot B^{-1}) \cdot g \;.$$

Sei F eine Fundamentallösung von (*), dann ist entsprechend $B \cdot F(t) \cdot B^{-1}$ eine Fundamentallösung des obigen Systems. Wir können also die Jordan-Normalform von A einsetzen, um die Fundamentallösung für (*) zu bestimmen.

Für festes t gilt

$$\frac{d}{dx} e^{tx} = t\, e^{tx} \;.$$

Für einen Jordanblock $J(\lambda, \ell)$ erhalten wir also

$$\exp\bigl(t\,J(\lambda,\ell)\bigr) = \begin{pmatrix} \frac{1}{0!}\,e^{t\lambda} & & & 0 \\ \frac{t}{1!}\,e^{t\lambda} & \frac{1}{0!}\,e^{t\lambda} & & \\ \vdots & \ddots & \ddots & \\ \frac{t^{\ell-1}}{(\ell-1)!}\,e^{t\lambda} & \cdots & \frac{t}{1!}\,e^{t\lambda} & \frac{1}{0!}\,e^{t\lambda} \end{pmatrix} .$$

Man überprüft jetzt leicht, dass dann tatsächlich

$$\frac{d}{dt} \exp\bigl(t\,J(\lambda,\ell)\bigr) = J(\lambda,\ell) \cdot \exp\bigl(t\,J(\lambda,\ell)\bigr) \;.$$

Da wir jede reelle Matrix über $\mathbb{C}$ in Jordan-Normalform bringen können, erhalten wir das folgende Ergebnis.

6.43. Proposition. *Es sei $A \in M_n(\mathbb{R})$. Dann ist*

$$t \longmapsto \exp(tA)$$

die Fundamentallösung für das Differentialgleichungssystem (*).

6.44. Beispiel. Solche Differentialgleichungssysteme kommen auch in der Physik gelegentlich vor. Sei beispielsweise $u''+a\,u'+b\,u=0$ die Bewegungsgleichung einer linear gedämpften Schwingung. Der Term $a\,u'$ mit der Reibungskonstante $a \geq 0$ beschreibt die Dämpfung des Systems, der Term $b\,u$ mit der Federkonstante $b > 0$ gibt die Rückstellkraft an. Wir setzen $f_1 = u$, führen eine neue Funktion $f_2 = u' = f_1'$ ein, und erhalten das Differentialgleichungssystem

$$f' = \begin{pmatrix} u' \\ u'' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix} \cdot \begin{pmatrix} u \\ u' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix} \cdot f\ .$$

In physikalisch sinnvollen Situationen gilt $a \geq 0$ und $b > 0$. Das charakteristische Polynom ist $\chi_A(X) = X^2 + aX + b$, und wir unterscheiden drei Fälle.

(1) Falls $a^2 > 4b$, hat A zwei reelle Eigenwerte $-\frac{a}{2} \pm \frac{\sqrt{a^2-4b}}{2}$. In der Tat erhalten wir für das ursprüngliche Problem zwei linear unabhängige Lösungen

$$u_1(t) = e^{-\left(\frac{a}{2}-\frac{\sqrt{a^2-4b}}{2}\right)t} \qquad \text{und} \qquad u_2(t) = e^{-\left(\frac{a}{2}+\frac{\sqrt{a^2-4b}}{2}\right)t}\ .$$

Alle anderen Lösungen sind Linearkombinationen dieser Lösungen. Im physikalisch relevanten Fall $a > 0$ und $0 < 4b < a^2$ klingen beide Lösungen exponentiell schnell ab.

(2) Falls $a^2 < 4b$, sind die obigen zwei Eigenwerte komplex. Wir erhalten zwei reelle Lösungen

$$u_1(t) = e^{-\frac{at}{2}} \cos\left(\frac{\sqrt{4b-a^2}}{2}\,t\right) \quad \text{und} \quad u_2(t) = e^{-\frac{at}{2}} \sin\left(\frac{\sqrt{4b-a^2}}{2}\,t\right)\ .$$

Diese Lösungen sind Schwingungen mit der Frequenz $\frac{\sqrt{4b-a^2}}{2}$, deren Amplitude im relevanten Fall $a > 0$ exponentiell abklingt.

(3) Im Grenzfall $a^2 = 4b$ erhalten wir ebenfalls zwei linear unabhängige Lösungen

$$u_1(t) = e^{-\frac{at}{2}} \qquad \text{und} \qquad u_2(t) = t\,e^{-\frac{at}{2}}\ ,$$

denn in der Tat gilt auch

$$\begin{aligned} u_2'(t) &= e^{-\frac{at}{2}} - \frac{at}{2}\,e^{-\frac{at}{2}}\ , \\ u_2''(t) &= -a\,e^{-\frac{at}{2}} + \frac{a^2t}{4}\,e^{-\frac{at}{2}} = -a\,e^{-\frac{at}{2}} + \left(\frac{a^2}{2} - b\right) t\,e^{-\frac{at}{2}} \\ &= -a\,u_2'(t) - b\,u_2(t)\ . \end{aligned}$$

Im Grenzfall (3) klingt die Amplitude genauso schnell wie in (2) ab, ohne dass es zu Schwingungen kommt. Außerdem klingen die Lösungen in (3) schneller ab als die „langsame“ Lösung u_1 im Fall (1). Daher versucht man in technischen Anwendungen, zum Beispiel bei Stoßdämpfern, möglichst nahe an den Grenzfall (3) heranzukommen.

Wir geben noch die Fundamentallösung im Fall (3) an. Es gilt

$$\begin{pmatrix} 0 & 1 \\ -\frac{a^2}{4} & -a \end{pmatrix} = \begin{pmatrix} 1 & \frac{a}{2} \\ 0 & -\frac{a^2}{4} \end{pmatrix} \cdot \begin{pmatrix} -\frac{a}{2} & 0 \\ 1 & -\frac{a}{2} \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{2}{a} \\ 0 & -\frac{4}{a^2} \end{pmatrix} ,$$

$$\begin{aligned} \exp\left(t \begin{pmatrix} 0 & 1 \\ -\frac{a^2}{4} & -a \end{pmatrix}\right) &= \begin{pmatrix} 1 & \frac{a}{2} \\ 0 & -\frac{a^2}{4} \end{pmatrix} \cdot \begin{pmatrix} e^{-\frac{at}{2}} & 0 \\ t\,e^{-\frac{at}{2}} & e^{-\frac{at}{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{2}{a} \\ 0 & -\frac{4}{a^2} \end{pmatrix} \\ &= \begin{pmatrix} e^{-\frac{at}{2}} + \frac{at}{2}\,e^{-\frac{at}{2}} & t\,e^{-\frac{at}{2}} \\ -\frac{a^2 t}{4}\,e^{-\frac{at}{2}} & e^{-\frac{at}{2}} - \frac{at}{2}\,e^{-\frac{at}{2}} \end{pmatrix} . \end{aligned}$$

Man überprüft leicht, dass man für $t = 0$ die Einheitsmatrix erhält. In der ersten Zeile stehen zwei Linearkombinationen der unter (3) genannten Lösungen. In der zweiten Zeile stehen jeweils ihre Ableitungen.

KAPITEL 7

Vektorräume mit Skalarprodukt

In diesem Kapitel betrachten wir Vektorräume über $\Bbbk = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$ mit Skalarprodukt. Wir haben bereits in Abschnitt 1.4 über Euklidische Geometrie gesehen, dass man mit Hilfe des Standard-Skalarproduktes Längen und Winkel bestimmen kann. In diesem Abschnitt sehen wir, dass man auch Volumina von einfachen geometrischen Objekten mit Hilfe des Skalarproduktes definieren kann. Auch in der Physik spielen Skalarprodukte eine große Rolle.

Am Ende von Abschnitt 2.5 haben wir gesehen, dass man mit Orthonormalbasen besonders gut rechnen kann, insbesondere braucht man die Inverse der Basisabbildung nicht umständlich auszurechnen. In diesem Kapitel konstruieren wir systematisch Orthogonalbasen und entsprechende Basen für Vektorräume mit Skalarprodukt über $\mathbb{C}$ oder $\mathbb{H}$, mit denen man entsprechend einfach arbeiten kann.

Am Ende von Abschnitt 2.3 haben wir den Dualraum V^* eines Vektorraums V kennengelernt. Wir führen hier auch noch den sogenannten „Antidualraum“ ein und zeigen, wie beide über ein Skalarprodukt mit V identifiziert werden können, wenn V endlich-dimensional ist.

Lineare Abbildungen, die ein Skalarprodukt invariant lassen, heißen „lineare Isometrien“. Lineare Isometrien sind ein Spezialfall sogenannter „normaler Abbildungen“. Wir zeigen, dass normale Abbildungen bezüglich geeigneter Orthogonalbasen durch spezielle Matrizen dargestellt werden können. Dadurch erhalten wir einen einfacheren Zugang zur Klassifikation von Isometrien Euklidischer Vektorräume, vergleiche dazu die Überlegungen am Ende der Abschnitte 1.5 und 1.6.

Im folgenden sei stets $\Bbbk = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$, wenn nicht anders angegeben.

7.1. Skalarprodukte

In Definition 1.51 (1) haben wir das Standard-Skalarprodukt auf dem Vektorraum $\mathbb{R}^n$ eingeführt als

$$\langle x, y \rangle = \sum_{a=1}^{n} x_a y_a \qquad \text{für alle } x,\ y \in \mathbb{R}^n\ .$$

Wir haben in 1.51 (2) und (3) gesehen, wie man mit dem Skalarprodukt Längen von Vektoren und Winkel zwischen Vektoren definieren kann. Insbesondere ist

$$\|x\|^2 = \langle x, x\rangle = \sum_{a=1}^{n} x^2 \geq 0 ,$$

da die rechte Seite eine Summe von Quadraten ist.

Wenn wir diese Definition unverändert auf $\mathbb{C}^n$ oder $\mathbb{H}^n$ übertragen, haben wir ein Problem, denn für $z \in \mathbb{C}$ gilt $z^2 \in \mathbb{R}$ mit $z^2 \geq 0$ nur dann, wenn bereits z eine reelle Zahl ist, also eine Zahl mit $\operatorname{Im} z = 0$. Wir erinnern uns daher an die Überlegung, die zu Definition 1.62 geführt hat, siehe auch Bemerkung 1.63 (1): es gilt

$$\bar{z} \cdot z = (x - yi)(x + yi) = x^2 + y^2 = \|z\|^2 \geq 0 \qquad \text{für alle } z \in \mathbb{C} .$$

Völlig analog gilt nach Satz 1.71 (7), dass

$$\bar{q} \cdot q = \|q\|^2 \geq 0 \qquad \text{für alle } q \in \mathbb{H} ,$$

dabei bezeichnet $\|\cdot\|$ in beiden Gleichungen die Euklidische Norm auf $\mathbb{C} \cong \mathbb{R}^2$ beziehungsweise $\mathbb{H} \cong \mathbb{R}^4$.

7.1. Bemerkung. Wir führen auch auf $\mathbb{R}$ eine „Konjugation“ ein durch

$$\bar{t} = t \qquad \text{für alle } t \in \mathbb{R} .$$

Dann gilt für r, $s \in \mathbb{k} = \mathbb{R}$, $\mathbb{C}$ und $\mathbb{H}$ gleichermaßen

(1) $$\overline{r + s} = \bar{r} + \bar{s} ,$$

(2) $$\overline{r \cdot s} = \bar{s} \cdot \bar{r} ,$$

(3) $$\bar{\bar{r}} = r ,$$

(4) $$\bar{r} = r \qquad \Longleftrightarrow \qquad r \in \mathbb{R} \subset \mathbb{k} ,$$

(5) $$\bar{r} \cdot r \geq 0 \qquad \text{und} \qquad \bar{r} \cdot r = 0 \quad \Longleftrightarrow \quad r = 0 ,$$

Aufgrund der Eigenschaften (1) und (2) nennen wir die Konjugation einen *Antiautomorphismus* von $\mathbb{k}$, da sie die Reihenfolge der Faktoren in einem Produkt umdreht. Wegen (5) definieren wir den Absolutbetrag

$$|r| = \sqrt{\bar{r} \cdot r} \in \mathbb{R} ,$$

wie in der Cauchy-Schwarz-Ungleichung 1.53 für $\mathbb{k} = \mathbb{R}$, beziehungsweise in den Definition 1.62 und 1.72 für $\mathbb{k} = \mathbb{C}$ und $\mathbb{H}$. Eigenschaften des komplexen Absolutbetrages haben wir in Bemerkung 1.63 zusammengestellt. Die entsprechenden Aussagen über den quaternionischen Absolutbetrag lassen sich analog zeigen.

7.2. Definition. Es sei $\mathbb{k} = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$. Es sei V ein Rechts-$\mathbb{k}$-Vektorraum und W ein Links-$\mathbb{k}$-Vektorraum. Eine Abbildung $\varphi\colon V \to W$ heißt ($\mathbb{k}$-) *semilinear* oder *antilinear,* wenn für alle u, $v \in V$ und alle $r \in \mathbb{k}$ gilt

(L1) $$\varphi(u + v) = \varphi(u) + \varphi(v) \qquad (\textit{Additivität}),$$

($\overline{\text{L2}}$) $$\varphi(v \,.\, r) = \bar{r} \cdot \varphi(v) \qquad (\textit{Antihomogenität}).$$

Analog definieren wir anti- oder semilineare Abbildungen von einem Links- in einen Rechtsvektorraum.

Wir benuzten die obigen Begriffe, um Axiome für Skalarprodukte anzugeben.

7.3. Definition. Es sei V ein Rechts-$\Bbbk$-Vektorraum. Eine Abbildung $S\colon V \times V \to \Bbbk$ heißt *Sesquilinearform,* wenn für alle u, $v \in V$ die Abbildung

$$\text{(S1)}\qquad \begin{aligned} &S(u,\,\cdot\,)\colon V \to \Bbbk && \text{mit} && v \mapsto S(u,v) && \text{linear, und}\\ &S(\,\cdot\,,v)\colon V \to \Bbbk && \text{mit} && u \mapsto S(u,v) && \text{antilinear ist.} \end{aligned}$$

Eine Sesquilinearform $S\colon V \times V \to \Bbbk$ heißt *Hermitesche Form,* wenn für alle u, $v \in V$ gilt

$$\text{(S2)}\qquad S(v,u) = \overline{S(u,v)} \in \Bbbk\,.$$

Eine Hermitesche Form $S\colon V \times V \to \Bbbk$ heißt *positiv semidefinit* oder kurz $S \geq 0$, wenn

$$S(v,v) \geq 0 \qquad \text{für alle } v \in V\,.$$

Sie heißt *positiv definit* oder *positiv,* kurz $S > 0$, wenn für alle $v \in V$ gilt

$$\text{(S3)}\qquad S(v,v) \geq 0 \qquad \text{und} \qquad S(v,v) = 0 \quad\Longleftrightarrow\quad v = 0\,.$$

Ein *Skalarprodukt* oder auch eine *Hermitesche Metrik* auf V ist eine positive definite Hermitesche Form g auf V. Wir nennen (V,g) einen *Euklidischen Vektorraum,* wenn $\Bbbk = \mathbb{R}$, einen *unitären Vektorraum,* wenn $\Bbbk = \mathbb{C}$, und einen *quaternionisch-unitären Vektorraum,* wenn $\Bbbk = \mathbb{H}$.

Die lateinische Vorsilbe „semi“ bedeutet „halb“. Eine semilineare Abbildung erfüllt nur die Hälfte der Axiome, daher der Name. Die Vorsilbe „sesqui“ bedeutet „anderthalb“. Eine Sesquilinearform ist in einem Argument linear, im anderen nur halb, also insgesamt nur anderthalbfach linear.

Man beachte, dass es für $\Bbbk = \mathbb{R}$ keinen Unterschied zwischen semilinear und linear und zwischen sesquilinear und bilinear (also linear in beiden Argumenten) gibt, da die Konjugation auf $\mathbb{R}$ die Identität ist. Genausowenig gibt es über $\mathbb{R}$ einen Unterschied zwischen Hermitesch und symmetrisch ($S(u,v) = S(v,u)$ für alle u, $v \in V$). Um eine einheitliche Notation zu haben, schreiben wir trotzdem die Konjugation auch für $\Bbbk = \mathbb{R}$ immer mit.

7.4. Bemerkung. Wir wollen uns überlegen, dass die Definitionen 7.2 und 7.3 sinnvoll sind.

(1) Semilineare Abbildungen sind mit den Vektorraumaxiomen verträglich. Wir prüfen insbesondere die Verträglichkeit mit dem Assoziativgesetz (M1). Für $\varphi\colon V \to W$ wir oben und $v \in V$, r, $s \in \Bbbk$ gilt

$$\begin{aligned} \varphi\bigl((u\,.\,r)\,.\,s\bigr) = \bar{s} \cdot \varphi(u\,.\,r) = \bar{s} \cdot \bigl(u \cdot \varphi(v)\bigr)\\ = (\bar{s} \cdot \bar{r}) \cdot \varphi(v) = (\overline{r \cdot s}) \cdot \varphi(v) = \varphi\bigl(v\,.\,(r \cdot s)\bigr)\,. \end{aligned}$$

Dabei haben wir die Eigenschaft (2) der Konjugation aus Bemerkung 7.1 und die Antihomogenität ($\overline{\text{L2}}$) ausgenutzt. Die Verträglichkeit mit den anderen Axiomen zeigt man entsprechend.

(2) Sei jetzt S eine Sesquilinearform auf V. Die Homogenität (L2) im zweiten Argument ist mit der Antihomogenität ($\overline{\text{L2}}$) im ersten Argument verträglich, denn für u, $v \in V$ und r, $s \in \Bbbk$ gilt

$$\begin{aligned} S(u \,.\, r, v \,.\, s) &= \bar{r} \cdot S(u, v \,.\, s) = \bar{r} \cdot \big(S(u,v) \cdot s\big) \\ &= \big(\bar{r} \cdot S(u,v)\big) \cdot s = S(u \,.\, r, v) \cdot s = S(u \,.\, r, v \,.\, s) \;. \end{aligned}$$

Ohne Antihomogenität in einem der Argumente hätten wir für $\Bbbk = \mathbb{H}$ im mittleren Schritt Probleme bekommen, denn wir hätten r und s zur selben Seite herausziehen und dann die Reihenfolge von r und s im Produkt vertauschen müssen. Aber das ist nicht einzige Grund dafür, Sesquilinearformen zu betrachten.

(3) Wenn S Hermitesch und linear im zweiten Argument ist, folgt Semilinearität im ersten Argument. Wir überprüfen nur Antihomogenität mit der folgenden Rechnung: Für u, $v \in V$ und $r \in \Bbbk$ gilt

$$S(u \,.\, r, v) = \overline{S(v, u \,.\, r)} = \overline{S(v,u) \cdot r} = \bar{r} \cdot \overline{S(v,u)} = \bar{r} \cdot S(u,v) \;.$$

(4) Der Hauptgrund dafür, dass wir mit Sesquilinear- statt mit Bilinearformen arbeiten, ist der folgende. Wenn wir zweimal dasselbe Argument $v \in V$ einsetzen, gilt

$$S(v,v) = \overline{S(v,v)} \qquad \Longrightarrow \qquad S(v,v) \in \mathbb{R}$$

nach (S2) und Bemerkung 7.1 (4). Insbesondere können wir nun verlangen, dass $S(v,v) \geq 0$. Hätten wir $S(u,v) = S(v,u)$ gefordert, so erhielten wir für $\Bbbk = \mathbb{C}$ ein Element in $\mathbb{C}$, und die Relation „$\geq$“ ist auf $\mathbb{C}$ nicht definiert.

(5) Schließlich gilt für $v \in V$ und $r \in \Bbbk$ noch, dass

$$S(v \,.\, r, v \,.\, r) = \bar{r} \cdot \underbrace{S(v,v)}_{\in \mathbb{R}} \cdot r = (\bar{r} \cdot r) \cdot S(v,v) = \underbrace{|r|^2}_{\geq 0} \cdot S(v,v)$$

wegen Bemerkung 7.1 (5). Also bleibt die Eigenschaft $S(v,v) > 0$ erhalten, wenn man v mit einem Element aus $\Bbbk^\times = \Bbbk \setminus \{0\}$ multipliziert.

In Definition 1.51 haben wir das Standard-Skalarprodukt auf $\mathbb{R}^n$ kennengelernt. Wir wollen jetzt die Standard-Skalarprodukte auf $\mathbb{C}^n$ und $\mathbb{H}^n$ konstruieren. Dazu gehen wir einen kleinen Umweg über adjungierte Matrizen.

7.5. Bemerkung. In Definition 2.80 haben wir bereits die adjungierte Matrix $A^* \in M_{n,m}(\Bbbk)$ zu einer Matrix $A \in M_{m,n}(\Bbbk)$ definiert durch

$$A^* = (\bar{a}_{ji})_{i,j} \;, \qquad \text{wobei} \qquad A = (a_{ij})_{i,j} \;.$$

In Aufgabe 1 von Blatt 10 zur linearen Algebra haben wir bereits gesehen, dass

$$(1) \qquad (A \cdot B)^* = B^* \cdot A^* \qquad \text{für alle } A \in M_{n,m}(\Bbbk) \text{ und alle } B \in M_{m,\ell}(\Bbbk) \;,$$

außerdem gilt $(A^*)^* = A$. Somit hat die Bildung der Adjungierten ähnliche Eigenschaften wie die Konjugation, siehe Bemerkung 7.1 (1)–(3).

In Beispiel 2.31 haben wir den Rechts-$\Bbbk$-Vektorraum $\Bbbk^n = M_{n,1}(\Bbbk)$ der Spalten und den Links-$\Bbbk$-Vektorraum $M_{1,n}(\Bbbk)$ der Zeilen definiert. Nach Bemerkung 2.65 (3), (4) entspricht die Multiplikation mit Skalaren aus $\Bbbk$ genau der Multiplikation mit 1×1-Matrizen. Dann ist die Abbildung

$$(2) \qquad \cdot^* \colon \Bbbk^n \longrightarrow {}^n\Bbbk \qquad \text{mit} \qquad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \longmapsto v^* = (\bar{v}_1, \dots, \bar{v}_n)$$

semilinear, denn für 1×1-Matrizen r ist $r^* = \bar{r}$, und nach (1) oben gilt

$$(v \,.\, r)^* = r^* \cdot v^* = \bar{r} \cdot v^* \,.$$

Die Umkehrabbildung $\cdot^* \colon {}^n\Bbbk \to \Bbbk^n$ ist ebenfalls semilinear.

7.6. Beispiel. Seien u, $v \in \Bbbk^n$ für $\Bbbk = \mathbb{C}$ oder $\mathbb{H}$, dann definieren wir das komplexe und das quaternionische Standard-Skalarprodukt in Analogie zu Definition 1.51 (1) durch

$$\langle u, v \rangle = u^* \cdot v = \sum_{a=1}^{n} \bar{u}_a \cdot v_a \quad \in \quad \Bbbk = M_{1,1}(\Bbbk) \,.$$

Wir überprüfen die Axiome (S1)–(S3). Seien etwa u, v, $w \in \Bbbk^n$ und r, $s \in \Bbbk$, dann gilt

$$\begin{aligned} \langle u \,.\, r + v \,.\, s, w \rangle &= (u \,.\, r + v \,.\, s)^* \cdot w = (\bar{r} \cdot u^* + \bar{s} \cdot v^*) \cdot w \\ &= \bar{r} \cdot (u^* \cdot w) + \bar{s} \cdot (v^* \cdot w) = \bar{r} \, \langle u, w \rangle + \bar{s} \, \langle v, w \rangle \,, \end{aligned}$$

und Linearität im zweiten Argument ist noch einfacher zeigen, somit ist $\langle \,\cdot\,, \,\cdot\, \rangle$ sesquilinear. Aus (1) oben folgt

$$\overline{\langle u, v \rangle} = (u^* \cdot v)^* = v^* \cdot (u^*)^* = v^* \cdot u = \langle v, u \rangle \,,$$

also ist $\langle \,\cdot\,, \,\cdot\, \rangle$ Hermitesch. Schließlich gilt

$$\langle v, v >= v^* \cdot v = \sum_{a=1}^{n} \bar{v}_a \cdot v_a = \sum_{a=1}^{n} |v_a|^2 \geq 0$$

nach Bemerkung 7.1 (1), und Gleichhheit ist nur möglich, wenn alle $v_a = 0$ sind, also wenn $v = 0 \in \Bbbk^n$. Somit ist $\langle \,\cdot\,, \,\cdot\, \rangle$ auch positiv definit, also ein Skalarprodukt auf $\Bbbk^n$.

7.7. Beispiel. Wir geben ein Beispiel aus der Analysis. Dabei sei $V = C^\infty([0,1]; \Bbbk)$ der Raum der unendlich oft differenzierbaren Funktionen auf dem Intervall $[0,1]$ mit Werten in $\Bbbk = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$. Die folgenden Konstruktionen lassen sich auch auf anderen Intervallen an Stelle von $[0,1]$ durchführen.

(1) Das L^2-*Skalarprodukt* ist definiert durch

$$\langle f, g \rangle_{L^2} = \int_0^1 \overline{f(t)} \, g(t) \, dt \in \Bbbk \,.$$

Da f, g stetig sind, sind sie auf $[0,1]$ beschränkt, so dass das Riemann-Integral existiert. Das L^2-Skalarprodukt ist offensichtlich sesquilinear, Hermitesch und positiv semidefinit. Um zu sehen, dass es definit ist, sei $f \neq 0$. Also existiert $t \in [0,1]$ mit $f(t) \neq 0$. Wegen Stetigkeit existiert ein $\varepsilon > 0$, so dass $|f(t)| \geq \varepsilon$ auf $(t-\varepsilon, t+\varepsilon) \cap [0,1]$, somit

$$\langle f, f \rangle_{L^2} = \int_0^1 |f(t)|^2 \, dt \geq \int_{\max(0,t-\varepsilon)}^{\min(1,t+\varepsilon)} \varepsilon^2 \, dt \geq \varepsilon^3 > 0 \; .$$

(2) Wir versuchen es mit

$$\langle\!\langle f, g \rangle\!\rangle = \int_0^1 \overline{f'(t)} \, g'(t) \, dt = \langle f', g' \rangle_{L^2} \in \Bbbk \; .$$

Diese Hermitesche Sesquilinearform ist nur positiv semidefinit, denn für $f \equiv c \in \Bbbk$ konstant gilt $f'(t) \equiv 0$, somit $\langle\!\langle f, f \rangle\!\rangle = 0$.

(3) Wir addieren die beiden obigen Produkte zum (ersten) Sobolev-Skalarprodukt

$$\langle f, g \rangle_{H^1} = \langle f, g \rangle_{L^2} + \langle f', g' \rangle_{L^2} \in \Bbbk \; .$$

Die Summe ist wieder positiv definit, denn für $f \neq 0$ gilt

$$\langle f, f \rangle_{H^1} = \underbrace{\langle f, f \rangle_{L^2}}_{>0} + \underbrace{\langle f', f' \rangle_{L^2}}_{\geq 0} > 0 \; .$$

In Analysis lernen Sie, dass zwei Skalarprodukte $\langle \,\cdot\, , \,\cdot\, \rangle_1$ und $\langle \,\cdot\, , \,\cdot\, \rangle_2$ auf einem endlich-dimensionalen Vektorraum V vergleichbar sind, das heißt, es gibt eine Konstante C, so dass

$$\frac{1}{C} \, \langle v, v \rangle_1 \leq \langle v, v \rangle_2 \leq C \, \langle v, v \rangle_1 \; .$$

Die obigen zwei Skalarprodukte auf $C^\infty([0,1]; \Bbbk)$ sind nicht vergleichbar. Zwar gilt offensichtlich

$$\langle f, f \rangle_{L^2} \leq \langle f, f \rangle_{H^1} \; ,$$

aber für die Folge $f_n(x) = x^n$ gilt

$$\begin{aligned} \langle f_n, f_n \rangle_{L^2} &= \int_0^1 x^{2n} \, dx = \left(\frac{1}{2n+1} \, x^{2n+1} \right) \Big|_{x=0}^1 = \frac{1}{2n+1} \; , \\ \langle f_n', f_n' \rangle_{L^2} &= \int_0^1 n^2 x^{2n-2} \, dx = \left(\frac{n^2}{2n-1} \, x^{2n-1} \right) \Big|_{x=0}^1 = \frac{n^2}{2n-1} \; , \\ \langle f_n, f_n \rangle_{H^1} &= \langle f_n, f_n \rangle_{L^2} + \langle f_n', f_n' \rangle_{L^2} = \frac{2n^3 + n^2 + 2n - 1}{(2n+1)(2n-1)} \; , \end{aligned}$$

und man sieht leicht, dass die Folge

$$\left(\frac{\langle f_n, f_n \rangle_{H^1}}{\langle f_n, f_n \rangle_{L^2}} \right)_n = \left(\frac{2n^3 + n^2 + 2n - 1}{2n - 1} \right)$$

unbeschränkt ist für $n \to \infty$.

Ab sofort verwenden wir für Skalarprodukte die Buchstaben g, h, $\dots$, da wir den Buchstaben B später wieder für Basen und Basisabbildungen benutzen wollen.

7.8. Definition. Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Dann definieren wir die *Norm* zum Skalarprodukt g durch

$$\|v\|_g = \sqrt{g(v,v)} \in \mathbb{R} .$$

Im Falle $\Bbbk = \mathbb{R}$ nennt man diese Norm auch die *Euklidische Norm* zum Skalarprodukt g, vergleiche Definition 1.51 (2).

7.9. Bemerkung. Es sei (V, g) ein Rechts-$\Bbbk$-Vektorraum mit Skalarprodukt. Dann gelten die *Norm-Axiome*

$$\begin{aligned}
&\text{(N1)} && \|v\|_g \geq 0 \quad\text{und}\quad \|v\|_g = 0 \Longleftrightarrow v = 0 && (\textit{Positivität}),\\
&\text{(N2)} && \|v \,.\, r\|_g = |r| \cdot \|v\|_g && (\textit{Homogenität}),\\
&\text{(N3)} && \|v+w\|_g \leq \|v\|_g + \|w\|_g && (\textit{Dreiecksungleichung}),
\end{aligned}$$

für alle v, $w \in V$ und $r \in \Bbbk$. Jede Abbildung $\|\cdot\| \colon V \to \mathbb{R}$, die (N1)–(N3) erfüllt heißt eine *Norm* auf V.

Da g nach (S3) positiv definit ist, folgt (N1). Aus Bemerkung 7.4 (5) ergibt sich unmittelbar (N2). Für jede Zahl $r \in \Bbbk$ gilt $r + \bar r \in \mathbb{R}$ wegen Bemerkung 7.1 (3) und (4). Außerdem folgt

$$|r+\bar r|^2 + |r-\bar r|^2 = (r+\bar r)^2 - (r-\bar r)^2 = 2r\bar r + 2\bar r r = 4\,|r|^2 \ ,$$

so dass insbesondere $r + \bar r \leq 2\,|r|$ gilt. Dabei haben wir $|\bar r| = |r|$ benutzt, siehe Bemerkung 1.63 (5) im Falle $\Bbbk = \mathbb{C}$. Mit der Cauchy-Schwarz-Ungleichung, siehe Satz 1.53 und Satz 7.10 unten, ergibt sich

$$\begin{aligned}
\|v+w\|^2 = g(v+w, v+w) &= \|v\|_g^2 + g(v,w) + g(w,v) + \|w\|_g^2\\
&\leq \|v\|_g^2 + 2\,|g(v,w)| + \|w\|_g^2\\
&\leq \|v\|_g^2 + 2\,\|v\|_g\,\|w\|_g + \|w\|_g^2 = \big(\|v\|_g + \|w\|_g\big)^2 .
\end{aligned}$$

Wurzelziehen liefert die Dreiecksungleichung (N3).

7.10. Satz (Cauchy-Schwarz-Ungleichung)**.** *Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Dann gilt für alle Vektoren v, $w \in V$, dass*

$$|g(v,w)| \leq \|v\| \cdot \|w\| \ .$$

Gleichheit gilt genau dann, wenn v und w linear abhängig sind.

Beweis. Wir passen den Beweis von Satz 1.53 an.

Fall 1: Es sei $v = 0$. Dann gilt

$$g(v,w) = g(0,w) = 0 = \|0\|_g \cdot \|w\|_g = \|v\|_g \cdot \|w\|_g \ .$$

Also gilt Gleichheit, und v und w sind offensichtlich linear abhängig.

Fall 2: Es sei $v \neq 0$, dann folgt $\|v\|_g^2 = g(v,v) > 0$, und wir berechnen

$$\begin{aligned} 0 &\le \left\| w - v \,.\, \frac{g(v,w)}{\|v\|_g^2} \right\|_g = g\left(w - v \,.\, \frac{g(v,w)}{\|v\|_g^2}, w - v \,.\, \frac{g(v,w)}{\|v\|_g^2} \right) \\ &= \|w\|_g^2 - \frac{\overline{g(v,w)}}{\|v\|_g^2}\, g(v,w) - \underbrace{g(w,v)}_{=\overline{g(v,w)}} \frac{g(v,w)}{\|v\|_g^2} + \frac{\overline{g(v,w)}}{\|v\|_g^2} \underbrace{\|v\|_g^2}_{\in \mathbb{R}} \frac{g(v,w)}{\|v\|_g^2} \\ &= \|w\|_g^2 - \frac{|g(v,w)|^2}{\|v\|_g^2} \,. \end{aligned}$$

Hieraus ergibt sich die Ungleichung durch elementare Umformungen.

Wenn Gleichheit gilt, dann folgt aus (S3) (oder äquivalent aus (N1)), dass

$$0 = w - v \,.\, \frac{g(v,w)}{\|v\|_g^2} \,,$$

insbesondere sind v und w dann linear abhängig. Seien umgekehrt v und w linear abhängig, dann gilt $w = v \,.\, r$, da $v \neq 0$ nach Annahme. Wir erhalten also

$$|g(v,w)| = |g(v, v \,.\, r)| = \left| \|v\|_g^2 \, r \right| = \|v\|_g^2 \, |r| = \|v\|_g \cdot \|v \,.\, r\| \,. \qquad \square$$

7.11. Bemerkung. Es sei (V,g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Dann erfüllt die Norm $\|\cdot\|_g$ für alle v, $w \in V$ die *Parallelogramm-Identität*

$$\|v+w\|_g^2 + \|v-w\|_g^2 = 2\,\|v\|_g^2 + 2\,\|w\|_g^2 \,, \tag{1}$$

wie man leicht nachrechnet.

Man kann das Skalarprodukt aus der Norm $\|\cdot\|_g$ zurückgewinnen mit Hilfe der *Polarisationsformeln*

$$g(v,w) = \frac{1}{4}\left(\|v+w\|_g^2 - \|v-w\|_g^2\right) \qquad \text{falls } \Bbbk = \mathbb{R} \,, \tag{2}$$

$$\begin{aligned} g(v,w) &= \frac{1}{4}\left(\|v+w\|_g^2 - \|v-w\|_g^2\right) \\ &\quad - \frac{i}{4}\left(\|v + w \,.\, i\| - \|v - w \,.\, i\|_g^2\right) \qquad \text{falls } \Bbbk = \mathbb{C} \,, \end{aligned} \tag{3}$$

$$\begin{aligned} g(v,w) &= \frac{1}{4}\left(\|v+w\|_g^2 - \|v-w\|_g^2\right) \\ &\quad - \frac{i}{4}\left(\|v + w \,.\, i\| - \|v - w \,.\, i\|_g^2\right) \\ &\quad - \frac{j}{4}\left(\|v + w \,.\, j\| - \|v - w \,.\, j\|_g^2\right) \\ &\quad - \frac{k}{4}\left(\|v + w \,.\, k\| - \|v - w \,.\, k\|_g^2\right) \qquad \text{falls } \Bbbk = \mathbb{H} \,. \end{aligned} \tag{4}$$

Darüberhinaus kann man zeigen, dass jede Norm auf einem $\Bbbk$-Vektorraum V, die die Parallelogrammidentität (1) erfüllt, von einem Skalarprodukt auf V herkommt, das man mit Hilfe der passenden Polarisationsformel berechnen kann.

7.2. Skalarprodukte als Matrizen

In diesem Abschnitt stellen wir Sesquilinearformen auf endlich-dimensionalen Vektorräumen bezüglich einer Basis als Matrizen dar. Wir untersuchen die Eigenschaften dieser Matrizen und geben Kriterien dafür, dass eine solche Matrix ein Hermitesches und positiv definites Skalarprodukt darstellt.

Die darstellende Matrix hat eine besonders einfache Gestalt, wenn die Basisvektoren alle Länge 1 haben und paarweise aufeinander senkrecht stehen. Solche Orthonormalbasen haben wir bereits in Abschnitt 2.5 kennengelernt. Wir lernen ein Verfahren kennen, das Orthonormalbasen mit speziellen Eigenschaften produziert. In diesem Zusammenhang beweisen wir auch ein Kriterium dafür, ob eine Matrix positiv definit ist.

7.12. Definition. Es sei (V, g) ein endlich-dimensionaler $\mathbb{k}$-Vektorraum mit Skalarprodukt, und es sei $B = (b_1, \dots, b_n)$ eine Basis von V. Dann definieren wir die *Gramsche Matrix* $G \in M_n(\mathbb{k})$ von g durch

$$G = \big(g(b_i, b_j)\big)_{i,j} \in M_n(\mathbb{k}) \;.$$

7.13. Beispiel. Wir schränken das L^2-Skalarprodukt aus Beispiel 7.7 (1) auf dem Raum $C^\infty([0,1]; \mathbb{k})$ auf den $(n+1)$-dimensionalen Raum der Polynome P vom Grad $\deg P \leq n$ ein. Als Basis wählen wir die Polynome $f_0(x) = x^0, \dots, f_n(x) = x^n$. Dann erhalten wir als Gramsche Matrix

$$\begin{aligned} G = \big(\langle f_i, f_j \rangle\big)_{i,j} &= \left(\int_0^1 x^i \cdot x^j \, dx\right)_{i,j} = \left(\frac{x^{i+j+1}}{i+j+1}\bigg|_{x=0}^1\right) = \left(\frac{1}{i+j+1}\right)_{i,j} \\ &= \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n+1} \\ \frac{1}{2} & \frac{1}{3} & & \frac{1}{n+2} \\ \vdots & & \ddots & \vdots \\ \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n+1} \end{pmatrix} . \end{aligned}$$

7.14. Definition. Eine quadratische Matrix $A = (a_{ij})_{i,j} = M_n(R)$ heißt *symmetrisch,* wenn $A = A^t$, das heißt, wenn $a_{ij} = a_{ji}$ für alle i, j gilt. Eine quadratische Matrix $A = (a_{ij})_{i,j} = M_n(\mathbb{k})$ heißt *selbstadjungiert* oder *Hermitesch,* wenn $A = A^*$, das heißt, wenn $a_{ij} = \bar{a}_{ji}$ für alle i, j gilt.

Eine Hermitesche Matrix $A = (a_{ij})_{i,j} = M_n(\mathbb{k})$ heißt *positiv semidefinit,* wenn

$$x^* \cdot A \cdot x \geq 0 \qquad \text{für alle } x \in \mathbb{k}^n \;.$$

Sie heißt *positiv definit,* wenn

$$x^* \cdot A \cdot x \geq 0 \quad \text{und} \quad x^* \cdot A \cdot x = 0 \iff x = 0 \qquad \text{für alle } x \in \mathbb{k}^n \;.$$

Die Definition von positiv (semi-) definit ist sinnvoll, da für eine Hermitesche Matrix A gilt, dass

$$x^* \, A \, x = x^* \, A^* \, (x^*)^* = (x^* \, A \, x)^* = \overline{x^* \, A \, x} \quad \in \mathbb{R} \subset \mathbb{k}$$

nach Bemerkung 7.1 (4).

Wir beachten, dass die Begriffe „Hermitesch“, „selbstadjungiert“ und „symmetrisch“ für $\Bbbk = \mathbb{R}$ gleichbedeutend sind. Wir werden im Folgenden auch für $\Bbbk = \mathbb{R}$ die Begriffe „Hermitesch“ und „selbstadjungiert“ verwenden. Dabei benutzt man „Hermitesch“ eher für Matrizen, die Skalarprodukte darstellen, und „selbstadjungiert“ für Matrizen, die Endomorphismen darstellen.

7.15. Bemerkung. Es sei S eine Sesquilinearform auf einem endlich-dimensionalen $\Bbbk$-Vektorraum, und es sei $B = (b_1, \dots, b_n)$ eine Basis von V. Wie in Definition 7.12 betrachten wir die Matrix

$$A = \bigl(S(b_p, b_q)\bigr)_{p,q} \in M_n(\Bbbk) .$$

Wenn S ein Skalarprodukt ist, handelt es sich dabei gerade um die Gramsche Matrix.

(1) Die Matrix A legt S eindeutig fest. Denn seien $v, w \in V$, dann existieren Koordinaten $(x_p)_p$, $(y_q)_q \in \Bbbk^n$, so dass

$$v = B(x) = \sum_{p=1}^{n} b_p . x_p \qquad \text{und} \qquad w = B(y) = \sum_{q=1}^{n} b_q . y_q ,$$

siehe Proposition 2.32. Da S eine Sesquilinearform ist, folgt

$$S(v, w) = S\biggl(\sum_{p=1}^{n} b_p . x_p, \sum_{q=1}^{n} b_q . y_q\biggr) = \sum_{p,q=1}^{n} \bar{x}_p \, S(b_p, b_q) \, y_q = x^* \, A \, y .$$

Umgekehrt liefert die obige Formel zu jeder Matrix $A \in M_n(\Bbbk)$ eine Sesquilinearform S auf V.

(2) Die Sesquilinearform S ist genau dann Hermitesch, wenn die Matrix A Hermitesch ist. Da $S(b_q, b_p) = \overline{S(b_p, b_q)}$, ist die Richtung „$\Rightarrow$“ klar.

Zu „$\Leftarrow$“ seien $v = B(x)$ und $w = B(y) \in V$ wie oben und A sei Hermitesch. Aus (1) folgt

$$S(w, v) = y^* \, A \, x = y^* \, A^* \, (x^*)^* = (x^* \, A \, y)^* = \overline{S(v, w)} .$$

(3) Sei S Hermitesch, dann ist S genau dann positiv (semi-) definit, wenn A positiv (semi-) definit ist. Die Basisabbildung $B \colon \Bbbk^n \to V$ ist bijektiv, also gilt $S(v, v) \geq 0$ wegen (1) genau dann für alle $v \in V$, wenn

$$x^* \, A \, x = S\bigl(B(x), B(x)\bigr)$$

für alle $x \in \Bbbk^n$ gilt. Entsprechend gilt $S(v, v) = 0$ genau dann nur für $v = 0$, wenn $x^* A x = 0$ nur für $x = 0$ gilt. Später lernen wir noch ein etwas griffigeres Kriterium für positive Definitheit kennen, bei dem wir $x^* A x$ nicht für alle $x \in \Bbbk^n$ testen müssen.

(4) Zum Schluss betrachten wir noch das Verhalten der darstellenden Matrix A unter Basiswechsel. Dazu seien $B = (b_1, \dots, b_n)$ und $C = (c_1, \dots, c_n)$ Basen von V. Dann existiert eine Matrix $M = (m_{pq})_{p,q} \in GL(n, \Bbbk)$, so dass

$$c_q = \sum_{p=1}^{n} b_p . m_{pq} ,$$

siehe Bemerkung 2.76. Es sei A wie oben die darstellende Matrix zur Basis B, dann erhalten wir zur Basis C die darstellende Matrix

$$\begin{aligned}\big(S(c_p, c_q)\big)_{p,q} &= \left(S\bigg(\sum_{r=1}^n b_r \,.\, m_{rp}, \sum_{s=1}^n b_s \,.\, m_{sq}\bigg)\right)_{p,q} \\ &= \bigg(\sum_{r,s=1}^n \bar m_{rp}\, S(b_r, b_s)\, m_{sq}\bigg)_{p,q} = M^*\, A\, M \;.\end{aligned}$$

Sei $v = C(s)$ mit $s \in \Bbbk^n$, dann folgt

$$v = \sum_{q=1}^n c_q \,.\, s_q = \sum_{p,q=1}^n b_p \,.\, m_{pq} \,.\, s_q = \sum_{p=1}^n b_p\, x_p \;,$$

wobei $x = M \cdot s \in \Bbbk^n$. In der Tat gilt für $v = C(s) = B(x)$ und $w = C(t) = B(y)$ mit $x = M\, s$ und $y = M\, t$, dass

$$S(v, w) = x^*\, A\, y = (M\, s)^*\, A\, (M\, t) = s^*(M^* A M) t \;.$$

Das Standardskalarprodukt $\langle\, \cdot \,, \cdot \,\rangle$ auf $\Bbbk^n$ aus Beispiel 7.6 wird bezüglich der Standardbasis durch die Einheitsmatrix dargestellt: $\langle e_i, e_j\rangle = \delta_{ij}$. Somit ist die Standardbasis eine Orthonormalbasis für das Standardskalarprodukt, siehe Definition 2.81. Wir wollen den Begriff der Orthonormalbasis jetzt auf beliebige Vektorräume mit Skalarprodukt ausdehnen.

7.16. Definition. Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Ein Tupel $(v_1, \dots, v_k)$ von Elementen von V heißt *orthogonal* oder auch (*paarweise*) *senkrecht,* wenn

$$g(v_i, v_j) = 0 \qquad \text{für alle } i,\ j \text{ mit } i \neq j \;.$$

Wenn $(v_1, \dots, v_k)$ außerdem eine Basis bildet, nennt man diese eine *Orthogonalbasis.*

Dann heißt eine Basis $B = (b_1, \dots, b_n)$ von V eine *Orthonormalbasis* von V, wenn

$$g(b_i, b_j) = \delta_{ij} \;.$$

Eine Orthonormalbasis eines $\Bbbk$-Vektorraums heißt manchmal auch *unitäre Basis* ($\Bbbk = \mathbb{C}$), beziehungsweise *quaternionisch-unitäre Basis* ($\Bbbk = \mathbb{H}$) .

7.17. Bemerkung. Es sei (V, g) ein endlich-dimensionaler Vektorraum mit Skalarprodukt.

(1) Jedes orthogonale Tupel $(v_1, \dots, v_k)$ mit $v_i \neq 0$ für alle i ist linear unabhängig, denn sei

$$0 = \sum_{p=1}^k v_p \,.\, r_p \;,$$

dann folgt für alle q, dass

$$0 = g\Bigl(v_q, \sum_{p=1}^k v_p \,.\, r_p\Bigr) = \sum_{p=1}^k g(v_p, v_q) \cdot r_p = \|v_q\|_g^2 \, r_q \,.$$

Aus $v_q \neq 0$ folgt $\|v_q\|_g \neq 0$, und somit $r_q = 0$. Also sind $v_1, \dots, v_k$ linear unabhängig.

(2) Sei $\dim V = n$, dann bildet ein orthogonales n-Tupel von Vektoren eine Basis, wenn keiner der Vektoren verschwindet, also eine Orthogonalbasis. Das folgt aus (1) und dem Basissätzen 3.3 und 3.4 von Steinitz, siehe auch Aufgabe 2 von Blatt 11 zur Linearen Algebra I.

(3) In Definition 2.33 hatten wir die Koordinatenabbildung als Inverse der Basisabbildung $B\colon \Bbbk^n \to V$ eingeführt. Es sei $B = (b_1, \dots, b_n)$ eine Orthogonalbasis, dann wird die Koordinatenabbildung $B^{-1}\colon V \to \Bbbk^n$ beschrieben durch die Formel

$$B^{-1}(v) = \begin{pmatrix} g(b_1, v) \\ \vdots \\ g(b_n, v) \end{pmatrix} ,$$

denn sei $v = B(x)$ mit $x \in \Bbbk^n$, dann gilt

$$g(b_p, v) = g\Bigl(b_p, \sum_{q=1}^n b_q \,.\, x_q\Bigr) = \sum_{q=1}^n g(b_p, b_q)\, x_q = x_p \,.$$

Eine ähnliche Aussage hat wir in Proposition 2.82 bereits bewiesen. Allerdings haben wir damals umständlich zeigen müssen, dass

$$v = \sum_{p=1}^n b_p \,.\, g(b_p, v) ,$$

da wir die Steinitzsätze und damit auch (2) nicht zur Verfügung hatten.

Im Folgenden bezeichnen wir das Erzeugnis von $v_1, \dots, v_p$ mit $\langle v_1, \dots, v_p \rangle$. Für das Skalarprodukt verwenden wir wieder den Buchstaben g, um Verwechselungen zu vermeiden. Die Axiome (S1)–(S3) bleiben gültig, wenn man g auf einen Unterraum einschränkt. Insbesondere ist also $g|_{\langle v_1,\dots,v_p\rangle \times \langle v_1,\dots,v_p\rangle}$ wieder ein Skalarprodukt, das wir der Kürze halber wieder mit g bezeichnen.

7.18. Satz (Gram-Schmidt-Orthonormalisierungsverfahren)**.** *Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt und $(v_1, \dots, v_n)$ sei eine Basis von V. Dann existieren eindeutig bestimmte Vektoren $b_1, \dots, b_n \in V$, so dass für alle $p = 1, \dots, n$ gilt:*

(1) *$(b_1, \dots, b_p)$ ist eine g-Orthonormalbasis von $\langle v_1, \dots, v_p \rangle \subset V$, und*
(2) *es gilt $g(b_p, v_p) \in \mathbb{R}$ und $g(v_p, b_p) > 0$.*

Dazu konstruiert man b_p induktiv durch

$$b_p = w_p \,.\, \frac{1}{\|w_p\|_g} , \qquad \textit{wobei} \qquad w_p = v_p - \sum_{q=1}^{p-1} b_q \,.\, g(b_q, v_p) \,.$$

Insbesondere erhalten wir am Ende eine Orthonormalbasis $(b_1, \dots, b_n)$ von (V, g). Für viele Anwendungen reicht das, aber manchmal möchten wir die volle Stärke der Eigenschaften (1) und (2) ausnutzen.

Beweis. Wir beweisen den Satz durch Induktion. Für $p = 0$ ist nichts zu zeigen.

Sei also $p \geq 1$, und seien b_1, ..., b_{p-1} bereits konstruiert. Wir beginnen mit der Existenzaussage und definieren w_p wie oben. Nach Voraussetzung liegen b_1, ..., $b_{p-1} \in \langle v_1, \dots, v_{p-1} \rangle$, also folgt

$$w_p = v_p - \sum_{q=1}^{p-1} b_q \,.\, g(b_q, v_p) \in \langle v_1, \dots, v_p \rangle \,.$$

Da die v_q linear unabhängig sind, gilt

$$v_p \notin \langle v_1, \dots, v_{p-1} \rangle = \langle b_1, \dots, b_{p-1} \rangle \,,$$

also auch $w_p \notin \langle v_1, \dots, v_{p-1} \rangle$, insbesondere $w_p \neq 0$, so dass wir b_p wie oben definieren dürfen. Für $q \leq p - 1$ berechnen wir

$$g(b_q, b_p) = g\biggl(b_q, v_p - \sum_{r=1}^{p-1} b_r \,.\, g(b_r, v_p)\biggr) \frac{1}{\|w_p\|_g} = \frac{g(b_q, v_p) - g(b_q, v_p)}{\|w_p\|_g} = 0 \,.$$

Außerdem gilt $\|b_p\| = 1$ nach Konstruktion, und die Vektoren b_1, ..., b_{p-1} sind nach Induktionsvoraussetzung orthogonal und normiert, also ist (1) erfüllt.

Aus (1) und der Konstruktion von b_p folgern wir (2), denn es gilt

$$g(b_p, v_p) = g\biggl(b_p, v_p - \sum_{q=1}^{p-1} b_q \,.\, g(b_q, v_p)\biggr) = g(b_p, w_p) = \|w_p\|_g > 0 \,.$$

Damit ist die Existenz von b_p mit den gewünschten Eigenschaften bewiesen.

Wir kommen zur Eindeutigkeit. Da b_1, ..., b_{p-1} durch (1) und (2) bereits eindeutig bestimmt sind, brauchen wir im Induktionsschritt nur noch die Eindeutigkeit von b_p zu beweisen. Es sei also $v \in \langle v_1, \dots, v_p \rangle$ ein weiterer Vektor, so dass $g(b_q, v) = 0$ für $1 \leq q < p$, $\|v\|_g = 1$ und $g(v, v_p) > 0$. Wir stellen v in der Orthonormalbasis $(b_1, \dots, b_p)$ von $\langle v_1, \dots, v_p \rangle$ dar als

$$v = \sum_{q=1}^{p} b_q \,.\, x_q \,.$$

Dann folgt als erstes $x_q = g(b_q, v) = 0$ für alle $1 \leq q < p$, so dass $v = b_p \cdot x_p$. Es folgt

$$|x_p| = |x_p| \, \|b_p\|_g = \|v\|_g = 1 \,.$$

Da $g(v_p, b_p) = \overline{g(b_p, v_p)} > 0$, gilt außerdem

$$\bar{x}_p \, g(b_p, v_p) = g(b_p \,.\, x_p, v_p) = g(v, v_p) > 0 \,,$$

also auch $\bar{x}_p > 0$, und daher $x_p > 0$. Aber die einzige Zahl $x_p \in \Bbbk$ mit $|x_p| = 1$, $x_p \in \mathbb{R}$ und $x_p > 0$ ist 1. Also folgt $v = b_p$, und die Eindeutigkeit ist ebenfalls gezeigt. $\square$

Es sei $A \in M_n(\mathbb{k})$ eine quadratische Matrix und $r \leq n$, dann schreiben wir $A_r \in M_r(\mathbb{k}_k)$ für den oberen linken $r \times r$-Block

$$A_r = \big((a_{p,q})_{p,q \leq r}\big) = \begin{pmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} \end{pmatrix} .$$

In Folgerung 4.15 (1) haben wir gesehen, dass $\det A^t = \det A$ für alle quadratischen Matrizen $M_n(\mathbb{k})$ gilt, wenn $\mathbb{k} = \mathbb{R}$ oder $\mathbb{C}$ ein Körper ist. Da die Konjugation mit den Rechenoperationen in $\mathbb{k}$ verträglich ist, sieht man anhand der Leibniz-Formel aus Satz 4.13, dass $\det \bar{A} = \overline{\det A}$ gilt. Insgesamt folgt daraus

$$\det A^* = \det \bar{A}^t = \det \bar{A} = \overline{\det A} .$$

7.19. Folgerung. *Es sei $A = (a_{pq})_{p,q} \in M_n(\mathbb{k})$ eine quadratische Matrix. Dann sind die folgenden Aussagen äquivalent.*

(1) *Die Matrix A ist Hermitesch und positiv definit.*
(2) *Es gibt eine obere Dreiecksmatrix B mit reellen, positiven Diagonaleinträgen, so dass $A = B^*B$.*
(3) *Es gibt eine invertierbare Matrix $B \in GL(n, \mathbb{k})$ mit $A = B^*B$.*

Falls $\mathbb{k} = \mathbb{R}$ oder $\mathbb{C}$, sind die obigen Aussagen außerdem äquivalent zu

(4) Sylvester- *oder auch* Hurwitz-Kriterium. *Die Matrix A ist Hermitesch, und für alle $r = 1, \ldots, n$ gilt* $\det A_r > 0$.

In der Analysis benötigt man analog zu (4) ein Kriterium für negative Definitheit. Dazu betrachten wir anstelle einer Hermiteschen Matrix A die Matrix $-A$ und sehen, dass

$$\big(v^*Av \leq 0 \quad \text{und} \quad v^*Av = 0 \Leftrightarrow v = 0\big) \qquad \Longleftrightarrow \qquad (-1)^r \det A_r > 0 \quad \text{für alle } r = 1, \ldots, n .$$

Achtung: Das Sylvester-Kriterium funktioniert nicht für positiv semidefinite Matrizen. Beispielsweise sei

$$A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} ,$$

dann gilt $\det A_1 = a_{11} = 0 \geq 0$ und $\det A_2 = \det A = 0 \geq 0$, aber die Matrix A ist nicht positiv semidefinit, denn $e_2^* A e_2 = -1$.

Beweis. Zu „(1) $\Longrightarrow$ (2)“ fassen wir A als Gramsche Matrix eines Skalarproduktes

$$g(x, y) = x^*Ay \in \mathbb{k}$$

auf $\Bbbk^N$ auf. Wir konstruieren eine Orthonormalbasis $(v_1, \dots, v_n)$ von V mit dem Gram-Schmidt-Verfahren, beginnend mit der Standardbasis $(e_1, \dots, e_n)$. Es sei $B \in M_n(\Bbbk)$ die Basiswechselmatrix, so dass

$$e_q = \sum_{p=1}^{n} v_p \,.\, b_{pq} \,.$$

Dann ist B eine obere Dreiecksmatrix nach Satz 7.18 (1), denn aus $e_q \in \langle e_1, \dots, e_q \rangle = \langle v_1, \dots, v_q \rangle$ folgt $b_{pq} = 0$ für $p > q$. Die Diagonaleinträge sind reell und positiv nach Satz 7.18 (2), denn

$$b_{qq} = g(v_q, v_q \,.\, b_{qq}) = g\left(v_q, \sum_{p=1}^{n} v_p \,.\, b_{pq}\right) = g(v_q, e_q) > 0 \,.$$

Schließlich gilt $A = B^*B$ nach Bemerkung 7.15 (4).

Der Schritt „(2) $\Longrightarrow$ (3)“ folgt für $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$, da B nach Folgerung 4.17 (2) positive Determinante hat und somit invertierbar ist. Über $\mathbb{H}$ überlegen wir uns stattdessen, dass Dreiecksmatrizen mit von 0 verschiedenen Diagonaleinträgen mit dem Gauß-Verfahren 3.25 immer invertiert werden können.

Zu „(3) $\Longrightarrow$ (1)“ überlegen wir uns, dass A Hermitesch ist, da

$$A^* = (B^*B)^* = B^*(B^*)^* = B^*B = A \,.$$

Da B invertierbar ist, ist A positiv definit, denn

$$x^*Ax = x^*B^*Bx = (Bx)^*(Bx) \geq 0 \quad \text{und} \quad x^*Ax = 0 \Longleftrightarrow Bx = 0 \Longleftrightarrow x = 0 \,.$$

Es sei jetzt $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$ ein Körper, so dass wir Determinanten bilden können. Wir schließen „(2) $\Longrightarrow$ (4)“, denn für p, $q \leq r$ gilt

$$a_{pq} = (Be_p)^*(Be_q) = \sum_{s=1}^{p} \sum_{t=1}^{q} \bar{b}_{sp} b_{sq} = \sum_{s,t=1}^{r} \bar{b}_{sp} b_{sq} \,,$$

so dass $A_r = B_r^* B_r$, und daher

$$\det A_r = \det B_r^* \det B_r = |\det B_r|^2 > 0 \,,$$

da der obere linke $r \times r$-Block B_r von B aus dem gleichen Grund wie B oben positive Determinante hat.

Zu „(4) $\Longrightarrow$ (1)“ beweisen wir durch Induktion über r, dass A_r ein Skalarprodukt auf $\Bbbk^r$ definiert. Für $r = 1$ ist das klar, da $a_{11} = \det A_1 > 0$.

Es sei also $r \geq 1$, und A_r definiere ein Skalarprodukt auf $\Bbbk^r$. Wir konstruieren wie oben eine Orthonormalbasis $(v_1, \dots, v_r)$ mit dem Gram-Schmidt-Verfahren, beginnend mit der Standardbasis. Wir definieren

$$w_{r+1} = e_{r+1} - \sum_{p=1}^{r} v_p \,.\, (v_p^* A e_{r+1}) \,,$$

so dass $v_p^* A w_{r+1} = 0$. Da A Hermitesch ist, gilt ebenfalls $w_{r+1}^* A v_p = 0$ für alle $p \le r$. Dann bilden $(v_1, \dots, v_r, w_{r+1})$ eine Basis von $\langle e_1, \dots, e_{r+1} \rangle$. Es sei C_{r+1} die zugehörige Basiswechselmatrix, so dass

$$e_q = \sum_{p=1}^{r} v_p \,.\, c_{pq} + w_{r+1} \,.\, c_{p,r+1} \;.$$

Aus Bemerkung 7.15 (4) folgt
(*)

$$A_{r+1} = C_{r+1}^* D_{r+1} C_{r+1} \;, \qquad \text{wobei } D_{r+1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & w_{r+1}^* A w_{r+1} \end{pmatrix} \;,$$

somit

$$0 < \det A_{r+1} = \left|\det C_{r+1}\right|^2 \, w_{r+1}^* A w_{r+1} \;.$$

Man überprüft jetzt leicht, dass D_{r+1} positiv definit ist, und damit auch A_{r+1}, denn

$$x^* A_{r+1} x = (C_{r+1} x)^* D_{r+1} (C_{r+1} x) \;.$$

Also beschreibt A_{r+1} ebenfalls ein Skalarprodukt auf $\Bbbk^{r+1}$. Damit ist die Behauptung bewiesen. □

7.20. Bemerkung. Wir geben noch eine geometrische Deutung der diversen Konstruktion in den Beweisen von Satz 7.18 und Folgerung 7.19.

(1) Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt und $U \subset V$ ein Unterraum. Wir wählen eine Orthonormalbasis $(e_1, \dots, e_m)$ von U und definieren eine Abbildung

$$p\colon V \to U \qquad \text{durch} \qquad p(v) = \sum_{p=1}^{m} e_p \,.\, g(e_p, v) \in U \;.$$

Für alle $u \in U$ gilt $p(u) = u$, somit gilt $p^2 = p$. Eine Abbildung mit dieser Eigenschaft heißt *Projektion.*

Es gilt $g\big(u, p(v)\big) = g(u, v)$ zunächst einmal für $u = e_1, \dots, e_m$, wie man leicht nachrechnet. Da $(e_1, \dots, e_m)$ eine Basis von U bilden, gilt $g\big(u, p(v)\big) = g(u, v)$ sogar für alle u, mit anderen Worten

$$g\big(u, v - p(v)\big) = 0 \qquad \text{für alle } u \in U \text{ und alle } v \in V \;.$$

Aus diesem Grund nennt man p die *orthogonale* oder *senkrechte* Projektion von V auf den Unterraum U. Man kann zeigen, dass $p(v)$ derjenige Punkt in U ist, für den der Abstand $\|v - p(v)\|_g$ minimal wird (Übung).

(2) Es sei jetzt $\Bbbk = \mathbb{R}$, und es sei A die Gramsche Matrix eines Skalarproduktes g bezüglich einer Basis $(v_1, \dots, v_n)$ von V. Wir wollen durch Induktion über r motivieren, dass $\det A_r$ das Quadrat des Volumens des von den Vektoren $v_1, \dots, v_r$ aufgespannten r-dimensionalen Parallelotops P_r ist. Dabei erinnern wir uns an den Anfang von Abschnitt 4.1,

wo wir entsprechende Überlegungen für Parallelotope maximaler Dimension $r = n$ in $\mathbb{R}^n$ angestellt haben.

In Dimension $r = 1$ sollte das „Volumen“ des Vektors v_1 seine Länge sein. In der Tat gilt

$$\det A_1 = a_{11} = g(v_1, v_1) = \|v_1\|_g^2 \ .$$

Sei jetzt $r \geq 1$. Dann hat das von v_1, …, v_{r+1} aufgespannte Parallelotop P_{r+1} als „Grundfläche“ das Parallelotop P_r vom Volumen $\operatorname{vol}(P_r) = \sqrt{\det A_r}$ nach Induktionsvoraussetzung, und als Höhe den Vektor $w_{r+1} = v_{r+1} - p(v_{r+1})$, dabei ist $p\colon V \to \langle v_1, \dots, v_r\rangle$ die orthogonale Projektion aus (1).

Wie im Beweis der Folgerung 7.19, Schritt (4) $\Rightarrow$ (1), sei e_1, …, e_r eine Orthonormalbasis von $\langle v_1, \dots, v_r\rangle$, und C_{r+1} sei die dortige Basiswechselmatrix. Dann hat C_{r+1} die Blockgestalt

$$C_{r+1} = \begin{pmatrix} B_r & * \\ 0 & 1 \end{pmatrix} \ .$$

Nach Induktionsvoraussetzung und (*) hat also die Grundfläche das Volumen

$$\operatorname{vol}(P_r) = \sqrt{\det A_r} = |\det B_r| = |\det C_{r+1}| \ .$$

Die Länge der Höhe ist $\|w_{r+1}\|_g = \sqrt{w_{r+1}^* A_r w_{r+1}}$, und somit erhalten wir mit (*), dass

$$\operatorname{vol}(P_{r+1}) = |\det C_{r+1}| \cdot \sqrt{w_{r+1}^* A w_{r+1}} = \sqrt{\det A_{r+1}} \ .$$

Damit ist unsere Behauptung gezeigt, allerdings unter der Annahme, dass man das Volumen mit Hilfe der Formel „Grundfläche $\times$ Höhe“ berechnen darf.

Einen alternativen Zugang zur Behauptung $\operatorname{vol} P_r = \sqrt{\det A_r}$ finden Sie in den Übungen.

(3) Im $\mathbb{R}^n$ mit dem Standard-Skalarprodukt hat das von den Vektoren v_1, …, v_r aufgespannte Parallelotop also das Volumen

$$\sqrt{\det\bigl((\langle v_p, v_q\rangle)_{p,q \leq r}\bigr)} = \sqrt{\det\bigl((v_p^* v_q)_{p,q \leq r}\bigr)} \ .$$

Beispielsweise hatten wir in Bemerkung 1.69 eine geometrische Interpretation des Kreuz- und des Spatproduktes gegeben. Dazu hatten wir die Fläche des von $u, v \in \mathbb{R}^3$ aufgespannten Parallelogramms berechnet als

$$\|u \times v\| = \sqrt{\|u\|^2 \, \|v\|^2 - \langle u, v\rangle^2} = \det \begin{pmatrix} \langle u, u\rangle & \langle u, v\rangle \\ \langle v, u\rangle & \langle v, v\rangle \end{pmatrix}^{\frac{1}{2}} \ .$$

7.3. Dualräume und adjungierte Abbildungen

Wir erinnern uns an die Definition 2.43 des Dualraumes $V^* = \operatorname{Hom}_{\Bbbk}(V, \Bbbk)$ eine Links-$\Bbbk$-Vektorraums V. Der Dualraum eines Links-Vektorraums ist ein

Rechtsvektorraum und umgekehrt (wobei wir anstelle von ${}^*V$ oft einfach wieder V^* schreiben). Elemente des Dualraumes heißen auch *Linearformen* $\alpha \in V^*$. Linearformen auf V sind also lineare Abbildungen $\alpha \colon V \to \mathbb{k}$.

7.21. Beispiel. Wir kennen Beispiele von Linearformen.

(1) Für $p = 1, \dots, n$ ist die Abbildung $\varepsilon_p \colon \mathbb{k}^n \to \mathbb{k}$ mit $\varepsilon_p(x) = x^p$ eine Linearform. Für die Standardbasisvektoren $e_1, \dots, e_n$ gilt

$$\varepsilon_p(e_q) = \delta_{pq} \;,$$

und man nennt $(\varepsilon_1, \dots, \varepsilon_n)$ die zu $(e_1, \dots, e_n)$ duale Basis von ${}^n\mathbb{k}$, siehe Bemerkung 2.72.

Allgemeiner sei V ein Rechts-$\mathbb{k}$-Vektorraum und $B = (b_1, \dots, b_n)$ eine Basis. Die dazu duale Basis $(\beta_1, \dots, \beta_n)$ mit $\beta_p = \varepsilon_p \circ B^{-1} \colon V \to \mathbb{k}$ haben wir in Proposition 2.79 konstruiert, so dass wieder

$$\beta_p(b_q) = \delta_{pq} \;.$$

(2) Wir betrachten den Raum $C^\infty([0,1], \mathbb{k})$ der beliebig oft differenzierbaren, $\mathbb{k}$-wertigen Funktionen auf dem Intervall $[0,1]$. Typische Linearformen auf $C^\infty([0,1], \mathbb{k})$ sind zum Beispiel

$$f \longmapsto f(x_0) \;, \qquad f \longmapsto f'(x_0) \;, \dots$$

für $x_0 \in [0,1]$, sowie Linearkombinationen solcher Linearformen.

(3) Auch die Abbildung

$$f \longmapsto \int_0^1 f(x)\, dx$$

ist linear. Allgemeiner betrachten wir das L^2-Skalarprodukt aus Beispiel 7.7 (1). Die obige Linearform entspricht der Abbildung

$$\langle 1, \cdot\,\rangle_{L^2} \colon C^\infty([0,1], \mathbb{k}) \longrightarrow \mathbb{k} \quad \text{mit} \quad f \longmapsto \langle 1, f\rangle_{L^2} = \int_0^1 \bar{1} \cdot f(x)\, dx \;,$$

wobei 1 die konstante Abbildung $x \mapsto 1$ bezeichne. Wenn wir das L^2-Skalarprodukt mit der gleichen Definition auf beschränkte und stückweise stetige Funktionen erweitern, existiert für ein beschränktes und stückweise stetiges $g \colon [0,1] \to \mathbb{k}$ die Abbildung

$$f \longmapsto \langle f, g\rangle_{L^2} = \int_0^1 \overline{g(x)} \cdot f(x)\, dx \;.$$

Einschränken auf $f \in C^\infty([0,1], \mathbb{k})$ liefert wieder eine Linearform auf $C^\infty([0,1], \mathbb{k})$.

(4) Wenn g stetig und differenzierbar ist, dann erhalten wir entsprechend mit der Erweiterung des Sobolev-Skalarproduktes aus Beispiel 7.7 (3) auf C^1-Funktionen eine Linearform

$$g \longmapsto \langle f, g\rangle_{H^1} = \langle f, g\rangle_{L^2} + \langle f', g'\rangle_{L^2} \;.$$

(5) Es sei $U \subset \mathbb{R}^n$ offen und $f\colon U \to \mathbb{R}$ eine C^1-Funktion (total differenzierbar würde auch ausreichen). Dann ist die *totale Ableitung* an der Stelle $x_0 \in U$ die lineare Abbildung $df(x_0)\colon \mathbb{R}^n \to \mathbb{R}$, die jedem Vektor $v \in \mathbb{R}^n$ die *Richtungsableitung*
$$df(x_0)(v) = \lim_{t\to 0} \frac{f(x_0 + v\,.\,t) - f(x_0)}{t}$$
zuordnet. Die Koordinaten von $df(x_0)$ bezüglich der dualen Basis ε_1, $\dots$, ε_n von ${}^n\mathbb{R} = (\mathbb{R}^n)^*$ heißen auch die *partiellen Ableitungen* von f.

Wir erinnern uns auch an anti- (oder semi-) lineare Abbildungen, siehe Definition 7.2.

7.22. Proposition. *Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Dann induziert g eine injektive antilineare Abbildung $g\colon V \to V^*$ durch*
$$v \longmapsto g(v) \in V^* \qquad \textit{mit} \qquad g(v)(w) = g(v, w) \qquad \textit{für alle } v,\, w \in V\ .$$

Dann heißt eine Linearform $\alpha \in V^*$ *darstellbar* bezüglich g, wenn es einen Vektor $v \in V$ mit $\alpha = g(v)$ gibt. Man sagt auch, dass $v \in V$ die Linearform α *darstellt.*

Beweis. Es sei $v \in V$, dann ist nach (S1) in Definition 7.3 die Abbildung
$$g(v) = g(v, \,\cdot\,)\colon V \to \Bbbk$$
linear, also gilt $g(v) \in V^*$. Nach Definition 2.43 ist V^* ein Links-$\Bbbk$-Vektorraum. Aus (S1) folgt, dass $g\colon V \to V^*$ antilinear ist, denn für u, v, $w \in V$ und r, $s \in \Bbbk$ gilt
$$g(u\,.\,r+v\,.\,s)(w) = g(u\,.\,r+v\,.\,s, w) = \bar r\, g(u, w)+\bar s\, g(v, w) = \big(\bar r\, g(u)+\bar s\, g(v)\big)(w)\ .$$
Zur Injektivität nehmen wir an, dass $g(u) = g(v)$, das heißt, es gilt $g(u)(w) = g(v)(w)$ für alle $w \in V$. Dann folgt
$$0 = g(u)(u - v) - g(v)(u - v) = \|u - v\|_g^2\ ,$$
also gilt $u = v$ wegen (S3), und $g\colon V \to V^*$ ist injektiv. □

7.23. Beispiel. Wir wollen wissen, ob die Linearformen aus dem obigen Beispiel darstellbar sind.

(1) Es sei $(e_1, \dots, e_n)$ eine Orthonormalbasis, dann wird ε_p im Beispiel 7.21 (1) durch den Vektor e_p dargestellt, siehe Proposition 2.82 und Bemerkung 7.17 (3).
(2) Im Beispiel 7.21 (3) wird die Linearform
$$f \longmapsto \int_0^1 \overline{g(x)} \cdot f(x)\, dx$$
auf den stückweise stetigen Funktionen durch g dargestellt. Auf dem Unterraum $C^\infty([0,1], \Bbbk)$ wird sie nur genau dann durch g dargestellt, wenn $g \in C^\infty([0,1], \Bbbk)$. Also gibt es viele Linearformen auf $C^\infty([0,1], \Bbbk)$, die bezüglich des L^2-Skalarproduktes nicht (das

heißt, nicht durch C^∞-Funktionen) darstellbar sind. Auch die Linearformen aus 7.21 (2) sind nicht durch C^∞-Funktionen darstellbar.

(3) Wie oben ist die Linearform $\langle g,\,\cdot\,\rangle_{H^1}$ aus 7.21 (4) auf C^1 bezüglich des ersten Sobolev-Skalarproduktes durch g darstellbar. Auf $C^\infty([0,1],\Bbbk)$ ist sie genau dann darstellbar, wenn $g \in C^\infty([0,1],\Bbbk)$. Wenn darüberhinaus $g'(0) = g'(1) = 0$ gilt, ist $\langle g,\,\cdot\,\rangle_{H^1}$ sogar bezüglich des L^2-Skalarproduktes darstellbar, denn partielle Integration liefert

$$\begin{aligned}\langle g, f\rangle_{H^1} &= \int_0^1 \overline{g(x)}\, f(x)\, dx + \int_0^1 \overline{g'(x)}\, f'(x)\, dx \\ &= \int_0^1 \overline{g(x)}\, f(x)\, dx + \overline{g'(x)}\, f(x)\Big|_{x=0}^1 - \int_0^1 \overline{g''(x)}\, f(x)\, dx \\ &= \langle g - g'', f\rangle_{L^2} \;.\end{aligned}$$

(4) Die totale Ableitung $df(x_0) \in {}^n\mathbb{R} = (\mathbb{R}^n)^*$ aus Beispiel 7.21 (5) wird bezüglich des Standardskalarproduktes auf $\mathbb{R}^n$ dargestellt durch den *Gradienten*

$$\operatorname{grad} f(x_0) = \big(df(x_0)\big)* = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{pmatrix} \;.$$

Um zu sehen, dass alle Linearformen auf einem endlich-dimensionalen Vektorraum mit Skalarprodukt darstellbar sind, führen wir als Hilfsmittel noch einen weiteren Vektorraum ein.

7.24. Definition. Es sei V ein Rechts-$\Bbbk$-Vektorraum, Dann ist eine *Antilinearform* eine antilineare Abbildung $\gamma\colon V \to \Bbbk$. Wir definieren wir den *Antidualraum* von V als

$$\bar V^* = \big\{\, \gamma\colon V \to \Bbbk \;\big|\; \gamma \text{ ist } \Bbbk\text{-antilinear} \,\big\} \;.$$

Analog definieren wir den Antidualraum eines Links-$\Bbbk$-Vektorraums.

7.25. Bemerkung. Wir sammeln einige einfache Eigenschaften.

(1) Der Antidualraum $\overline{V}^*$ eines Rechts-$\Bbbk$-Vektorraums V ist wieder ein Rechts-$\Bbbk$-Vektorraum. Sei $\gamma\colon V \to \Bbbk$ antilinear, und seien $v \in V$ und r, $s \in \Bbbk$, dann definieren wir

$$(\gamma \,.\, r)(v) = \gamma(v) \cdot r \in \Bbbk \;.$$

mit Definition 7.2 erhalten wir

$$(\gamma \,.\, r)(v \,.\, s) = \gamma(v \,.\, s) \,.\, r = \bar s \cdot \gamma(v) \,.\, r = \bar s \cdot (\gamma \,.\, r)(v) \;.$$

(2) Wir können aus jeder Linearform $\alpha \in V^*$ eine Antilinearform $\gamma = \bar\alpha \in \overline{V}^*$ machen und umgekehrt, wobei

$$\bar\alpha(v) = \overline{\alpha(v)} \in \Bbbk \;.$$

Das liefert eine antilineare Abbildung $V^* \to \overline{V}^*$ mit einer antilinearen Umkehrabbildung. Ähnlich wie in Bemerkung 7.4 geht dabei die Links-$\Bbbk$-Vektorraumstruktur von V^* in die Rechts-$\Bbbk$-Vektorraumstruktur von $\overline{V}^*$ über und umgekehrt.

(3) Wir definieren eine Abbildung $\bar{g}\colon V \to \overline{V}^*$ durch

$$\bar{g}(v)(w) = \overline{g(v)}(w) = \overline{g(v,w)} = g(w,v)$$

für alle v, $w \in V$. Aus Proposition 7.22 folgt, dass $\bar{g}$ eine injektive lineare Abbildung ist. Antilinearformen im Bild von $\bar{g}$ heißen wieder *darstellbar.*

Die Frage, welche Linearformen sich durch Elemente spezieller Funktionenräume darstellen lassen, ist ein wichtiges Thema in der Funktionalanalysis. Das folgende Lemma ist ein elementarer Spezialfall des Rieszschen Darstellungssatzes.

7.26. Lemma. *Es sei (V,g) ein endlich-dimensionaler $\Bbbk$-Vektorraum, dann sind die Abbildungen $g\colon V \to V^*$ und $\bar{g}\colon V \to V^*$ bijektiv, und wir erhalten Umkehrabbildungen $g^{-1}\colon V^* \to V$ und $\bar{g}^{-1}\colon \overline{V}^* \to V$.*

Insbesondere ist jede Linearform und jede Antilinearform auf einem endlich-dimensionalen Vektorraum V bezüglich g darstellbar. Dieses Lemma erklärt also insbesondere die Beispiele 7.23 (1) und (4).

Der Dualraum eines unendlich-dimensionalen $\Bbbk$-Vektorraums ist (unter geeigneten mengentheoretischen Annahmen) stets mächtiger als der Vektorraum selbst, so dass g für unendlich-dimensionale Vektorräume nie invertierbar ist. Aus diesem Grund betrachtet man in der Funktionalanalysis stattdessen den Raum der stetigen Linearformen bezüglich der zum Skalarprodukt gehörigen Norm. Dadurch wird g auch für zahlreiche wichtige unendlich-dimensionale Vektorräume mit Skalarprodukt invertierbar.

Beweis. Es sei $\dim V = n$. Nach Proposition 2.79 ist V^* ein n-dimensionaler Links $\Bbbk$-Vektorraum. Nach Bemerkung 7.25 (2) ist $\overline{V}^*$ ein n-dimensionaler Rechts-$\Bbbk$-Vektorraum. Die Abbildung $\bar{g}\colon V \to \overline{V}^*$ ist nach Proposition 7.22 und Bemerkung 7.25 (3) injektiv. Aus dem Rangsatz 3.13 folgt, dass $\bar{g}$ ein Isomorphismus ist. Aber dann ist auch g bijektiv. □

7.27. Bemerkung. Es sei $(e_1, \dots, e_n)$ eine Orthonormalbasis von V, siehe Satz 7.18. Wir können sie benutzen, um einen alternativen, konstruktiven Beweis von Lemma 7.26 zu geben. Nach Beispiel 7.23 (1) werden die Vektoren der dualen Basis $(\varepsilon_1, \dots, \varepsilon_n)$ durch die Vektoren $e_1, \dots, e_n$ dargestellt. Sei also

$$\alpha = \sum_{p=1}^{n} a_p \cdot \varepsilon_p \qquad \text{mit} \qquad a \in {}^n\Bbbk\,,$$

dann wird α dargestellt durch den Vektor

$$v = \sum_{p=1}^{n} e_p \,.\, \bar{a}_p\,,$$

denn für alle $w \in V$ gilt

$$\alpha(w) = \sum_{p=1}^{n} a_p \cdot \varepsilon_p(w) = \sum_{p=1}^{n} a_p \cdot g(e_p, w) = g\Bigl(\sum_{p=1}^{n} e_p \,.\, \bar a_p, w\Bigr) \,.$$

Also gilt $\alpha = g(v)$ und analog $\bar\alpha = \bar g(v)$.

7.28. Definition. Es seien (V, g) und (W, h) Vektorräume über $\Bbbk$ mit Skalarprodukt. Eine lineare Abbildung $F\colon V \to W$ heißt *adjungierbar,* wenn es eine Abbildung $G\colon W \to V$ gibt, so dass

$$g\bigl(G(w), v\bigr) = h\bigl(w, F(v)\bigr) \qquad \text{für alle } v \in V \text{ und } w \in W \,.$$

In diesem Fall heißt G die zu F *adjungierte Abbildung,* und wir schreiben $G = F^*$.

7.29. Bemerkung. Es seien (V, g), (W, h) und $F\colon V \to W$ wie oben.

(1) Falls F adjungierbar ist, ist die adjungierte Abbildung G von F eindeutig bestimmt. Denn sei H eine weitere adjungierte Abbildung von F, dann gilt für alle $w \in W$, dass

$$\begin{aligned} 0 &= h\bigl(w, F(G(w) - H(w))\bigr) - h\bigl(w, F(G(w) - H(w))\bigr) \\ &= g\bigl(G(w), G(w) - H(w)\bigr) - g\bigl(H(w), G(w) - H(w)\bigr) \\ &= \|G(w) - H(w)\|_g^2 \ , \end{aligned}$$

und somit $G(w) = H(w)$ wegen der Eigenschaft (S3) des Skalarproduktes g. Daher dürfen wir F^* für die Adjungierte Abbildung schreiben, wenn Sie existiert.

(2) Die adjungierte Abbildung $F^*\colon W \to V$ ist linear, denn für alle $v \in V$ und alle u, $w \in W$ und alle r, $s \in \Bbbk$ gilt

$$\begin{aligned} g\bigl(F^*(u.r{+}w.s), v\bigr) &= h\bigl(u.r{+}w.s, F(v)\bigr) = \bar r\, h\bigl(u, F(v)\bigr) + \bar s\, h\bigl(w, F(v)\bigr) \\ &= \bar r\, g\bigl(F^*(u), v\bigr) + \bar s\, g\bigl(F^*(w), v\bigr) = g\bigl(F^*(u) \,.\, r + F^*(w) \,.\, s, v\bigr) \,. \end{aligned}$$

Indem wir $v = F^*(u \,.\, r + w \,.\, s) - F^*(u) \,.\, r - F^*(w) \,.\, s$ wählen, erhalten wir

$$0 = \|F^*(u \,.\, r + w \,.\, s) - F^*(u) \,.\, r - F^*(w) \,.\, s\|_g^2 \ ,$$

und wegen (S3) gilt somit $F^*(u \,.\, r + w \,.\, s) = F^*(u) \,.\, r + F^*(w) \,.\, s$.

(3) Wenn G zu F adjungiert ist, ist auch F zu G adjungiert, denn wegen (S2) gilt

$$h\bigl(F(v), w\bigr) = \overline{h\bigl(w, F(v)\bigr)} = \overline{g\bigl(G(w), v\bigr)} = g\bigl(v, G(w)\bigr)$$

für alle $v \in V$ und $w \in W$. Wegen (1) gilt also $F = G^*$ genau dann, wenn $G = F^*$, insbesondere folgt $(F^*)^* = F$.

7.30. Beispiel. Wir geben Beispiele adjungierter Abbildungen.

(1) Wir betrachten das Standardskalarprodukt auf den Räumen $\Bbbk^m$ und $\Bbbk^n$. Sei $F\colon \Bbbk^n \to \Bbbk^m$ gegeben durch eine Matrix $C \in M_{m,n}(\Bbbk)$,

dann wird die adjungierte Abbildung F^* gegeben durch die adjungierte Matrix, denn für alle $x \in \Bbbk^m$ und alle $y \in \Bbbk^n$ gilt

$$\langle F^*(x), y\rangle = \langle x, F(y)\rangle = x^*Ay = x^*(A^*)^*y = (A^*x)^*y = \langle A^*x, y\rangle \;.$$

Aus diesem Grund benutzen wir in beiden Fällen den Begriff „adjungiert“.

(2) Etwas allgemeiner sei $(e_1, \dots, e_n)$ eine Orthonormalbasis von (V, g) und $(f_1, \dots, f_m)$ eine Orthonormalbasis von (W, h). Wenn $F\colon V \to W$ bezüglich dieser Basen durch eine Matrix $A \in M_{m,n}(\Bbbk)$ dargestellt wird, dann wird F^* durch A^* dargestellt, denn für alle $p = 1, \dots, m$ und alle $q = 1, \dots, n$ gilt

$$\langle e_p, F^*(f_p)\rangle = \langle F(e_p), f_q\rangle = \overline{\langle f_q, F(e_p)\rangle} = \bar a_{pq} \;.$$

(3) Wir betrachten wieder den Raum $V = C^\infty([0,1];\Bbbk)$ mit dem L^2-Skalarprodukt. Multiplikation mit einer Funktion $f \in V$ definiert eine lineare Abbildung Die adjungierte Abbildung ist Multiplikation mit $\bar f$, denn

$$\langle g, fh\rangle_{L^2} = \int_0^1 \overline{g(x)}\, f(x)\, h(x)\, dx = \int_0^1 \overline{\overline{f(x)}\, g(x)}\, h(x)\, dx = \langle \bar f g, h\rangle_{L^2} \;.$$

(4) Es sei V wie oben, und es sei $f \in V$ eine Funktion mit $f(0) = f(1) = 0$. Dann betrachten wir den Differentialoperator $F \in \operatorname{End}(V)$ mit

$$F(g) = f \cdot g'$$

und bestimmen den adjungierten Differentialoperator durch partielle Integration als

$$\begin{aligned}
\langle F^*(g), h\rangle_{L^2} = \langle g, F(h)\rangle &= \int_0^1 \overline{g(x)}\, f(x)\, h'(x)\, dx \\
&= \Bigl(\overline{g(x)}\, f(x)\, h(x)\Bigr)\Big|_{x=0}^1 - \int_0^1 \bigl(\bar g\, f\bigr)'(x)\, h(x)\, dx \\
&= \bigl\langle (\bar f\, g)', h\bigr\rangle_{L^2} = \bigl\langle \bar f\, g' + \bar f'\, g, h\bigr\rangle_{L^2} \;.
\end{aligned}$$

(5) Wenn wir in (4) einfach nur den Ableitungsoperator $F(g) = g'$ betrachten, zeigt eine analoge Rechnung, dass F nicht adjungierbar ist, da sich die Randterme $\bigl(\overline{g(x)}\, h(x)\bigr)\big|_{x=0}^1$ nicht durch ein Integral beschreiben lassen. In der Analysis umgeht man dieses Problem, indem man den Begriff des adjungierten Operators etwas anders definiert und dann Randbedingungen stellt wie etwa $g(0) = g(1) = 0$, um keine Randterme mehr zu erhalten.

Die adjungierte Abbildung ist eng verwandt mit dem folgenden Konzept.

7.31. Definition. Es seien V und W Vektorräume über einem Körper $\Bbbk$, und es sei $F\colon V \to W$ eine lineare Abbildung. Die zu F *duale Abbildung* $F^*\colon W^* \to V^*$ ist definiert durch

$$F^*\beta = \beta \circ F \in V^* \qquad \text{für alle } \beta \in W^* \;.$$

Man beachte, dass die duale Abbildung im Gegensatz zur adjungierten Abbildung immer existiert und nach Definition eindeutig bestimmt ist. Wir verwenden für beide die Bezeichnung F^*, man muss also aufpassen, welche der beiden Abbildungen jeweils gemeint ist: $F^*\colon W^* \to V^*$ ist die duale Abbildung, $F^*\colon W \to V$ die adjungierte. Aus diesem Grund verwenden manche Autoren für duale Moduln, Vektorräume und Abbildungen das Symbol $\cdot'$ oder $\cdot^\vee$.

7.32. Bemerkung. Wir sammeln ein paar elementare Eigenschaften. Seien dazu U, V und W Vektorräume.

(1) Es gilt stets $\mathrm{id}_V^* = \mathrm{id}_{V^*}$, denn $\alpha \circ \mathrm{id}_V = \alpha \in V^*$ für alle $\alpha \in V^*$.
(2) Seien $F\colon V \to W$ und $G\colon U \to V$ linear, dann gilt $(F \circ G)^* = G^* \circ F^*$, denn

$$(F \circ G)^*\beta = \beta \circ F \circ G = G^*(\beta \circ F) = G^*\big(F^*(\beta)\big) .$$

(3) Die duale Abbildung ist linear. Das lässt sich nachrechnen, da $\Bbbk$ auf $\beta \in W^*$ durch $(r \cdot \beta)(w) = r \cdot \beta(w)$ wirkt.
(4) Es seien $B = (v_1, \dots, v_n)$ und $C = (w_1, \dots, w_m)$ Basen von V beziehungsweise W, und es seien $B^* = (\varphi_1, \dots, \varphi_n)$ und $C^* = (\psi_1, \dots, \psi_m)$ die dualen Basen von V^* und W^*. Es sei $F\colon V \to W$ bezüglich der obigen Basen dargestellt durch die Abbildungsmatrix $A = M_{m,n}(\Bbbk)$, dann gilt

$$\psi_p\big(F(v_q)\big) = \psi_p\left(\sum_{r=1}^{m} w_r \,.\, a_{rq}\right) = \sum_{r=1}^{m} \psi_p(w_r) \,.\, a_{rq} = a_{pq}$$

für alle $p = 1, \dots, m$ und alle $q = 1, \dots, n$. Dabei geht der Vektor $v = B(x)$ in den Vektor $C(Ax)$ über, wobei $x \in \Bbbk^n$.

Es sei jetzt $\eta \in {}^m\Bbbk$ eine Zeile. Für die duale Matrix rechnen wir

$$\begin{aligned} F^*\big(C^*(\eta)\big)(v_r) &= F^*\left(\sum_{p=1}^{m} \eta_p \cdot \psi_p\right)(v_r) = \left(\sum_{p=1}^{m} \eta_p \cdot (\psi_p \circ F)\right)(v_r) \\ &= \sum_{p=1}^{m} \eta_p \cdot a_{pr} = \left(\sum_{p=1}^{m} \sum_{q=1}^{n} \eta_p \cdot a_{pq} \cdot \varphi_q\right)(v_r) = \big(B^*(\eta A)\big)(v_r) . \end{aligned}$$

Die duale Abbildung wird durch also dieselbe Matrix A dargestellt, allerdings werden jetzt Zeilen in $\Bbbk^m$ von rechts mit A multipliziert.

Mit Hilfe von Lemma 7.26 können wir einen Zusammenhang zwischen der adjungierten Abbildung und der dualen Abbildung herstellen.

7.33. Proposition. *Es seien (V, g) und (W, h) Vektorräume über $\Bbbk$ mit Skalarprodukt und $F\colon V \to W$ sei linear. Wenn F adjungierbar ist, kommutiert das Diagramm*

$$\begin{tikzcd} W \arrow[r, "F^*"] \arrow[d, "h"'] & V \arrow[d, "g"] \\ W^* \arrow[r, "F^*"] & V^* . \end{tikzcd}$$

Insbesondere ist F immer adjungierbar, wenn V endlich-dimensional ist.

Man beachte, dass die beiden waagerechten Pfeile lineare Abbildungen sind, während die senkrechten Pfeile antilinear sind. Somit sind beide Wege von W nach V^* durch antilineare Abbildungen gegeben.

Die letzte Behauptung erklärt insbesondere die Beispiele 7.30 (1) und (2). Die Beispiele 7.30 (3) und (4) zeigen, dass die zusätzliche Bedingung $\dim V < \infty$ nicht notwendig ist.

Beweis. Es seien $v \in V$ und $w \in W$, dann folgt

$$g\big(F^*(w)\big)(v) = g\big(F^*(w), v\big) = h\big(w, F(v)\big) = h(w)\big(F(v)\big) = F^*\big(h(w)\big)(v) .$$

Da das für alle $v \in V$ gilt, folgt $g \circ F^* = F^* \circ h$, wobei links die adjungierte und rechts die duale Abbildung gemeint ist. Damit ist die erste Behauptung bewiesen.

Wenn g invertierbar ist, können wir die adjungierte Abbildung als $g^{-1} \circ F^* \circ h$ schreiben. Nach Lemma 7.26 gilt das, wenn V endlich-dimensional ist. $\square$

7.34. Definition. Es seien V und W Vektorräume über $\Bbbk$, es sei $F \colon V \to W$ linear, und S sei eine Sesquilinearform auf W. Dann definiert man die mit F *zurückgeholte Sesquilinearform* F^*S auf V durch

$$(F^*S)(u, v) = S\big(F(v), F(w)\big) \qquad \text{für alle } u,\, v \in V .$$

Wir haben jetzt die Notation F^* mit einer weiteren Bedeutung versehen. Aus dem Zusammenhang muss man jeweils erkennen, wofür F^* gerade steht.

7.35. Bemerkung. Die ersten drei der folgenden Eigenschaften rechnet man leicht nach.

(1) Die Form F^*S ist wieder sesquilinear (S1).
(2) Wenn S Hermitesch ist, dann ist auch F^*S Hermitesch (S2).
(3) Wenn S außerdem positiv semidefinit ist, dann ist auch F^* positiv semidefinit.
(4) Wenn S positiv definit (S3) und F injektiv ist, dann ist auch F^*S ein Skalarprodukt, denn dann gilt

$$(F^*S)(v, v) = S\big(F(v), F(v)\big) = 0 \quad\iff\quad F(v) = 0 \quad\iff\quad v = 0 .$$

In diesem Fall heißt F^*S auch das zurückgeholte Skalarprodukt. Auf die Injektivität von F kann man leider nicht verzichten, denn für alle $v \in \ker F$ gilt $(F^*S)(v) = 0$.
(5) Es seien $(v_1, \dots, v_n)$ und $(w_1, \dots, w_m)$ Basen von V und W. Sei $A \in M_{M,n}(\Bbbk)$ die Abbildungsmatrix von F, und sei S dargestellt durch die Gramsche Matrix G, dann wird F^*S dargestellt durch die Matrix A^*GA, denn

$$(F^*S)(v_p, v_q) = S\bigg(\sum_{r=1}^m w_r \,.\, a_{rp}, \sum_{s=1}^m w_s \,.\, a_{sq}\bigg) = \sum_{r,s=1}^m \bar a_{rp}\, g_{rs}\, a_{sq} .$$

7.36. Bemerkung. Genauso können wir eine Sesquilinearform S auf W mit einer antilinearen Abbildung $F\colon V \to W$ zurückholen durch

$$(F^*S)(u,v) = S\bigl(F(v), F(u)\bigr) \qquad \text{für alle } u,\, v \in V\ .$$

Durch das Vertauschen der Argumente stellen wir sicher, dass F^*S wieder sesquilinear ist. Die Punkte (2)–(4) aus Bemerkung 7.35 gelten analog.

Zum Beispiel sei (V,g) ein endlich-dimensionaler Vektorraum mit Skalarprodukt, dann ist die antilineare Abbildung $g\colon V \to V^*$ invertierbar, und wir können das Skalarprodukt g mit der Inversen Abbildung g^{-1} auf V^* zurückholen. Dieses Skalarprodukt nennen wir das zu g *duale Skalarprodukt* g^* auf V^*. Sei dazu $(e_1, \dots, e_n)$ eine Orthonormalbasis von V, dann ist $(\varepsilon_1, \dots, \varepsilon_n) = \bigl(g(e_1), \dots, g(e_n)\bigr)$ die duale Basis von V^* nach Beispiel 7.23 (1). Somit gilt

$$\bigl((g^{-1})^*g\bigr)(\varepsilon_p, \varepsilon_q) = g\bigl(g^{-1}(\varepsilon_q), g^{-1}(\varepsilon_p)\bigr) = g(e_q, e_p) = \delta_{pq}\ ,$$

also ist die duale Basis einer Orthonormalbasis wieder eine Orthonormalbasis. Im Allgemeinen sei A die Gramsche Matrix von g, dann kann man zeigen, dass g^* durch die Inverse Matrix A^{-1} dargestellt wird.

7.37. Bemerkung. Es seien (V,g) und (W,h) Vektorräume über $\Bbbk$ mit Skalarprodukt. Wir nennen eine lineare Abbildung $F\colon V \to W$ eine *isometrische Einbettung,* wenn $F^*h = g$ gilt. In diesem Fall gilt also für alle $u,\, v \in V$, dass

$$g(u,v) = (F^*h)(u,v) = h\bigl(F(u), F(v)\bigr) \qquad \text{und} \qquad \|v\|_g = \|F(v)\|_g\ .$$

Insbesondere ist F immer injektiv.

Seien $(v_1, \dots, v_n)$ und $(w_1, \dots, w_m)$ Orthonormalbasen von V und W, und sei $A \in M_{m,n}(\Bbbk)$ die Abbildungsmatrix von F. Wegen Bemerkung 7.35 (5) gilt dann

$$E_n = A^*E_mA = A^*A\ .$$

Falls $n = m$ ist, ist A insbesondere invertierbar, und es gilt $A^{-1} = A^*$. In diesem Fall nennen wir F eine *lineare Isometrie.*

7.4. Normale Endomorphismen

In diesem Kapitel betrachten wir bestimmte Endomorphismen von endlichdimensionalen $\Bbbk$-Vektorräumen mit Skalarprodukt und zeigen, dass sie sich bezüglich einer geeigneten Orthonormalbasis durch besonders einfache Matrizen darstellen lassen. Diese Resultate haben zahlreiche Anwendungen, unter anderem in Analysis, Geometrie und Physik.

7.38. Definition. Es sei (V,g) ein $\Bbbk$-Vektorraum mit Skalarprodukt und es sei $F \in \operatorname{End}_{\Bbbk}(V)$ adjungierbar. Dann heißt F

(1) *selbstadjungiert,* wenn $F^* = F$,
(2) *schief,* wenn $F^* = -F$,
(3) *normal,* wenn $F^* \circ F = F \circ F^*$, und
(4) *lineare Isometrie,* oder *isometrischer* oder auch *unitärer Automorphismus,* wenn F invertierbar ist mit $F^{-1} = F^*$.

7.39. Bemerkung. Man sieht leicht, dass selbstadjungierte und schiefe Endomorphismen und isometrische Automorphismen allesamt normal sind. Stellt man F wie oben bezüglich einer Orthonormalbasis als Matrix $A \in M_n(\mathbb{k})$ dar, so gilt jeweils $A^* = A$, $A^* = -A$, $A^*A = AA^*$, beziehungsweise $A^{-1} = A^*$.

7.40. Satz (Hauptsatz über normale Abbildungen)**.** *Es sei (V, g) ein endlichdimensionaler komplexer Vektorraum mit Skalarprodukt und $F \in \operatorname{End}_{\mathbb{C}} V$. Dann existiert genau dann eine unitäre Basis von (V, g) aus Eigenvektoren von F, wenn F normal ist.*

Insbesondere sind normale Endomorphismen über $\mathbb{C}$ immer diagonalisierbar; die Aussage im Satz ist aber noch etwas stärker, da wir sogar eine unitäre (also eine Orthonormal-) Basis erhalten. Die Darstellung als Diagonalmatrix ist eindeutig bis auf die Reihenfolge der Einträge nach Satz 5.34. In der Funktionalanalysis heißt der entsprechende Satz auch der „Spektralsatz für normale Operatoren“.

Beweis. Zu „$\Longrightarrow$“ sei $(e_1, \dots, e_n)$ eine Orthonormalbasis aus Eigenvektoren von F. Nach Proposition 5.4 wird F bezüglich dieser Basis durch eine Diagonalmatrix A dargestellt. Nach Beispiel 7.30 (2) wird F^* durch A^* dargestellt, und A^* ist auch eine Diagonalmatrix. Man sieht leicht, dass $A^*A = AA^*$ gilt, somit ist F normal.

Wir beweisen „$\Longleftarrow$“ durch Induktion über die Dimension n von V. Im Fall $n = 0$ ist nichts zu zeigen. Es sei also $n \geq 1$. Da $\mathbb{C}$ algebraisch abgeschlossen ist, hat das charakteristische Polynom χ_F eine Nullstelle λ. Es sei v ein Eigenvektor von F zum Eigenwert λ. Dann ist v auch ein Eigenvektor von F^* zum Eigenwert $\bar\lambda$, denn

$$\begin{aligned}\|F^*(v) - v \,.\, \lambda\|_g^2 &= g\big((F^* - \bar\lambda \operatorname{id}_V)(v), (F^* - \bar\lambda \operatorname{id}_V)(v)\big)\\ &= g\big((F - \lambda \operatorname{id}_V)(F^* - \bar\lambda \operatorname{id}_V)(v), v\big)\\ &= g\big((F^* - \bar\lambda \operatorname{id}_V)(F - \lambda \operatorname{id}_V)(v), v\big) = 0\end{aligned}$$

Es sei

$$W = \left\{\, w \in V \;\middle|\; g(v, w) = 0 \,\right\}$$

das orthogonale Komplement vom Vektor v, dann ist $W \subset V$ ein Untervektorraum, siehe Übungen. Der Unterraum W ist sowohl unter F als auch unter F^* invariant, denn sei $w \in W$, dann folgt

$$\begin{aligned} & g\big(v, F(w)\big) = g\big(F^*(v), w\big) = g(v \,.\, \mu, w) = 0\\ \text{und}\qquad & g\big(v, F^*(w)\big) = g\big(F(v), w\big) = g(v \,.\, \lambda, w) = 0\;,\end{aligned}$$

somit liegen mit w auch $F(w)$ und $F^*(w)$ wieder in W.

Insbesondere ist $F^*|_W$ gleichzeitig die adjungierte Abbildung zu $F|_W$ bezüglich des auf W eingeschränkten Skalarproduktes, und $F|_W$ ist nach wie vor normal, also existiert nach Induktionsannahme eine unitäre Basis $v_2, \dots,$ v_n von W aus Eigenvektoren von $F|_W$. Wir dürfen $\|v\|_g = 1$ annehmen. Dann ist $(v, v_2, \dots, v_n)$ eine unitäre Basis von V aus Eigenvektoren von F, und wir sind fertig. $\square$

Über den reellen Zahlen verhalten sich normale Abbildungen etwas komplizierter. Man überprüft, dass Matrizen der Form

$$(*) \qquad \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

normal sind, denn

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \cdot \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \cdot \begin{pmatrix} a & b \\ -b & a \end{pmatrix} .$$

Das charakteristische Polynom $(X-a)^2+b^2$ hat jedoch keine reellen Nullstellen, falls $b \neq 0$. Nach Lemma 5.33 hat die obige Matrix also keine Eigenvektoren.

7.41. Folgerung. *Es sei (V, g) ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt und $F \in \operatorname{End}_{\mathbb{R}} V$. Dann existiert genau dann eine Orthonormalbasis von (V, g), bezüglich der F durch eine Block-Diagonalmatrix aus 1×1-Blöcken und aus 2×2-Blöcken der Gestalt $(*)$ mit $b > 0$ dargestellt wird, wenn F normal ist. In diesem Fall ist die Matrix bis auf die Reihenfolge der Blöcke eindeutig.*

Beweis. Die Richtung „$\Longrightarrow$“ folgt wie im Beweis von Satz 7.40 durch Nachrechnen.

Zu „$\Longleftarrow$“ wählen wir zunächst eine beliebige Orthonormalbasis von V und stellen F durch eine normale Matrix $A \in M_n(\mathbb{R})$ dar. Dann betrachten wir A als normale komplexe Matrix, also als normalen Endomorphismus von $\mathbb{C}^n$ mit dem Standardskalarprodukt. Der Beweis geht wieder durch Induktion über n.

Wir finden einen gemeinsamen Eigenvektor v von A zum Eigenwert λ und von A^* zum Eigenwert $\bar{\lambda}$. Wenn λ reell ist, können wir $v \in \mathbb{R}^n \subset \mathbb{C}^n$ wählen wegen Lemma 5.33. Danach betrachten wir das orthogonale Komplement W von v und machen weiter wie im obigen Beweis.

Wenn λ nicht reell ist, betrachten wir den Vektor $\bar{v} \in \mathbb{C}^n$ und rechnen nach, dass

$$A\,\bar{v} = \bar{A}\,\bar{v} = \overline{Av} = \overline{v \,.\, \lambda} = \bar{v} \,.\, \bar{\lambda} ,$$

so dass $\bar{v}$ ein Eigenvektor von A zum Eigenwert $\bar{\lambda} \neq \lambda$ ist. Dabei haben wir benutzt, dass $\bar{A} = A$, da A eine reelle Matrix ist. Wir schreiben $\lambda = a + bi$ mit $a, b \in \mathbb{R}$ und $v = w + ui$ mit $u, w \in \mathbb{R}^n$, und erhalten

$$\begin{aligned} A\,u &= \frac{i}{2}\,(A\bar{v} - Av) = \frac{i}{2}\,(\bar{v}\,\bar{\lambda} - v\lambda) \\ &= \frac{i}{2}\,\big((w-ui)(a-bi) - (w+ui)(a+bi)\big) = wa + ub , \\ \text{und} \qquad A\,w &= \frac{1}{2}\,(A\bar{v} + Av) = \frac{1}{2}\,(\bar{v}\,\bar{\lambda} + v\lambda) \\ &= \frac{1}{2}\,\big((w-ui)(a-bi) + (w+ui)(a+bi)\big) = wa - ub . \end{aligned}$$

Wir nehmen an, dass $b = \operatorname{Im}\lambda > 0$, andernfalls vertauschen wir die Rollen von v und $\bar{v}$. Dann hat A auf dem von u und w aufgespannten Unterraum U bezüglich der Basis (u, w) gerade die Gestalt (*).

Als nächstes überlegen wir uns, dass v und $\bar{v}$ aufeinander senkrecht stehen, da sie Eigenvektoren zu verschiedenen Eigenwerten sind, und somit wie im Beweis von Satz 7.40 der Vektor $\bar{v}$ im orthogonalen Komplement von v liegt. Wir nehmen an, dass $\|v\|_g = \|\bar{v}\|_g = 2$, dann gilt

$$\begin{aligned} 2 &= \|v\|_g^2 = g(w + ui, w + ui) = \|w\|_g^2 + \|v\|_g^2 + \big(g(w, u) - g(u, w)\big)\, i \\ 0 &= g(w - ui, w + ui) = \|w\|_g^2 - \|v\|_g^2 + 2g(w, u)\, i\,. \end{aligned}$$

Da u, w reell sind, sind auch alle einzelnen Skalarprodukte rechts reell. Hieraus folgt $g(u, w) = 0$ und $\|u\|_g^2 = \|w\|_g^2 = 1$, so dass u, w eine Orthonormalbasis von U bilden. Wie im Beweis von Satz 7.40 ist das orthogonale Komplement

$$W = \left\{\, z \in \mathbb{C}^n \;\middle|\; g(w, z) = g(u, z) = 0 \,\right\} = \left\{\, z \in \mathbb{C}^n \;\middle|\; g(v, z) = g(\bar{v}, z) = 0 \,\right\}$$

invariant unter F und F^*, und wir können den Beweis wie oben fortsetzen.

Wir erhalten also eine Blockmatrix aus 1×1-Blöcken, die genau den reellen Nullstellen von χ_F entsprechen, und aus 2×2-Blöcken der Gestalt (*), so dass $a \pm bi$ echt komplexe Nullstellen von χ_F sind. Somit können wir die gesuchte Matrix aus den komplexen Nullstellen des charakteristischen Polynoms ablesen, was die Eindeutigkeitsaussage beweist. □

Da die Quaternionen nicht kommutativ sind, ist der Begriff des Eigenraums nicht sinnvoll, siehe Übung 4 von Blatt 1. Dennoch erhalten können wir normale Abbildungen auch über den Quaternionen charakterisieren.

7.42. Folgerung. *Es sei (V, g) ein Rechts-$\mathbb{H}$-Vektorraum mit Skalarprodukt und $F \in \operatorname{End}_{\mathbb{H}} V$. Dann existiert genau dann eine quaternionisch unitäre Basis von V, bezüglich der F durch eine Diagonalmatrix mit Einträgen der Form $a+bi$ mit $b \geq 0$ dargestellt wird, wenn F normal ist. Diese Matrix ist eindeutig bis auf Reihenfolge der Einträge.*

Beweis. Die Richtung „$\Longrightarrow$“ überprüft man wieder durch Nachrechnen.

Wir beweisen „$\Longleftarrow$“ wieder durch Induktion über $n = \dim_{\mathbb{H}} V$. Dazu betrachten wir $\mathbb{C} = \mathbb{R} + i\mathbb{R} \subset \mathbb{H}$ als Teilkörper und fassen V für einen Moment als komplexen Vektorraum auf. Wir erhalten ein komplexes Skalarprodukt, indem wir die j- und k-Komponenten von g vergessen. Bezüglich dieses Skalarproduktes ist F als komplex lineare Abbildung immer noch normal mit der selben adjungierten Abbildung F^*. Also existiert wie oben ein simultaner Eigenvektor v zum Eigenwert $\lambda = a + bi \in \mathbb{C}$ von F und zum Eigenwert $\bar{\lambda} = a - bi$ von F^*. Wenn $b \geq 0$ gilt, dann ist das orthogonale Komplement

$$W = \left\{\, w \in V \;\middle|\; g(v, w) = 0 \in \mathbb{H} \,\right\}$$

ein invarianter quaternionischer Unterraum der Dimension $n-1$, und wir fahren fort wie im Beweis von Satz 7.40.

Falls $b < 0$, betrachten wir den Vektor $v \,.\, j$. Es gilt

$$F(v \,.\, j) = F(v) \,.\, j = v \,.\, \big((a+bi)\,j\big) = v \,.\, \big(j\,(a-bi)\big) = (v \,.\, j) \,.\, \bar\lambda$$

und genauso $F^*(v \,.\, j) = (v \,.\, j) \,.\, \lambda$. Anstelle von v betrachten wir also $v \,.\, j$ und machen weiter wie oben und erhalten die gesuchte quaternionisch unitäre Basis.

Zur Eindeutigkeit überlegen wir uns, dass das charakteristische Polynom χ_F von F als Endomorphismus über dem Körper $\mathbb{C}$ aufgrund der obigen Überlegung in Faktoren der Form

$$(X-\lambda)(X-\bar\lambda) = (X-a)^2 + b^2$$

zerfällt. Dadurch sind die Diagonaleinträge bis auf ihre Reihenfolge eindeutig festgelegt. □

Wir können „i“ in der Darstellung $a + bi$ auch durch j, k oder einen beliebigen anderen imaginären Einheitsquaternion q ersetzen. Dadurch ändern sich die Matrix und die zugehörige Basis, aber nicht die Paare (a, b) in $a + bq$. Im Beweis arbeiten wir dann mit einem Teilkörper $\mathbb{R} + q\mathbb{R} \cong \mathbb{C}$.

Wir kommen jetzt zu wichtigen Spezialfällen normaler Abbildungen.

7.43. Folgerung (Hauptachsentransformation)**.** *Es sei (V, g) ein endlichdimensionaler $\Bbbk$-Vektorraum mit Skalarprodukt und $F \in \operatorname{End}_{\Bbbk} V$. Dann existiert genau dann eine Orthonormalbasis von V, bezüglich der F durch eine Diagonalmatrix mit reellen Einträgen dargestellt wird, wenn F selbstadjungiert ist.*

Aus den obigen Ergebnisses folgt dann auch die Eindeutigkeit dieser Diagonalmatrix bis auf die Reihenfolge der Diagonaleinträge. In der Funktionalanalysis heißt der entsprechende Satz auch der „Spektralsatz für selbstadjungierte Operatoren“.

Beweis. Die Richtung „$\Longrightarrow$“ ergibt sich wieder durch Nachrechnen.

Zu „$\Longleftarrow$“ wenden wir Satz 7.40 oder eine der Folgerungen 7.41 oder 7.42 an. Da F selbstadjungiert ist, ist auch die Matrix $A \in M_n(\Bbbk)$, die wir so erhalten, selbstadjungiert. Im Fall $\Bbbk = \mathbb{R}$ ist ein 2×2-Block vom Typ * nur dann selbstadjungiert, wenn $b = 0$ gilt. In den Fällen $\Bbbk = \mathbb{C}$ oder $\mathbb{H}$ muss $\lambda = \bar\lambda$ für jeden Diagonaleintrag $\lambda \in \Bbbk$ gelten, und wegen Bemerkung 7.1 (4) folgt $\lambda \in \mathbb{R}$. □

7.44. Beispiel. Wir betrachten einen physikalischen Körper K im $\mathbb{R}^3$, der sich ohne Einfluss äußerer Kräfte bewegt. Dabei nehmen wir an, dass der Schwerpunkt für alle Zeiten im Nullpunkt liegt. Dann dreht sich der Körper um sich selbst.

Um diese Drehung zu beschreiben, betrachtet man zu einer festen Zeit t den Trägheitstensor $F_t \colon \mathbb{R}^3 \to \mathbb{R}^3$ mit

$$F_t(v) = \int_{K_t} \rho_t(p)\, p \times (v \times p)\, d^3x \;,$$

wobei $K_t \subset \mathbb{R}^3$ den Körper zur Zeit t und $\rho_t(x)$ seine (Massen-) im Punkt x bezeichne. Dabei bezeichne die Richtung von v die Drehachse und $\|v\|$ die Drehgeschwindigkeit, dann beschreibt $v \times p$ die tatsächliche Geschwindigkeit im Punkt p, und $\rho_t(p)\, p \times (v \times p)$ den Beitrag zum Drehimpuls. Insgesamt ist $F_t(v_t)$ also der Drehimpuls zur Zeit t, wenn v_t wie oben die Drehung zur Zeit t beschreibt.

Die Abbildung F_t ist selbsadjungiert. Am einfachsten überlegt man sich das für den Integranden: für alle $w \in \mathbb{R}^3$ gilt

$$\langle p \times (v \times p), w\rangle = \langle p, p\rangle\langle v, w\rangle - \langle p, v\rangle\langle p, w\rangle = \langle v, p \times (w \times p)\rangle$$

nach Satz 1.68 (2). Also existiert eine Orthonormalbasis $\bigl(e_1(t), e_2(t), e_3(t)\bigr)$ von $\mathbb{R}^3$ aus Eigenvektoren von F_t. Wegen der Cauchy-Schwarz-Ungleichung 7.10 sind alle Eigenwerte $\lambda_1 \le \lambda_2 \le \lambda_3$ positiv, falls der Körper sich in jeder Raumrichtung ausdehnt.

Die Richtungen der Eigenvektoren heißen die *Hauptachsen* des Körpers K. Bei einem achsenparallelen Quader sind das beispielsweise gerade die Koordinatenachsen. Wenn sich der Körper zu einer festen Zeit um eine der Hauptachsen dreht, dann tut er das für alle Zeit. Dreht er sich hingegen um eine andere Achse, dann verändert sich die Drehachse selbst im Laufe der Zeit; der Körper scheint zu taumeln. Es gibt jedoch einen konstanten Drehimpulsvektor $L = F_t(v_t) \in \mathbb{R}^3$, so dass die Drehachse zu jedem Zeitpunkt in Richtung $F_t^{-1}(L)$ zeigt.

7.45. Folgerung. *Es sei (V, g) ein endlich-dimensionaler $\Bbbk$-Vektorraum, und S sei eine Sesquilinearform auf V. Dann existiert genau dann eine g-Orthonormalbasis von V, bezüglich der S durch eine Diagonalmatrix mit reellen Einträgen dargestellt wird, wenn S Hermitesch ist.*

Wie in Folgerung 7.43 sind die Diagonaleinträge bis auf ihre Reihenfolge eindeutig.

Beweis. Die Richtung „$\Longrightarrow$“ ergibt sich aus Bemerkung 7.15 (2).

Zu „$\Longleftarrow$“ fassen wir zunächst S als antilineare Abbildung $S\colon V \to V^*$ wie in Proposition 7.22 auf. Da wir nichts über die Definitheit von S wissen, können wir allerdings nicht schließen, dass S injektiv ist. Sei $g^{-1}\colon V^* \to V$ die antilineare Umkehrabbildung aus Lemma 7.26, dann setzen wir $F = g^{-1} \circ S \in \operatorname{End}_\Bbbk V$, so dass

$$S(v, w) = (g \circ F)(v)(w) = g\bigl(F(v), w\bigr) \qquad \text{für alle } v,\, w \in V\ .$$

Da S Hermitesch ist, ist F selbstadjungiert, und Folgerung 7.43 liefert eine Orthonormalbasis $(e_1, \dots, e_n)$ aus Eigenwerten von F. Man überlegt sich leicht, dass F und S bezüglich $(e_1, \dots, e_n)$ durch die selbe Matrix dargestellt werden, also durch eine Diagonalmatrix mit reellen Eigenwerten. $\square$

7.46. Beispiel. Ein Brillenglas ist eine gekrümmte Fläche. Die Wirkung des Glases auf Lichtstrahlen hängt von der Krümmung ab. Wenn wir das Glas in einem Punkt p flach auf den Tisch legen, können wir eine Seite des Glases als Graph einer Funktion $f\colon U \to \mathbb{R}$ mit $U \subset \mathbb{R}^2$ darstellen. Wenn f mindestens zweimal stetig differenzierbar ist, beschreibt die zweite Ableitung bei p die Krümmung an der Stelle p. Nach dem Satz von Schwarz ist die zweite Ableitung an der Stelle p eine reelle symmetrische Bilinearform (also eine reelle Hermitesche Sesquilinearform) $f''(p)\colon \mathbb{R}^2 \to \mathbb{R}$, und die Krümmung in Richtung $v \in \mathbb{R}^2$ wird gegeben als

$$f''(p)(v,v) \qquad \text{für alle } v \in \mathbb{R}^2 \text{ mit } \|v\| = 1\;.$$

Nach Folgerung 7.45 können wir $f''(p)$ bezüglich einer Orthogonalbasis (v_1, v_2) des $\mathbb{R}^2$ als Diagonalmatrix mit Einträgen κ_1, κ_2 schreiben. Dann nennt man κ_1 und κ_2 die *Hauptkrümmungen* im Punkt p, und v_1, v_2 die *Hauptkrümmungsrichtungen.* Wir dürfen $\kappa_1 \le \kappa_2$ annehmen. Bei einem gewöhnlichen Brillenglas sollten die Hauptkrümmungen und die Hauptkrümmungen über das ganze Glas in etwa konstant bleiben. In diesem Fall muss der Augenarzt dem Optiker die Krümmungen (als Werte in Dioptrien) und eine Hauptkrümmungsrichtung mitteilen. Die andere Hauptkrümmungsrichtung ergibt sich, da beide senkrecht aufeinander stehen.

7.47. Folgerung (Singuläre Werte)**.** *Es seien (V,g) und (W,h) endlichdimensionale $\Bbbk$-Vektorräume mit Skalarprodukten, und es sei $F\colon V \to W$ linear. Dann existieren Orthonormalbasen von V und von W, so dass F bezüglich dieser Basen dargestellt wird durch eine Matrix der Form*

$$\begin{pmatrix} a_1 & 0 & & \cdots & & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ & \ddots & a_{\operatorname{rg} F} & 0 & \cdots & 0 \\ \vdots & & 0 & 0 & \cdots & 0 \\ & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

mit eindeutig bestimmten reellen Einträgen $a_1 \ge \cdots \ge a_{\operatorname{rg} F} > 0$.

Diese Folgerung ist eine Verfeinerung des Rangsatzes 3.13, in dem anstelle von Orthonormalbasen beliebige Basen erlaubt sind. Die Normalform hier sieht äußerlich der Smith-Normalform sehr ähnlich.

Beweis. Die Abbildung $F^*F\colon V \to V$ ist offensichtlich selbstadjungiert und hat nicht-negative Eigenwerte, denn die zugehörige Hermitesche Bilinearform

$$S(u,v) = g\bigl(u, (F^* \circ F)(v)\bigr) = h\bigl(F(u), F(v)\bigr) = (F^*h)(u,v)$$

ist positiv semidefinit nach Bemerkung 7.35 (3). Die Hauptachsentransformation 7.43 liefert eine Orthonormalbasis $(v_1, \dots, v_n)$ aus Eigenvektoren von $F^* \circ F$ zu den Eigenwerten $\lambda_1, \dots, \lambda_n$; dabei sortieren wir die Basisvektoren so, dass $\lambda_1 \ge \cdots \ge \lambda_\ell > 0 = \lambda_{\ell+1} = \cdots = \lambda_n$.

Jetzt betrachten wir die Vektoren $w_1 = F(v_1) \,.\, \lambda_1^{-\frac{1}{2}}$, …, $w_\ell = F(v_\ell) \,.\, \lambda_\ell^{-\frac{1}{2}}$. Da die Faktoren $\lambda_p^{-\frac{1}{2}}$ reell sind, folgt

$$h(w_p, w_q) = \frac{h\bigl(F(v_p), F(v_q)\bigr)}{\sqrt{\lambda_p}\sqrt{\lambda_q}} = \frac{g\bigl(v_p, (F^*F)(v_q)\bigr)}{\sqrt{\lambda_p \lambda_q}} = \delta_{pq}\,\frac{\sqrt{\lambda_q}}{\sqrt{\lambda_p}} = \delta_{pq}\;.$$

Nach Bemerkung 7.17 (1) sind die Vektoren w_1, …, w_ℓ linear unabhängig. Also ergänzen wir mit dem Basisergänzungssatz 3.3 zu einer Basis von W, die wir mit dem Gram-Schmidt-Verfahren 7.18 in eine Orthonormalbasis $(w_1, \dots, w_m)$ überführen. Bezüglich der so konstruierten Basen hat F die angegebene Abbildungsmatrix, wobei $a_p = \sqrt{\lambda_p}$ für alle $p = 1$, …, $\ell = \operatorname{rg} F$.

Die Eindeutigkeitsaussage ergibt sich, indem man aus der gegebenen Abbildungsmatrix die Abbildungsmatrix von F^*F ableitet und die Eindeutigkeitsaussage aus Folgerung 7.43 benutzt. □

7.48. Bemerkung. Die singulären Werte geben also an, wie stark die Abbildung F die Längen in unterschiedlichen Richtungen verzerrt. Wenn wir beispielsweise eine Gummifolie als Fläche im Raum betrachten, dann können wir das als eine Abbildung $f\colon U \to \mathbb{R}^3$ mit $U \subset \mathbb{R}^2$ betrachten. Es sei $p \in U$, dann gibt die Ableitung $F = df(p)\colon \mathbb{R}^2 \to \mathbb{R}^3$ an, wie f am Punkt p die Richtungen im $\mathbb{R}^2$ in den $\mathbb{R}^3$ abbildet. Die singulären Werte $a_1 \geq a_2$ geben das Maximum und das Minimum der Längenverzerrung an. Nach Folgerung 7.47 stehen die zugehörigen Richtungen immer senkrecht aufeinander.

Die singulären Werte a_1, …, $a_{\operatorname{rg} F}$ heißen manchmal auch *verallgemeinerte Eigenwerte* von F. Diese Bezeichnung ist etwas unglücklich, da für eine Matrix A die Eigenwerte von A mit den verallgemeinerten Eigenwerten, also den Eigenwerten von A^*A, nichts zu tun haben müssen. Als Beispiel betrachte einen Jordan-Block $M_\ell(\lambda) \in M_\ell(\Bbbk)$ der Grösse ℓ zum Eigenwert λ. Dann hat

$$M_\ell(\lambda)^* M_\ell(\lambda) = \begin{pmatrix} \lambda^2+1 & \lambda & & 0 \\ \lambda & \ddots & \ddots & \\ & \ddots & \lambda^2+1 & \lambda \\ 0 & & \lambda & \lambda^2 \end{pmatrix}$$

für $\ell = 2$ die Eigenwerte $\lambda^2 + \frac{1}{2} \pm \sqrt{\lambda^2 + \frac{1}{4}}$, und die singulären Werte sind die positiven Wurzeln davon.

Ähnliche Aussagen wie in Folgerung 7.43 lassen sich auch für schiefe Endomorphismen $F \in \operatorname{End}_\Bbbk V$ eines endlich-dimensionalen $\Bbbk$-Vektorraums beweisen. Dazu muss man wieder nur untersuchen, welche der möglichen Normalformen in Satz 7.40 und den Folgerungen 7.41 und 7.42 schiefe Endomorphismen beschreiben.

7.49. Folgerung. *Es sei (V, g) ein endlich-dimensionaler $\Bbbk$-Vektorraum mit Skalarprodukt und $F \in \operatorname{End}_\Bbbk V$. Dann existiert genau dann eine Orthonormalbasis von V, bezüglich der F dargestellt wird*

(1) *durch eine Block-Diagonalmatrix aus* 1×1*-Blöcken* 0 *und* 2×2*-Blöcken vom Typ* (*) *mit* $a = 0$ *falls* $\mathbb{k} = \mathbb{R}$,
(2) *durch eine Diagonalmatrix mit rein imaginären Einträgen falls* $\mathbb{k} = \mathbb{C}$, *beziehungsweise*
(3) *durch eine Diagonalmatrix mit Einträgen der Form* bi *mit* $b \geq 0$ *falls* $\mathbb{k} = \mathbb{H}$,

wenn F *schief ist.*

Beweis. Analog zum Beweis von Folgerung 7.43. □

Besonders interessant ist der Fall, dass F eine Isometrie ist. Aus Kapitel 1 kennen wir Spiegelungen und Drehungen.

7.50. Folgerung. *Es sei* (V, g) *ein endlich-dimensionaler* $\mathbb{k}$*-Vektorraum mit Skalarprodukt und* $F \in \operatorname{End}_{\mathbb{k}} V$*. Dann existiert genau dann eine Orthonormalbasis von* V*, bezüglich der* F *dargestellt wird*

(1) *durch eine Block-Diagonalmatrix aus* 1×1*-Blöcken* ± 1 *und* 2×2*-Blöcken vom Typ* (*) *mit* $a^2 + b^2 = 1$ *falls* $\mathbb{k} = \mathbb{R}$,
(2) *durch eine Diagonalmatrix mit Einträgen vom Betrag* 1 *falls* $\mathbb{k} = \mathbb{C}$, *beziehungsweise*
(3) *durch eine Diagonalmatrix mit Einträgen der Form* $a+bi$ *vom Betrag* 1 *mit* $b \geq 0$ *falls* $\mathbb{k} = \mathbb{H}$,

wenn F *eine lineare Isometrie ist.*

Beweis. Analog zum Beweis von Folgerung 7.43. □

7.51. Bemerkung. In Aufgabe 2 von Blatt 14 zur linearen Algebra I und Bemerkung 4.30 haben wir die Untergruppen

$$\begin{aligned} O(n) &= \left\{\, A \in M_n(\mathbb{R}) \;\middle|\; A^t \cdot A = E_n \,\right\} \\ \text{und} \qquad SO(n) &= \left\{\, A \in O(n) \;\middle|\; \det A = 1 \,\right\} \end{aligned}$$

der Gruppe $GL(n, \mathbb{R})$ kennengelernt. Die Elemente von $O(n)$ sind dadurch charakterisiert, dass sie das Standard-Skalarprodukt erhalten, somit ist die *orthogonale Gruppe* $O(n)$ die Gruppe der linearen Isometrien des $\mathbb{R}^n$. Gleichzeitig ist $O(n)$ auch die Gruppe der Basiswechselmatrizen zwischen Orthonormalbasen, siehe Proposition 2.77 und Bemerkung 7.15 (4).

Die *spezielle orthogonale Gruppe* $SO(n)$ ist die Gruppe der orientierungserhaltenden Isometrien. Gleichzeitig ist sie die Gruppe der Basiswechselmatrizen zwischen gleich orientierten Orthonormalbasen.

Analog betrachten wir die *unitäre* und die *spezielle unitäre Gruppe*

$$\begin{aligned} U(n) &= \left\{\, A \in M_n(\mathbb{C}) \;\middle|\; A^* \cdot A = E_n \,\right\} \\ \text{und} \qquad SU(n) &= \left\{\, A \in U(n) \;\middle|\; \det A = 1 \,\right\} . \end{aligned}$$

Die unitäre Gruppe $U(n)$ ist die Gruppe der linearen Isometrien des $\mathbb{C}^n$ mit dem Standardskalarprodukt, und gleichzeitig die Gruppe der Basiswechselmatrizen zwischen unitären Basen. Für Elemente $A \in U(n)$ gilt

$$1 = \det(A^*A) = |\det a|^2 \ ,$$

und das Beispiel $(e^{it}) \in U(1)$ zeigt, dass alle komplexen Zahlen vom Betrag 1 als Determinante einer unitären Matrix auftreten können. Da wir Orientierungen für komplexe Vektorräume nicht eingeführt haben, ist $SU(n)$ einfach nur die Untergruppe der Isometrien mit Determinante 1.

Über den Quaternionen definieren wir nur die (*kompakte*) *symplektische Gruppe*

$$Sp(n) = \left\{ A \in M_n(\mathbb{H}) \;\middle|\; A^* \cdot A = E_n \right\}$$

der linearen Isometrien des $\mathbb{H}^n$ mit Standardskalarprodukt, beziehungsweise der Basiswechselmatrizen zwischen quaternionisch unitären Basen. Da die Quaternionen nicht kommutativ sind, gibt es keine Determinante, und wir definieren nur diese eine Gruppe.

7.52. Bemerkung. Wir haben wieder eine Reihe von Normalformen kennengelernt und auch ein paar Anwendungen gesehen.

(1) Sei (V, g) ein n-dimensionaler $\Bbbk$-Vektorraum mit Skalarprodukt, dann ist für jede Orthonormalbasis B von V die Basisabbildung ein Isomorphismus $B\colon \Bbbk^n \to V$, so dass B^*g gerade das Standardskalarprodukt auf $\Bbbk^n$ ist. Insbesondere ist die Dimension eine vollständige Invariante für endlich-dimensionale $\Bbbk$-Vektorräume mit Skalarprodukt, ähnlich wie in Bemerkung 3.17 für endlich-dimensionale Vektorräume.

(2) Es sei (V, g) ein $\Bbbk$-Vektorraum mit Skalarprodukt. Wir betrachten zwei Endomorphismen F, $G \in \operatorname{End}_\Bbbk V$ als *metrisch äquivalent,* wenn es eine lineare Isometrie $U \in \operatorname{Aut}_\Bbbk V$ gibt, so dass $G = U^{-1}FU$. Somit sind F und G genau dann äquivalent, wenn es Orthonormalbasen B und C gibt, so dass F bezüglich B die gleiche Darstellung hat wie G bezüglich C. Dann haben wir in Satz 7.40 und den Folgerungen 7.41 7.42 eine Normalform für normale Endomorphismen kennengelernt. Spezialfälle haben wir in den Folgerungen 7.43, 7.49 und 7.50 betrachtet. Für selbstadjungierte Matrizen beispielsweise erhalten wir als vollständige Invariante die Dimension $\dim V$ und das Tupel der nach Größe geordneten reellen Eigenwerte.

   Man beachte, dass nicht nur die Auswahl der betrachteten Endomorphismen spezieller ist als in Kapitel 6, sondern auch die Äquivalenzrelation.

(3) Wir nennen zwei lineare Abbildungen F, $G\colon V \to W$ zwischen Vektorräumen *metrisch äquivalent,* wenn es lineare Isometrien $P \in \operatorname{Aut}_\Bbbk V$ und $Q \in \operatorname{Aut}_\Bbbk W$ gibt, so dass $Q \circ F = G \circ P$, siehe Folgerung 3.16. In diesem Fall liefert Folgerung 7.47 eine Normalform, und $(\dim, \dim W, \operatorname{rg} F)$ bildet zusammen mit dem Tupel der nach Größe geordneten singulären Werte eine vollständige Invariante.

7.53. Bemerkung. Es sei $F \in \operatorname{End}_{\mathbb{R}}(V)$ eine Isometrie eines endlich-dimensionalen Euklidischen Vektorraums (V, g).

(1) Gemäß der Orthonormalbasis aus Folgerung 7.50 zerlegen wir V in eine direkte Summe von Unterräumen, die paarweise zueinander senkrecht stehen. Dann operiert F auf den eindimensionalen Unterräumen U_i mit Eigenwert ± 1, also als $\pm \operatorname{id}_{U_i}$, das heißt als Identität oder als Spiegelung.

   Auf den zweidimensionalen Eigenräumen V_j wirkt V durch eine Matrix vom Typ (*) mit $a^2 + b^2 = 1$ und $b > 0$. Also finden wir einen Winkel $\varphi = \arccos a \in [0, \pi]$, so dass

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} .$$

   Diese Matrix beschreibt eine Drehung des Raumes V_j um den Winkel φ und heißt daher auch einfach *Drehmatrix.*

(2) Gemäß Definition 4.28 heißt F genau dann orientierungserhaltend, wenn $\det F > 0$. Nach Folgerung 4.17 (1) ergibt sich die Determinante als das Produkt der Determinanten der einzelnen Blöcke. Ein 1×1-Block ± 1 hat Determinante ± 1, während eine Drehmatrix stets Determinante $a^2 + b^2 = \cos^2\varphi + \sin^2\varphi = 1$ hat. Somit ist eine Isometrie genau dann orientierungserhaltend, wenn die Anzahl der Spiegelungen in (1), also die Dimension des -1-Eigenraumes, gerade ist.

(3) Wenn V eine feste Orientierung trägt, versuchen wir, in Folgerung 7.50 eine orientierte Basis anzugeben. Das geht immer, wenn ± 1 Eigenwert ist, da wir dann das Vorzeichen des zugehörigen Eigenvektors frei wählen können. In einem Drehblock legt jedoch die Wahl des Winkels $\varphi \in (0, \pi)$ eine Orientierung fest. Wenn wir also nur Drehmatrizen zu Winkeln $\varphi_i \in (0, \pi)$ haben, müssen wir unter Umständen einen Winkel φ durch $-\varphi$ ersetzen. In diesem Fall können wir sagen, dass F entgegen dem mathematischen Drehsinn wirkt. In der Ebene ist eine Drehung im mathematischen Drehrsinn eine Drehung gegen den Uhrzeigersinn, und umgekehrt.

7.5. Affine Räume

Wenn wir bei einem Vektorraum den Nullpunkt „vergessen“, erhalten wir einen affinen Raum. In Definition 3.18 hatten wir bereits affine Unterräume von Vektorräumen kennengelernt. In diesem Abschnitt wollen wir affine Räume etwas abstrakter einführen und auch als metrische Räume betrachten.

Für den Anfang betrachten wir wieder beliebige Schiefkörper $\mathbb{k}$.

7.54. Definition. Es sei V ein $\mathbb{k}$-Vektorraum. Ein *affiner Raum* über V ist eine Menge A, zusammen mit einer Abbildung $+\colon A \times V \to A$, so dass gilt:

(1) für alle $a \in A$ und alle $v, w \in V$ gilt $(a + v) + w = a + (v + w)$,
(2) zu je zwei Punkten $a, b \in A$ existiert genau ein $v \in V$ mit $b = a + v$.

Die *Dimension* von A ist gerade die Dimension von V.

Es sei B ein weiterer affiner Raum über einem $\Bbbk$-Vektorraum W. Eine Abbildung $F\colon A \to B$ heißt *affin,* wenn es eine lineare Abbildung $L\colon V \to W$ gibt, so dass

$$F(a+v) = F(a) + L(v)$$

für alle $a \in A$ und alle $v \in V$. In diesem Fall nennt man F auch lineare Abbildung *über* L.

Eine nichtleere Teilmenge $C \subset A$ heißt *affiner Unterraum,* wenn es einen Untervektorraum $U \subset V$ gibt, so dass für alle $c \in C$ und alle $v \in V$ der Punkt $c + v$ genau dann in C liegt, wenn $v \in U$. Man nennt C dann auch affinen Unterraum *über* U. Zwei affine Unterräume heißen *parallel,* wenn sie über dem gleichen linearen Unterraum von V liegen.

7.55. Beispiel. Wir kennen schon einfache Beispiele.

(1) Jeder Vektorraum ist ein affiner Raum über sich selbst. Die affinen Unterräume im Sinne von Definition 3.18 sind genau die affinen Unterräume im obigen Sinne. Sei $F\colon V \to W$ eine affine Abbildung über der linearen Abbildung $L\colon V \to W$, dann folgt

$$A(v) = A(0+v) = A(0) + L(v) \qquad \text{für alle } v \in V\ .$$

Also haben affine Abbildungen zwischen Vektorräumen stets die Gestalt

$$A(v) = L(v) + w\ ,$$

wobei L linear ist Umgekehrt ist jede Abbildung dieser Form affin.

(2) Es sei V ein Vektorraum und $U \subset V$ ein Unterraum. Dann ist jeder zu U parallele affine Unterraum A selbst ein affiner Raum über dem Untervektorraum U. Beispielsweise ist die Lösungsmenge eines inhomogenen Gleichungssystems ein affiner Raum über der Lösungsmenge des zugehörigen homogenen Gleichungssystems, siehe Proposition 3.21 (3).

7.56. Bemerkung. Wir sammeln ein paar elementare Eigenschaften.

(1) Für jeden Punkt $a \in A$ ist die Zuordnung $v \mapsto a + v$ eine Bijektion von V nach A. Die Umkehrabbildung schreiben wir als Subtraktion

$$-\colon A \times A \to V\ ,$$

so dass $b-a = v$ genau dann, wenn $b = a+v$. Eine andere Bezeichnung ist $\overrightarrow{ab} = b - a$.

(2) Es sei A ein affiner Raum über einem $\Bbbk$-Vektorraum V. Wenn wir einen *Ursprung* $o \in A$ wählen, können wir A und V identifizieren, indem wir $a \in A$ mit dem Vektor $a - o \in V$ und $v \in V$ mit dem Punkt $a + v \in A$ gleichsetzen.

Wir setzen wieder $\Bbbk = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$ und erinnern uns an den Begriff einer Norm auf einem $\Bbbk$-Vektorraum, siehe Bemerkung 7.9. In Definition 7.8 haben wir speziell die Norm $\|\cdot\|_g$ zu einem Skalarprodukt g auf V eingeführt.

7.57. Definition. Es sei A ein affiner Raum über einem $\Bbbk$-Vektorraum V und $\|\cdot\|$ eine Norm auf V. Dann definieren wir die *affine Metrik* $d\colon A\times A\to\mathbb{R}$ zu $\|\cdot\|$ auf A durch

$$d(a,b)=\|a-b\| \qquad \text{für alle } a,\, b\in A\,.$$

Wenn $\|\cdot\|=\|\cdot\|_g$ die Euklidische Norm zu einem Skalarprodukt g auf V ist, nennen wir $d_g=d$ eine *Euklidische Metrik* auf V. Ein affiner Raum mit einer Euklidischen Metrik heißt auch *Euklidischer Raum* (A,d).

Eine affine Abbildung $F\colon A\to B$ zwischen Euklidischen Räumen (A,d) und (B,e) heißt (*affine*) *isometrische Einbettung,* wenn

$$e\bigl(F(a),F(b)\bigr)=d(a,b) \qquad \text{für alle } a,\, b\in A\,,$$

und (*affine*) *Isometrie,* wenn sie darüberhinaus invertierbar ist.

Obwohl es hier nicht gefordert haben, ist es für Studium Euklidischer Räume (A,d) am sinnvollsten, anzunehmen, dass der die zugrundeliegenden Vektorräume reell sind, also über $\Bbbk=\mathbb{R}$ zu arbeiten. Mehr dazu später.

7.58. Beispiel. In der Schule haben Sie die Geometrie der Euklidischen Ebene $(\mathbb{R}^2,d_g)$ studiert, wobei d_g zum Standard-Skalarprodukt auf $\mathbb{R}^2$ gehört. Analog kann man Euklidische Räume $(\mathbb{R}^n,d_g)$ beliebiger Dimension betrachten. Wir nennen d_g später die Standardmetrik.

In der klassischen Newtonschen Mechanik geht man davon aus, dass uns ein dreidimensionaler Euklidischer Raum umgibt. In diesem Raum ist weder ein Ursprung festgelegt (obwohl er von manchen Leuten auf der Erde, von anderen im Mittelpunkt der Sonne oder gar im Mittelpunkt der Galaxie gesehen wird), noch gibt es ausgezeichnete Richtung (wenn wir einen festen Punkt auf der Erde als Ursprung wählen, könnten wir als Richtungen zum Beispiel „Norden“, „Westen“ und „oben“ wählen, aber diese Wahl hängt dann von der Wahl unseres Ursprungs ab).

Auf der anderen Seite gibt es in der klassischen Mechanik die Vorstellung, dass es eine Euklidische Metrik d unabhängig vom Bezugspunkt gibt. Selbst, wenn sich der Ursprung entlang einer Geraden mit konstanter Geschwindigkeit bewegt, soll sich an dieser Metrik nichts ändern. Die zweite dieser Annahmen wird in Einsteins spezieller Relativitätstheorie durch die etwas komplizierteren Lorenzschen Transformationsformeln ersetzt. In der allgemeinen Relativitätstheorie schließlich wird aus dem „flachen“ Euklidischen Raum eine gekrümmte Raumzeit.

7.59. Bemerkung. Wir können Euklidische Räume als metrische Räume betrachten.

(1) Eine Metrik auf einer Menge M ist eine Funktion $d\colon M\times M\to\mathbb{R}$, so dass für alle a, b, $c\in M$ die folgenden Axiome gelten:

(D1) $d(a,b)\ge 0 \quad\text{und}\quad d(a,b)=0\Longleftrightarrow a=b$ (*Positivität*),

(D2) $d(b,a)=d(a,b)$ (*Symmetrie*),

(D3) $d(a,c)\le d(a,b)+d(b,c)$ (*Dreiecksungleichung*).

Dann nennt man (M, d) einen *metrischen Raum.*

Für eine affine Metrik zu einer Norm $\|\cdot\|$ auf V folgen diese Axiome jeweils aus den entsprechenden Axiomen (N1)–(N3) für $\|\cdot\|$.

Auf der anderen Seite kommt nicht jede Metrik auf A von einer Norm, beispielsweise gehört zu keiner Norm die „diskreten Metrik“

$$d(a,b) = \begin{cases} 0 & \text{falls } a = b, \text{ und} \\ 1 & \text{sonst.} \end{cases}$$

Also ist nicht jede Metrik auf einem affinen Raum eine affine Metrik. Das lässt sich auch dadurch erklären, dass die Homogenität (N2) zum Beweis der Symmetrie nur für die Skalare ± 1 benutzt wird.

(2) Es sei $d = d_g$ eine Euklidische Metrik auf A. In der Dreiecksungleichung gilt Gleichheit genau dann, wenn es reelle Zahlen r, $s \geq 0$ gibt, die nicht beide verschwinden, so dass

$$(b-a)\,r = (c-b)\,s \in V\ .$$

Somit zeigen beide Vektoren „in die gleiche Richtung“. Zur Begründung schreiben wir $v = b - a$ und $w = c - b \in V$ und betrachten den Beweis der Dreiecksungleichung in Bemerkung 7.9, wonach

$$\begin{aligned} \|v+w\|_g^2 &= \|v\|_g^2 + 2\operatorname{Re} g(v,w) + \|w\|_g^2 \\ &\leq \|v\|_g^2 + 2\,|g(v,w)| + \|w\|_g^2 \\ &\leq \|v\|_g^2 + 2\,\|v\|_g\,\|w\|_g + \|w\|_g^2 = \big(\|v\|_g + \|w\|_g\big)^2\ . \end{aligned}$$

Wegen der Cauchy-Schwarz-Ungleichung 7.10 wird aus der zweiten Ungleichung genau dann eine Gleichung, wenn v und w linear abhängig sind. Wir wollen annehmen, dass $r \in \mathbb{k}$ mit $w = v.r$ existiert, ansonsten vertauschen wir die Rollen von v und w. Dann gilt

$$\operatorname{Re}\big(g(v, v\,.\,r)\big) = \operatorname{Re}(r)\ \underbrace{g(v,v)}_{\geq 0} \leq |r|\ g(v,v)\ ,$$

und Gleichheit gilt genau dann, wenn r eine nichtnegative reelle Zahl ist. Mit $s = 1$ erhalten wir die obige Behauptung.

(3) Eine *Isometrie* zwischen metrischen Räumen (M, d) und (N, e) ist eine invertierbare Abbildung $F\colon M \to N$, so dass

$$e\big(F(a), F(b)\big) = d(a,b) \qquad \text{für alle } a,\ b \in M\ .$$

Es seien wieder (A, d) und (B, e) Euklidische Räume über $\mathbb{k}$. Wenn es eine Isometrie $F\colon A \to B$ gibt, kann man daraus folgern, dass F linear über $\mathbb{R}$ ist. Der Beweis ist nicht ganz einfach und benutzt unter anderem (2).

Die Abbildung F muss jedoch nicht $\mathbb{k}$-linear sein, falls $\mathbb{k} = \mathbb{C}$ oder $\mathbb{H}$. Aus diesem Grund ist es vom Standpunkt der metrischen Geometrie (also der Geometrie von Mengen M mit einer Metrik d wie in (1)) nicht besonders sinnvoll, Euklidische Räume über $\mathbb{C}$ oder $\mathbb{H}$ zu betrachten.

7.60. Proposition. *Es seien* (A, d_g) *und* (B, d_h) *endlich-dimensionale Euklidische Räume der gleichen Dimension über* $\mathbb{k}$*-Vektorräumen* (V, g) *und* (W, h) *mit Skalarprodukten. Dann gibt es eine affine Isometrie* $F\colon A \to B$.

Mit anderen Worten ist die Dimension eine vollständige Invariante für endlich-dimensionale Euklidische Räume über einem festen Körper $\mathbb{k}$ bis auf affine Isometrie, und $(\mathbb{k}^n, d_g)$ ist eine zugehörige Normalform, wenn d_g die Euklidische Metrik zum Standard-Skalarprodukt bezeichnet. Im Falle $\mathbb{k} = \mathbb{R}$ ist die Dimension wegen der obigen Bemerkung 7.59 (3) sogar eine vollständige Invariante endlich-dimensionaler Euklidischer Räume bis auf Isometrie.

Beweis. Wir wählen jeweils einen Ursprung $o \in A$ und $p \in B$ und identifizieren A und B mit den zugrundeliegenden $\mathbb{k}$-Vektorräumen (V, g) und (W, h) mit Skalarprodukten wie in Bemerkung 7.56 (2). Wegen Bemerkung 7.52 (1) gibt es eine lineare Isometrie $L\colon V \to W$. Dann definieren wir $F\colon A \to B$ durch

$$F(a) = p + L(a - o) .$$

Diese Abbildung ist eine affine Isometrie, denn für alle a, $b \in A$ gilt

$$\begin{aligned} d_h\bigl(F(a), F(b)\bigr) &= \|F(a) - F(b)\|_h = \|p + L(a + o) - p - L(b + o)\|_h \\ &= \|L(a - b)\|_h = \|a - b\|_g = d_g(a, b) . \qquad \square \end{aligned}$$

7.61. Bemerkung. Die *Euklidische Gruppe* oder auch (*Euklidische*) *Bewegungsgruppe* $E(n, \mathbb{k})$ ist die Gruppe der affinen Isometrien von $(\mathbb{k}^n, d)$, wobei d die Standardmetrik sei. Für $\mathbb{k} = \mathbb{R}$ schreiben wir kurz $E(n) = E(n, \mathbb{R})$.

(1) Nach Beispiel 7.55 (1) können wir jedes Element $F \in E(n, \mathbb{k})$ schreiben als

$$v \mapsto w + Av \qquad \text{mit } A \in M_n(\mathbb{k}) .$$

Da F eine Isometrie ist, muss für alle $v \in \mathbb{k}^n$ gelten, dass

$$\|Av\| = \|F(v) - F(0)\| = d\bigl(F(v), F(0)\bigr) = d(v, 0) = \|v\| .$$

Mit Hilfe der Polarisationsformeln aus Bemerkung 7.11 (2)–(4) folgt daraus $\langle Au, Av\rangle = \langle u, v\rangle$ für alle u, $v \in \mathbb{k}^n$, so dass $A \in U(n, \mathbb{k})$ mit $U(n, \mathbb{k}) = O(n)$, $U(n)$ beziehungsweise $Sp(n)$, je nachdem ob $\mathbb{k} = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{H}$. Umgekehrt sieht man leicht, dass die obige Abbildung F eine affine Isometrie, also eine *Bewegung* ist, wenn $A \in U(n, \mathbb{k})$ gilt.

(2) Wir schreiben $F = (w, A)$ für die obige Abbildung F. Wenn wir zwei solche Abbildungen $F = (w, A)$ und $G = (x, B)$ verketten, erhalten wir

$$(F \circ G)(v) = w + A(x + Bv) = (w + Ax) + ABv ,$$

also gilt $(w, A) \circ (x, B) = (w + Ax, AB)$. Somit werden die Matrizen in den zweiten Einträgen der Paare multipliziert, während die Vektoren im ersten Eintrag erst addiert werden, nachdem der zweite Vektor von links mit der Matrix aus dem ersten Paar multipliziert wurde.

Das heißt, als Menge gilt $E(n, \mathbb{k}) = \mathbb{k}^n \times U(n, \mathbb{k})$, aber für die Verknüpfung $\circ$ wird die Wirkung von $U(n, \mathbb{k})$ auf $\mathbb{k}^n$ benutzt. Man nennt daher $E(n,\Bbbk)$ das *semidirekte Produkt* von $\Bbbk^n$ und $U(n,\Bbbk)$ und schreibt entsprechend

$$\begin{aligned} E(n) &= \mathbb{R}^n \rtimes O(n) \;, \\ E(n,\mathbb{C}) &= \mathbb{C}^n \rtimes U(n) \\ \text{und} \qquad E(n,\mathbb{H}) &= \mathbb{H}^n \rtimes Sp(n) \;. \end{aligned}$$

(3) Für den Fall $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$ können wir auch die Untergruppen

$$\begin{aligned} SE(n) &= \mathbb{R}^n \rtimes SO(n) \subset E(n) \\ \text{und} \qquad SE(n,\mathbb{C}) &= \mathbb{C}^n \rtimes SU(n) \subset E(n,\mathbb{C}) \end{aligned}$$

betrachten. Dann ist $SE(n)$ die Gruppe der orientierungserhaltenden Bewegungen.

Wir wollen jetzt eine möglichst geometrische Beschreibung von affinen Isometrien geben. Zusammen mit den Normalformen für Isometrien auf Folgerung 7.50 können wir hieraus leicht eine Normalform für affine Isometrien herleiten.

7.62. Satz. *Es sei (A,d) ein affiner Raum über einem endlich-dimensionalen $\Bbbk$-Vektorraum (V,g) mit Skalarprodukt. Es sei $F\colon A \to A$ eine affine Isometrie über einer linearen Isometrie L. Dann existiert ein Punkt $o \in A$ und ein Vektor x aus dem Eigenraum U von L zum Eigenwert 1, so dass*

$$F(a) = o + x + L(a - o) \;.$$

Der Vektor $x \in U$ ist eindeutig durch A bestimmt. Der Punkt o kann beliebig gewählt werden aus einem affinen, F-invarianten Unterraum $B \subset A$ über U, auf dem F durch Addition von x wirkt.

Wir nennen B die *Achsenmenge* von F, und alle parallel affinen Geraden der Form

$$\bigl\{\, a + x \,.\, r \bigm| r \in \Bbbk \,\bigr\} \subset B$$

mit $a \in B$ heißen *Achsen* von F.

Beweis. Wir wählen zunächst einen beliebigen Punkt als Ursprung, identifizieren A mit V wie in Bemerkung 7.56 (2) und schreiben $F(v) = y + L(v)$ wie in Beispiel 7.55 (1). Es sei $U \subset V$ der Eigenraum zum Eigenwert 1 und $W \subset V$ das orthogonale Komplement von U. Wie im Beweis von Satz 7.40 sind U und W invariant unter L, und $L|_U = \mathrm{id}_U$.

Wir schreiben $y = (x,z) \in U \oplus W = V$. Die Abbildung $\mathrm{id}_W - L|_W$ ist invertierbar, denn 1 ist kein Eigenwert mehr von $L|_W$. Wir bestimmen $q \in W$ so, dass $q - L(q) = z$. Dann setzen wir $o = (p,q) \in U \oplus W \cong A$ für ein beliebiges $p \in U$. Für alle $v = (u,w) \in U \oplus W = V$ folgt

$$(*) \quad \begin{aligned} F(o+v) = (x,z) + L(p+u, q+w) &= \bigl(x+p+u, z+L(q)+L(w)\bigr) \\ &= \bigl(x+p+u, q+L(w)\bigr) = o + (x,0) + L(v) \;. \end{aligned}$$

Wir wählen also o als unseren neuen Ursprung und haben die gesuchte Darstellung von F gefunden.

Da F eine affine Abbildung über L ist, ist L durch F eindeutig bestimmt. Wir betrachten $o' = o + (p', q')$ als neuen Ursprung und $v = (u, w)$ mit p', $u \in U$ sowie q', $w \in W$. Dann betrachten wir den Vektor

$$\begin{aligned} x' &= F(o' + v) - o' - L(v) \\ &= F\bigl(o + (p' + u, q' + w)\bigr) - o - (p', q') - L(u, w) \\ &= \bigl(p' + u + x - p' - u, L(q' + w) - q' - L(w)\bigr) = \bigl(x, L(q') - q'\bigr) . \end{aligned}$$

Dann gilt

$$L(x') = \bigl(x, (L \circ L)(q') - L(q') = \bigl(x, L(q') - q'\bigr) = x'$$

genau dann, wenn

$$(\mathrm{id}_W - L|_W) \circ (\mathrm{id}_W - L|_W)(q') = 0 .$$

Nach Konstruktion ist $\mathrm{id}_W - L|_W$ invertierbar, also gilt das genau dann, wenn $q' = 0$, das heißt, wenn $x' = x$ ist und $o \in B$. Damit ist die Eindeutigkeitsaussage bewiesen. □

Wir wollen mit Hilfe dieses Satzes Normalformen von Isometrien verstehen

7.63. Beispiel. Es sei A ein zweidimensionaler reeller Euklidischer Raum über einem zweidimensionalen Euklidischen Vektorraum (V, g) und F eine affine Isometrie von A über eine linearen Isometrie L von V. Es sei wieder U der Eigenraum von L zum Eigenwert 1. Wir stellen L wie in Folgerung 7.50 (1) dar und unterscheiden folgende Fälle.

(1) Es sei F orientierungserhaltend.
   (a) Es sei $L = \mathrm{id}_V$, dann ist $U = V$, und $B = A$ ist die Achsenmenge. Falls $x = 0$ ist, ist $F = \mathrm{id}_A$ die Identität, ansonsten ist $F(a) = a + x$ eine *Verschiebung.*
   (b) Ansonsten ist L eine Drehung, also ist $U = \{0\}$ und daher $x = 0$. Die Achsenmenge B besteht aus einem einzigen Punkt o, und F ist eine *Drehung* um o.
(2) Wenn F nicht orientierungserhaltend ist, sind die Eigenräume zu den Eigenwerten ± 1 nach Bemerkung 7.53 (2) jeweils eindimensional, also ist die Achsenmenge B eine Gerade. Falls $x = 0$, ist F die *Spiegelung* an dieser Geraden, ansonsten eine *Gleitspiegelung.*

7.64. Beispiel. Sei A jetzt ein dreidimensionaler reeller Euklidischer Raum und V, F, L und $U \subset V$ wie oben.

(1) Es sei F orientierungserhaltend.
   (a) Es sei $L = \mathrm{id}_V$, dann ist $U = V$, und wie oben ist F entweder die *Identität* oder eine *Verschiebung.*
   (b) Ansonsten ist L eine Drehung, und U ist eindimensional. Also ist die Achsenmenge B eine Gerade. Falls $x = 0$, ist F eine *Drehung* um die Gerade B, ansonsten eine *Schraubung.*

(2) Wenn F orientierungsumkehrend ist, ist der Eigenraum von L zum Eigenwert 1 mindestens eindimensional.
   (a) Wenn L einen zweidimensionalen Eigenraum zum Eigenwert 1 hat, ist die Achsenmenge B eine Ebene. In diesem Fall ist F eine *Spiegelung* an B, falls $x = 0$, ansonsten eine *Gleitspiegelung.*
   (b) Wenn L in der Darstellung aus Folgerung 7.50 (1) durch einen Eigenwert -1 und einen Drehblock beschrieben wird, erhalten wir eine *Drehspiegelung.* Die Achsenmenge enthält nur einen Punkt o. Dabei wird zunächst an einer Ebene durch o gespiegelt, anschließend um die Gerade durch o senrecht zu dieser Ebene gedreht.
   (c) Einen Spezialfall davon erhalten wir, wenn der Eigenwert -1 Multiplizität 3 hat. In diesem Fall enthält die Achsenmenge ebenfalls nur einen Punkt o, und F ist eine *Punktspiegelung* an o.

7.6. Bilinearformen und quadratische Funktionen

In diesem Abschnitt betrachten wir Hermitesche Sesquilinearformen, die nicht notwendig positiv definit sind. Ein Beispiel dafür ist die Lorentz-Metrik in der speziellen Relativitätstheorie.

7.65. Definition. Es sei S eine Hermitesche Sesquilinearform auf einem $\Bbbk$-Vektorraum V. Der *Ausartungsraum* oder *Kern* von V ist definiert als

$$\ker S = \left\{ v \in V \;\middle|\; S(w,v) = 0 \text{ für alle } w \in V \right\} .$$

Seine Dimension heißt auch die *Nullität* $n_0(S)$ von S. Wenn $n_0(S) = 0$ gilt, heißt S *nicht ausgeartet,* sonst *ausgeartet.*

Ein Unterraum $U \subset V$ heißt *positiv* (*negativ*) bezüglich S, wenn $S|_U = S|_{U \times U}$ positiv (negativ) definit ist. Er heißt *maximal positiv* (*maximal negativ*), wenn kein Unterraum $W \subset V$ mit $U \subsetneqq W$ positiv (negativ) ist. Wir bezeichnen die Dimension eines maximalen positiven (negativen) Unterraums mit $n_\pm(S)$, dann heißt $n_-(S)$ auch der *Index* von S.

Wir werden später sehen, dass $n_+(S)$ und $n_-(S)$ nicht von der Wahl des maximalen Unterraums abhängen, und dass $\dim V = n_+(S) + n_-(S) + n_0(S)$.

7.66. Beispiel. Das *Lorentz-Produkt* auf $\Bbbk^{n+1}$ ist die Hermitesche Sesquilinearform zur Matrix

$$\begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} ,$$

wobei man die Standardbasisvektoren der Einfachheit halber mit $e_0, \dots, e_n$ durchnummeriert. Ein Vektor $v \in \Bbbk^n$ heißt *zeitartig,* wenn $S(v,v) < 0$ (Beispiel: e_0), *raumartig,* wenn $S(v,v) > 0$ (Beispiel: $e_1, \dots, e_n$), und *lichtartig,* wenn $S(v,v) = 0$ (Beispiel:. $e_0 \pm e_i$ mit $i \geq 1$). Man beachte, dass S nicht ausgeartet ist, und dennoch Vektoren mit $S(v,v) = 0$ existieren können.

Das Lorentz-Produkt hat Index 1, denn der von e_0 erzeugte Unterraum ist maximal negativ, und es kann keinen negativen Unterraum der Dimension ≥ 2 geben: Seien v, $w \in \mathbb{k}^{n+1}$ linear unabhängig, dann gibt es eine Linearkombination $v \,.\, r + w \,.\, s \neq 0$, deren nullte Koordinate verschwindet, also folgt $S(v \,.\, r + w \,.\, s, v \,.\, r + w \,.\, s) > 0$.

7.67. Satz (Sylvesterscher Trägheitssatz)**.** *Es sei S eine Hermitesche Sesquilinearform auf einem endlich-dimensionalen $\mathbb{k}$-Vektorraum V, dann existiert eine Basis, bezüglich der S durch eine Gramsche Matrix der Form*

$$(*)\qquad \begin{pmatrix} 1 & & & & & & & & 0 \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ 0 & & & & & & & & 0 \end{pmatrix}$$

dargestellt wird. Die Anzahlen der Einträge 1, -1 *und* 0 *sind gerade* $n_+(S)$, $n_-(S)$ *und* $n_0(S)$.

Man nennt das Tripel (n_+, n_-, n_0) auch die *Signatur* der Hermiteschen Sesquilinearform. Wenn S nicht ausgeartet ist, heißt das Paar (n_+, n_-) oder auch die Differenz $n_+ - n_-$ die Signatur von S. Die Signatur (n_+, n_-, n_0) bildet eine vollständige Invariante für $\mathbb{k}$-Vektorräume mit Hermitescher Sesquilinearform. Sei A die obige Matrix, dann ist $\mathbb{k}^n$ mit $S(x, y) = x^* A y$ die zugehörige Normalform.

Beweis. Die Existenz der Basis B lässt sich auf zweierlei Weisen zeigen. Zunächst, indem man ein beliebiges Skalarprodukt g auf V wählt und dann Folgerung 7.45 anwendet. Anschließend ersetzt man die Basisvektoren $v_i \neq \ker S$ durch $e_i = v_i \,.\, \frac{1}{\sqrt{|S(v_i,v_i)|}}$ (beachte, dass $S(v_i, v_i) \in \mathbb{R}$). Bezüglich dieser Basis hat die Gramsche Matrix von S die gewünschte Gestalt. Dieser Beweis ist nicht konstruktiv, da Satz 7.40 und die anschließenden Folgerungen nicht konstruktiv sind.

Alternativ wählt man zunächst eine Basis $e_{n-n_0+1}, \dots, e_n$ von $\ker S$ mit dem Gauß-Verfahren und fixiert ein Komplement $W \subset V$ vom Kern. Dann wählt man eine Basis $v_1, \dots, v_{n-n_0}$ von W. Als nächstes konstruieren wir induktiv Vektoren $e_1, \dots, e_{n-n_0}$ mit einem modifizierten Gram-Schmidt-Verfahren.

Seien dazu $e_1, \dots, e_{p-1}$ bereits konstruiert und $S(e_q, v_r) = 0$ für alle $q < p$ und alle $p \leq r \leq n - n_0$. Falls $S(v_p, v_p) = 0$, existiert v_r mit $r > p$ und $S(v_p, v_r) \neq 0$, da $v_p \notin \ker S$. Für $t \in \mathbb{k}$ betrachte den Vektor $v_p + v_r t$, dann gilt

$$S(v_p + v_r t, v_p + v_r t) = 2\,\mathrm{Re}\big(S(v_p, v_r)\, t\big) + |t|^2\; S(v_r, v_r)\;.$$

Für hinreichend kleine $t \neq 0$ ist $2\,|S(v_p,v_r)\,t| > \left|t^2\,S(v_r,v_r)\right|$. Wenn wir also für t ein kleines reelles Vielfaches von $\overline{S(v_p,v_r)}$ wählen, folgt

$$S(v_p + v_r t, v_p + v_r t) \neq 0 \; .$$

Wir ersetzen v_p durch $v_p + v_r t$, dann gilt nach wie vor $S(e_q, v_p) = 0$ für alle $q < p$.

Da $S(v_p, v_p) \in \mathbb{R} \setminus \{0\}$, können wir jetzt

$$e_p = v_p \,.\, \frac{1}{\sqrt{|S(v_p,v_p)|}}$$

definieren. Anschliessend machen wir die Vektoren $v_{p+1}, \dots, v_{n-n_0}$ orthogonal zu e_p bezüglich S, indem wir v_r für alle $r > p$ durch

$$v_r - e_p \,.\, \underbrace{S(e_p,e_p)}_{=\pm 1}\, S(e_p, v_r)$$

ersetzen. Falls $p < n - n_0$, ersetzen wir p durch $p + 1$ und machen weiter.

Zum Schluss sortieren wir die Basisvektoren so um, dass die Diagonaleinträge in der gewünschten Reihenfolge dastehen. Wir erhalten eine Diagonalmatrix A wie in (*). Die Eindeutigkeit von $n_0 = n_0(S) = \dim \ker S$ ist klar.

Zur Eindeutigkeit von n_+ sei $U \subset V$ ein positiver Unterraum. Falls $n_+ < \dim U$, finden wir aus Dimensionsgründen einen Vektor $v \in V_+ = \langle e_1, \dots, e_{n_+}\rangle$ mit $S(u,v) = 0$ für alle $u \in U$, also $v \in U^\perp \cap V_+$, insbesondere $U \oplus \langle v \rangle$ positiv und U daher nicht maximal positiv.

Sei umgekehrt $\dim U > n_+$, dann betrachte $V_- \oplus V_0 = \langle e_{n_+ +1}, \dots, e_n\rangle$. Aus Dimensionsgründen existiert $u \in U \cap (V_- \oplus V_0)$, also gilt $S(u,u) \leq 0$, und U ist nicht positiv. Also hat ein maximaler positiver Unterraum gerade die Dimension $n_+ = n_+(S)$. Analog hat ein maximaler negativer Unterraum Dimension $n_- = n_-(S)$. □

7.68. Bemerkung. Es sei V ein n-dimensionaler $\Bbbk$-Vektorraum mit einer Hermiteschen Sesquilinearform S. Nach Sylvesters Trägheitssatz 7.67 dürfen wir $V = \Bbbk^n$ annehmen, wobei S durch die obige Diagonalmatrix (*) gegeben wird. Es sei $p = n_+(S)$ und $q = n_-(S)$. Wir interessieren uns für die Untergruppe der Automorphismengruppe $GL(n,\Bbbk)$, die die Form S erhalten, also

$$G = \left\{ F \in GL(n,\Bbbk) \;\middle|\; F^*S = S \right\} .$$

(1) Falls $p + q = n$ gilt, ist S nicht ausgeartet. In diesem Fall heißt die entsprechende Gruppe $U(p,q;\Bbbk)$, beziehungsweise

$$\begin{aligned} O(p,q) &= U(p,q;\mathbb{R}) \,, \\ U(p,q) &= U(p,q;\mathbb{C}) \\ \text{und} \qquad Sp(p,q) &= U(p,q;\mathbb{H}) \,. \end{aligned}$$

Im Falle $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$ haben wir Determinanten zur Verfügung und definieren

$$\begin{aligned} SO(p,q) &= O(p,q) \cap SL(n,\mathbb{R}) \\ \text{und} \qquad SU(p,q) &= U(p,q) \cap SL(n,\mathbb{C}) \,. \end{aligned}$$

Es besteht eine gewisse formale Analogie zu den Gruppen aus Bemerkung 7.51, beispielsweise gilt

$$SO(1,1) = \left\{ \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} ,$$
$$SO(2) = \left\{ \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} .$$

Für diese Gruppen ist jedoch das Analogon zu Folgerung 7.50 im Allgemeinen nicht mehr richtig. Dazu betrachten wir für $0 \neq t \in \mathbb{R}$ die Matrix

$$A = \begin{pmatrix} 1+ti & t \\ t & 1-ti \end{pmatrix} \in M_2(\mathbb{C}) .$$

Man rechnet nach, dass

$$A^* \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ,$$

so dass $A \in U(1,1)$. Es gilt sogar $\det A = (1-ti)(1+ti) - t^2 = 1$, also $A \in SU(1,1)$. Das charakteristische Polynom von A ist

$$\chi_A(X) = X^2 - 2X + 1 = (X-1)^2 ,$$

aber da $A \neq E_2$, ist der 1-Eigenraum nicht zweidimensional, und somit ist A nicht diagonalisierbar.

(2) Sei jetzt S ausgeartet. Dann haben Elemente F der obigen Gruppe G die Blockgestalt

$$F = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \in M_n(\Bbbk)$$

mit $A \in U(p,q;\Bbbk)$, $C \in M_{n_0,p+q}(\Bbbk)$ beliebig, und $D \in GL(n_0,\Bbbk)$. Das liegt daran, dass $F(\ker S) \subset \ker S$ gelten muss, während sich umgekehrt das Skalarprodukt $S(F(v),F(w))$ nicht ändert, wenn man zu $F(v)$ oder $F(w)$ beliebige Elemente des Kerns hinzuaddiert.

Wir lassen in der Definition von Sesquilinearformen die Konjugation weg und erhalten den Begriff der Bilinearform. Im Moment können wir jeden beliebigen Körper $\Bbbk$ zulassen.

7.69. Definition. Es sei $\Bbbk$ ein Körper und V ein $\Bbbk$-Vektorraum. Eine Abbildung $B\colon V \times V \to \Bbbk$ heißt *Bilinearform,* wenn für alle $u, v \in V$ die Abbildungen

(B1)
$$B(u,\,\cdot\,)\colon V \to \Bbbk \quad \text{mit} \quad v \mapsto B(u,v) \quad \text{und}$$
$$B(\,\cdot\,,v)\colon V \to \Bbbk \quad \text{mit} \quad u \mapsto B(u,v)$$

linear sind. Eine Bilinearform heißt *symmetrisch,* wenn für alle $u, v \in V$ gilt

(B2)
$$B(v,u) = B(u,v) \in \Bbbk .$$

7.70. Bemerkung. Wegen Bemerkung 7.4 (2) sind Bilinearformen über nicht kommutativen Schiefkörpern nicht sinnvoll definiert. Daher haben wir oben nur Körper zugelassen.

(1) Über $\mathbb{R}$ sind Hermitesche Sesquilinearformen und symmetrische Bilinearformen das gleiche. Insbesondere beschreibt der Trägheitssatz 7.67 von Sylvester alle reellen symmetrischen Bilinearformen
(2) Jede symmetrische Bilinearform auf einem n-dimensionalen $\mathbb{C}$-Vektorraum wird bezüglich einer geeigneten Basis durch eine Matrix der Gestalt

$$\begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{pmatrix}$$

dargestellt. Die Anzahl der Nullen auf der Diagonalen ist gerade $n_0(B) = \dim \ker B$, dabei ist $\ker B$ analog zu Definition 7.65 definiert.

Zum Beweis gehen wir analog vor wie im zweiten Beweis des Trägheitssatzes 7.67. Beim Normieren ersetzen wir allerdings einen Vektor v_p mit $B(v_p, v_p) \neq 0$ durch

$$e_p = v_p \cdot \frac{1}{\sqrt{B(v_p, v_p)}} \,,$$

so dass jetzt stets $B(e_p, e_p) = 1$ gilt.
(3) Da $\mathbb{H}$ nicht kommutativ ist, können wir keine Bilinearformen über $\mathbb{H}$ definieren.

7.71. Definition. Es sei $\Bbbk$ ein Körper und V ein $\Bbbk$-Vektorraum. Eine Abbildung $q\colon V \to \Bbbk$ heißt *quadratische Funktion,* wenn eine symmetrische Bilinearform B auf V, eine Linearform $\alpha \in V^*$ und eine Konstante $c \in \Bbbk$ existieren, so dass

$$q(v) = B(v, v) + \alpha(v) + c \,.$$

Wir nennen q *nicht ausgeartet,* wenn B nicht ausgeartet ist. Die Nullstellenmenge Q einer quadratischen Funktion heißt auch *Quadrik* oder *Hyperfläche zweiten Grades.*

Eine quadratische Funktion q ist so etwas wie ein Polynom vom Grad ≤ 2 in einer Variablen aus dem Vektorraum V, und die Quadrik $Q = q^{-1}(0)$ ist ihre Nullstellenmenge. In der Analysis lernt man im Fall $\Bbbk = \mathbb{R}$, dass die Nullstellenmenge eine glatte Hyperfläche ist, wenn 0 ein regulärer Wert von q ist. Man beachte, dass es entartete Fälle gibt, in denen Q nicht glatt oder noch nicht einmal eine Hyperfläche ist, siehe Beispiele 7.74, 7.75. In diesen Fällen ist 0 kein regulärer Wert von q.

Im Folgenden müssen wir durch 2 teilen können, daher erinnern wir uns an die Charakteristik $\chi(\Bbbk)$ eines Körpers aus Definition 2.14.

7.72. Satz. *Es sei $\Bbbk$ ein Körper der Charakteristik $\chi(\Bbbk) \neq 2$, es sei V ein $\Bbbk$-Vektorraum und q quadratische Funktion auf V. Dann existiert eine invertierbare affine Abbildung $F\colon V \to V$, eine symmetrische Bilinearform B auf V,* *ein Komplement W von $U = \ker B$, eine Linearform $\alpha \in U^*$ und $c \in \Bbbk$, so dass*

$$(q \circ F)(u + w) = B(w, w) + \alpha(u) + c \qquad \textit{für alle } u \in \ker S \textit{ und } w \in W\ .$$

Beweis. Sei $q(v) = B(v, v) + \beta(v) + b$ für eine symmetrische Bilinearform B, eine Linearform $\beta \in V^*$ und eine Konstante $b \in \Bbbk$. Wir wählen ein Komplement W von $U = \ker B$, so dass $V = U \oplus W$. Wir definieren $\alpha \in U^*$ und $\gamma \in W^*$ durch

$$\beta(u + w) = \alpha(u) + \gamma(w) \qquad \text{für alle } u \in \ker B \text{ und } w \in W\ .$$

Ähnlich wie in Proposition 7.22 fassen wir $B|_W$ als linearen Isomorphismus $B\colon W \to W^*$ mit $B(w) = B(w, \cdot\,)$ auf. Dann existiert $x = B^{-1}(\gamma) \in W$ mit $2B(x, w) = \gamma(w) = \beta(w)$ für alle $w \in W$. Es sei $F\colon V \to V$ die Verschiebung $F(v) = v - x$. Für $v = u + w$ mit $u \in U$ und $w \in W$ gilt dann

$$\begin{aligned} (q \circ F)(v) &= B(w - x, w - x) + \alpha(u) + \gamma(w - x) + b \\ &= B(w, w) + \alpha(u) - 2B(x, w) + \gamma(w) + b + B(x, x) - \gamma(x) \\ &= B(w, w) + \alpha(u) + c \end{aligned}$$

mit $c = b + B(x, x) - \gamma(x)$. □

Im Beweis haben wir als affine Abbildung also nur eine Verschiebung gewählt, um eine quadratische Ergänzung durchzuführen. Im Falle $\Bbbk = \mathbb{R}$ oder $\mathbb{C}$ und $V = \Bbbk^n$ würden wir zusätzlich noch einen linearen Isomorphismus dazuschalten, so dass die Form B auf $\Bbbk^n$ durch eine der speziellen Formen aus Bemerkung 7.70 (1) oder (2) dargestellt wird, und so dass entweder $\alpha = 0$ oder $\alpha = \varepsilon_n$ gilt.

Um die Gestalt von $Q = q^{-1}(0) \subset V$ im Falle $\Bbbk = \mathbb{R}$ darzustellen, nehmen wir an, dass B tatsächlich durch die Matrix (*) wie im Trägheitssatz 7.67 von Sylvester dargestellt wird und die Signatur durch das Tripel (n_+, n_-, n_0) gegeben ist. Außerdem definieren wir noch folgende Mengen:

$$\begin{aligned} V^+ &= \langle e_1, \dots, e_{n_+} \rangle\ , \\ V^- &= \langle e_{n_++1}, \dots, e_{n-n_0} \rangle\ , \\ S^+ &= \bigl\{\, v \in V^+ \bigm| B(v, v) = 1 \,\bigr\}\ , \\ S^- &= \bigl\{\, v \in V^- \bigm| B(v, v) = -1 \,\bigr\}\ , \\ \text{und} \qquad U' &= \langle e_{n-n_0+1}, \dots, e_{n-1} \rangle\ , \end{aligned}$$

dann sind S^+, S^- gerade die „Einheitssphären“ im positiven beziehungsweise im negativen Unterraum, und $U' = \ker \alpha \subset U = \ker B$ falls $\alpha \neq 0$.

7.73. Folgerung. *Es sei V ein endlich-dimensionaler reeller Vektorraum, es sei B eine symmetrische Bilinearform auf V, und $W = V^+ \oplus V^- \subset V$ ein Komplement von $U = \ker S$. Es seien $\alpha \in U^*$, $c \in \mathbb{R}$ und*

$$q(u, w) = B(w, w) + \alpha(u) + c \qquad \textit{für alle } u \in \ker S \textit{ und } w \in W\ .$$

Dann hat $Q = q^{-1}(0)$ eine der folgenden Gestalten.

(1) *Falls $\alpha = 0$, gilt $Q = U \times Q'$, und*

(a) *falls* $c = 0 \dots$
   (i) *und* $n_+ = 0$ *oder* $n_- = 0$, *ist* $Q' = \{0\}$,
   (ii) *und* n_+, $n_- \geq 1$, *ist* Q' *ein* Doppelkegel
$$Q' = \left\{ (v_+ r, v_- r) \;\middle|\; v_+ \in S^+,\ v_- \in S^-,\ \textit{und}\ 0 \leq r \in \mathbb{R} \right\} ;$$
(b) *falls* $c > 0 \dots$
   (i) *und* $n_- = 0$, *ist* $Q' = \emptyset$,
   (ii) *und* $n_- \neq 0$, *wird* Q' *durch* $V^+ \times S^-$ *parametrisiert, wobei*
$$Q' = \left\{ \left(v_+, v_- \sqrt{c + B(v_+, v_+)}\right) \;\middle|\; v_+ \in V^+\ \textit{und}\ v_- \in S^- \right\} ;$$
(c) *falls* $c < 0$ *hat* Q' *eine entsprechende Gestalt wie in* (1.b), *aber mit den Rollen von* V^+ *und* V^- *vertauscht.*

(2) *Falls* $\alpha \neq 0$*:* $Q = \ker \alpha \times \Gamma$, *dabei ist* Γ *der Graph der nicht-ausgearteten quadratischen Funktion*
$$w \longmapsto -(B(w, w) + c)$$
über $W = V_+ \oplus V_-$.

Beweis. Man überzeugt sich, dass die Fallunterscheidung in der Folgerung vollständig ist. Es reicht also, Fall für Fall zu betrachten. Wir betrachten auf $V^\pm$ die Norm $\|v_\pm\| = \sqrt{\pm B(v_\pm, v_\pm)}$.

Im Fall (1) hängt $q(u, w)$ nicht von u ab, also sei
$$Q' = \left\{ w \in W \;\middle|\; q(0, w) = 0 \right\} ,$$
dann gilt $(u, w) \in Q$ genau dann, wenn $w \in Q'$, also gilt $Q = U \times Q'$. Ab sofort betrachten wir also nur noch die nicht-ausgeartete quadratische Form
$$q'(w) = q|_W(w) = B(w, w) + c$$
auf W.

Im Fall (1.a) ist $c = 0$. Falls (1.a.i) mit $n_- = 0$ vorliegt, folgt $B(w, w) \geq 0$, und $q'(w) = B(w, w) = 0$ gilt genau dann, wenn $w = 0$ ($B|_{W \times W}$ ist also positiv definit). Analoges gilt für $-q'$, falls $n_+ = 0$ gilt.

Im Fall (1.a.ii) sei $w = (v_+ r, v_- r) \in V^+ \oplus V^- = W$ mit $v_\pm \in S^\pm$ und $r \geq 0$, dann folgt
$$q'(w) = \|v_+\|^2\, r^2 - \|v_-\|^2\, r^2 = 0 ,$$
da $\|v_+\|^2 = \|v_-\|^2 = 1$ nach Annahme, also $(v_+ r, v_- r) \in Q'$. Sei umgekehrt $w = (w_+, w_-) \in Q' \subset V^+ \oplus V^-$, dann folgt
$$0 = q'(w) = \|w_+\|^2 - \|w_-\|^2 ,$$
also dürfen wir $r = \|w_+\| = \|w_-\|$ setzen. Falls $w_+ = w_- = 0$, dürfen wir $v_\pm \in S^\pm$ beliebig wählen; das geht, da $S^\pm \neq \emptyset$ falls $n_\pm \geq 1$. Andernfalls setzen wir $v_\pm = w_\pm \frac{1}{n} \in S^\pm$ und erhalten $w = (v_+ r, v_- r)$ wie oben.

Im Fall (1.b) ist $c > 0$. Im Fall (1.b.i) folgt $q'(w) > 0$ für alle $w \in W$, also $Q' = \emptyset$.

Im Fall (1.b.ii) gilt entsprechend $w_- \neq 0$ für alle $w = (w_+, w_-) \in Q'$, und es folgt
$$\|w_-\|^2 = \|w_+\|^2 + c \;,$$
also erhalten wir für jeden Vektor $v_+ \in V^+$ und jede Richtung $v_- \in V^-$ eine eindeutige Lösung $(v_+, v_- r) \in Q'$ mit
$$r = \sqrt{c + \|v_+\|} \;.$$

Im Fall (1.c) ersetzen wir q' durch $-q'$ und machen wie in (1.b) weiter. Dabei tauschen V^+ und V^- ihre Rollen.

Im Fall (2) gilt $n_0 \geq 1$, und wir dürfen wie oben gesagt annehmen, dass $\alpha = \varepsilon^n$. Für einen Vektor
$$v = (v_+, v_-, u', e_n h) \in V^+ \oplus V^- \oplus U' \oplus \langle e_n \rangle$$
gilt also $q(v) = 0$ genau dann, wenn
$$h = -B(w, w) - c = \|v_-\|^2 - \|v_+\|^2 - c \;.$$
Für jede Wahl von (v_+, v_-, u') gibt es also genau eine Zahl $h \in \mathbb{R}$, so dass $(v_+, v_-, u', e_n h) \in Q$, und h hängt nicht von $u' \in U' = \ker \alpha$ ab. Also hat Q die angegebene Gestalt. $\square$

7.74. Beispiel. Es sei $Q \subset \mathbb{R}^2$ eine Quadrik. Es sei $q\colon \mathbb{R}^2 \to \mathbb{R}$ eine quadratische Funktion in der obigen Normalform. Wir schreiben $v = (x, y) \in \mathbb{R}^2$. Wir geben im jeden einzelnen der Fälle aus Folgerung 7.73 die Gestalt von Q an.

Im Fall (1.a.i) ist $Q' = \{0\}$ ein Punkt. Falls $n_0 = 0$, ist auch Q ein Punkt. Andernfalls erhalten wir eine Gerade $Q = Q' \times \mathbb{R}$, falls $n_0 = 1$, oder den gesamten $\mathbb{R}^2 = Q' \times \mathbb{R}^2$, falls $n_0 = 2$.

Im Fall (1.a.ii) folgt $n_+ = n_- = 1$ und $n_0 = 0$. Die Quadrik Q besteht aus den beiden Geraden $y = x$ und $y = -x$.

Im Fall (1.b.i) ist $Q = Q' = \emptyset$.

Im Fall (1.b.ii) gibt es drei Möglichkeiten. Falls $n_- = 2$ und $n_- = n_0 = 0$, ist
$$Q = \left\{ \, (x, y) \in \mathbb{R}^2 \;\middle|\; x^2 + y^2 = c \, \right\}$$
ein Kreis. Falls $n_- = 1 = n_+$ und $n_0 = 0$, besteht
$$Q = \left\{ \, (x, y) \;\middle|\; y = \pm\sqrt{c + x^2} \, \right\}$$
aus den zwei Ästen einer Hyperbel. Falls $n_- = 1 = n_0$ und $n_+ = 0$, besteht Q nur aus den zwei Geraden $y = \pm\sqrt{c}$, da S^0 nur aus den zwei Punkten $\pm 1 \in \mathbb{R}$ besteht.

Der Fall (1.c) liefert die gleichen geometrischen Figuren wie (1.b).

Im Fall (2) sei $\alpha(x, y) = y$, so dass Q der Graph einer quadratischen Funktion $q'\colon \mathbb{R} \to \mathbb{R}$ ist. Wir unterscheiden drei Fälle. Falls $n_0 = 2$, ist q' konstant, und Q eine zur x-Achse parallele Grade. Falls $n_+ = 1 = n_0$ und $n_- = 0$, ist
$$Q = \left\{ \, (x, y) \;\middle|\; y = -c - x^2 \, \right\}$$

eine nach unten offene Parabel. Fall $n_- = 1 = n_0$ und $n_+ = 0$, ist Q entsprechend eine nach oben offene Parabel.

Man nennt alle diese Figuren auch Kegelschnitte, da sich die meisten (alle bis auf die leere Menge, den gesamten $\mathbb{R}^2$ und die zwei parallelen Geraden) als Schnitt eines Doppelkegels im $\mathbb{R}^3$ mit einer Ebene darstellen lassen. Man erhält umgekehrt jede Quadrik im $\mathbb{R}^2$ aus einem der obigen Beispiele durch eine invertierbare affine Abbildung. Wenn diese Abbildung keine affine Isometrie ist, kann sich das dadurch bemerkbar machen, dass aus dem runden Kreis eine Ellipse, aus der Hyperbel mit rechtem Winkel zwischen den Asymptoten eine Hyperbel mit einem anderen Asymptotenwinkel, aus der Einheitsparabel eine Parabel anderer Größe, und aus zwei sich rechtwinklig schneidenden Geraden zwei sich unter einem beliebigen Winkel $\neq 0$ schneidende Geraden werden.

7.75. Beispiel. Wir betrachten zum Schluss Quadriken im $\mathbb{R}^3$. Dabei listen wir aber nur noch die verschiedenen auftretenden Formen und in Klammern die Tripel (n_+, n_-, n_0) auf.

Im Fall (1.a.i) erhalten wir einen Punkt ((3,0,0) oder (0,3,0)), eine Gerade ((2,0,1) oder (0,2,1)), eine Ebene ((1,0,2) oder (0,1,2)) oder den gesamten $\mathbb{R}^3$ ((0,0,3)).

Im Fall (1.a.ii) erhalten wir einen Doppelkegel ((2,1,0) oder (1,2,0)) oder zwei sich schneidende Ebenen ((1,1,1)).

Im Fall (1.b.i) erhalten wir die leere Menge ((3,0,0), (2,0,1), (1,0,2) oder (0,0,3)).

Im Fall (1.b.ii) erhalten wir eine Kugel ((0,3,0)) ein einschaliges Rotationshyperboloid ((1,2,0)) einen Zylinder, also das Produkt aus einem Kreis und einer Geraden ((0,2,1)), ein zweischaliges Rotationshyperboloid ((2,1,0)), das Produkt aus einer Hyperbel und einer Geraden ((1,1,1)) oder zwei parallele Ebenen ((0,1,2)).

Der Fall (1.c) liefert wieder die gleichen Flächen wie (1.b).

Im Fall (2) erhalten wir ein Rotationsparaboloid ((2,0,1) oder (0,2,1)), ein hyperbolisches Paraboloid ((1,1,1)), ein Produkt aus einer Parabel und einer Geraden ((1,0,2) oder (0,1,2)), oder eine Ebene ((0,0,3)).

Allgemeine Quadriken im $\mathbb{R}^3$ entstehen aus den obigen durch invertierbare affine Abbildungen. Wenn wir nur affine Isometrien zulassen wollen, können wir die zugrundeliegende symmetrische Bilinearform B nicht auf die Normalform aus dem Sylvesterschen Trägheitssatz 7.67 bringen, aber wegen Folgerung 7.45 immerhin auf Diagonalgestalt. Hieraus folgt zum Beispiel, das ein Ellipsoid immer drei aufeinander senkrechte Hauptachsen hat, also bis auf eine affine Isometrie von der folgenden Form ist:

$$Q = \left\{ (x, y, z) \in \mathbb{R}^3 \;\middle|\; ax^2 + by^2 + cz^2 = 1 \right\} \qquad \text{mit } a,\, b,\, c > 0 \,.$$

Dieser geometrische Sachverhalt ist ein weiterer Grund, das Hauptergebnis aus Abschnitt 7.4 „Hauptachsentransformation“ zu nennen.

Notation

$\in$, 3
$\{\dots\}$, 4
$\emptyset$, 4
$\subset$, 5
$\subsetneq$, 5
$\cap$, 5
$\cup$, 5
$\setminus$, 5
$\times$, 5
$(\dots)$, 5
$\mathcal{P}$, 6
$\{\dots \mid \dots\}$, 6
$F\colon M \to N$, 6
$\Gamma(F)$, 6
Abb , 6
im , 6
F^{-1} , 6
id , 7
$\circ$, 7
$F|_U$, 7
$\mathbb{N}$, 9
$\underline{n}$, 10
$\underline{\mathbb{N}}$, 10
$\#$, 10
$\leq$, 11
$\mathbb{Z}$, 18
$\mathbb{Q}$, 19
$\mathbb{R}$, 20
$\mathbb{R}^n$, 21
$\langle \cdot, \cdot \rangle$, 21
$\| \cdot \|$, 21
$\measuredangle$, 21
i , 23
$\mathbb{C}$, 23
Re , 24
Im , 24
$\bar{\cdot}$, 24
$|\cdot|$, 25
$\times$, 28
$\mathbb{H}$, 30
Aut , 37
$\equiv \mod$, 39
$\mathbb{Z}/n$, 39
$\mathbb{k}^\times$, 42
$a \mid n$, 43
ggT , 44
$\sum_{i=1}^n$, 46
$(a_i)_{i\in I}$, 46
A^I , 46
$\sum_{i\in I}$, 48
$\langle E \rangle$, 48
δ_{ij} , 49
$R^{(I)}$, 50
${}^{(I)}R$, 51
R^n , 51
$\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$, 51
$e_1, \dots, e_n$, 51
${}^nR$, 51
$(r_1, \dots, r_n)$, 51
$\varepsilon_1, \dots, \varepsilon_n$, 51
Hom_R , ${}_R\mathrm{Hom}$, 53
Iso_R , ${}_R\mathrm{Iso}$, 56
End_R , ${}_R\mathrm{End}$, 56
Aut_R , ${}_R\mathrm{Aut}$, 56
M^* , ${}^*M$, 57

ker , 59
$U+V$, 62
$U \oplus V$, 62
$\sum_{i\in I} U_i$, 65
$\bigoplus_{i\in I} U_i$, 65
$\coprod_{i\in I} M_i$, 65
$\prod_{i\in I} M_i$, 65
$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$, 68
$M_{m,n}(R)$, 68
$M_n(R)$, 72
E_n , 72
A^{-1} , 72
$GL(n,R)$, 72
A^t , 77
A^* , 77
$(a_1,\ldots,\widehat{a_i},\ldots,a_n)$, 81
dim , 86
rg , 89
rg_S , rg_Z , 89
P_{ij} , 95
$M_i(k)$, 95
$E_{ij}(k)$, 95
vol , 101
$\Lambda^k M^*$, 102
F^* , 107
det , 108
S_n , 108
P_σ , 108
sign , 108
$\begin{pmatrix} 1 & \cdots & n \\ \sigma(1) & \ldots & \sigma(n) \end{pmatrix}$, 111
$R^\times$, 116
$O(n)$, $SO(n)$, 120
$GL(n,\mathbb{R})^+$, 120
$SL(n,\mathbb{R})$, 120
V_λ , 121
$R[X]$, 127
deg , 127
ev , 127
max , 130
$Q \mid P$, $Q \nmid P$, 132
$\mathrm{ord}_r\, P$, 133
$\chi_A(X)$, $\chi_F(X)$, 137
$\sigma_i(A)$, $\sigma_i(F)$, 137
$\mu_A(X)$, $\mu_F(X)$, 142
kgV , 143
(E) , (a) , 148
$r \mid s$, $r \nmid s$, 149
$\mathrm{Tor}\, M$, 155
$\mathrm{rg}\, M$, 155
$M(P)$, 158
$\mathcal{P}(R)$, 161
$\mu_p(r)$, 163
$\ell(M)$, 164
$\mathrm{Tor}_p\, M$, 173
$H_\lambda V$, 181
$\langle\cdot,\cdot\rangle_{L^2}$, 193
$\langle\cdot,\cdot\rangle_{H^1}$, 194
$\|\cdot\|_g$, 195
df , 207
$\mathrm{grad}\, f$, 208
F^* , 210
$U(n)$, $SU(n)$, 222
$Sp(n)$, 223
d , d_g , 226
$E(n)$, $E(n,\mathbb{k})$, 228
$U(n,\mathbb{k})$, 228
$\rtimes$, 229
$SE(n)$, $SE(n,\mathbb{C})$, 229
$U(p,q;\mathbb{k})$, 233
$O(p,q)$, $SO(p,q)$, 233
$U(p,q)$, $SU(p,q)$, 233
$Sp(p,q)$, 233