LINEARE ALGEBRA

PROF. DR. ANTON FREUND

VORLESUNG IM AKADEMISCHEN JAHR 2025/26
AN DER JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

Inhaltsverzeichnis

Dank. Ich bedanke mich herzlich für Anregungen und Korrekturen von Giacomo Cozzi, Antonia Dammrau, Dr. Jens Jordan, Prof. Dr. Madeleine Jotz, Tanja Neder, Prof. Dr. Silke Neuhaus-Eckhardt, Dr. Nicholas Pischke, Prof. Dr. Oliver Roth und Dr. Dominik Wehr sowie von den Studierenden meiner Vorlesung im Jahr 2025/26.

Anmerkungen gerne an anton.freund@uni-wuerzburg.de. Stand Juli 2026.

1. Vektorräume, lineare Abbildungen und fehlerkorrigierende Codes: grundlegende Begriffe und eine Anwendung

Die lineare Algebra findet in fast allen Bereichen der Mathematik Anwendung. Ein klassisches und wichtiges Anwendungsbeispiel ist die kartesische Geometrie, welche zu Beginn von vielen Lehrbüchern diskutiert wird. Wir betrachten hier ein anderes Beispiel aus einer Sammlung von Matoušek.[1] Gleichzeitig führen wir viele zentrale Grundbegriffe der linearen Algebra ein.

1.1. **Etwas Kodierungstheorie.** Angenommen, wir wollen eine Folge von Bits (Nullen und Einsen) verschicken. Diese zerlegen wir zunächst in Einheiten von – sagen wir – jeweils vier Bits. Wir rechnen damit, dass die Bits unabhängig voneinander mit einer Wahrscheinlichkeit von 95 % korrekt übertragen werden. Die Wahrscheinlichkeit für eine vollständig korrekte Einheit ist dann[2]

$$0{,}95^4 \approx 81\,\%.$$

Können wir es dem Empfänger ermöglichen, Fehler zu korrigieren? Eine Idee wäre, jeden Bit zu verdreifachen. Statt der Einheit 1010 würden wir dann beispielsweise die Bitfolge 111 000 111 000 verschicken. Tritt hierbei ein einzelner Fehler auf – sodass beim Empfänger etwa die Folge 111 001 111 000 ankommt –, so kann dieser die ursprüngliche Einheit dennoch ablesen. Allerdings verschicken wir nun zwölf Bits. Die Wahrscheinlichkeit, dass dabei höchstens ein Fehler auftritt ist mit

$$0{,}95^{12} + 12 \cdot 0{,}95^{11} \cdot 0{,}05 \approx 88\,\%$$

nicht allzu viel höher als oben. Wir werden in diesem Kapitel sehen, dass für die Korrektur eines einzelnen Fehlers nicht zwölf sondern nur sieben Bits pro Einheit nötig sind. Die Wahrscheinlichkeit für ein korrektes Ergebnis steigt dadurch auf

$$0{,}95^7 + 7 \cdot 0{,}95^6 \cdot 0{,}05 \approx 96\,\%,$$

was einen deutlichen und für die Praxis relevanten Fortschritt ausmacht (wobei die geringere Zahl an Bits zusätzlich noch die Übertragung effizienter macht). Zunächst bemühen wir uns um eine mathematische Modellierung der Situation.

Definition 1.1.1. Es sei $\mathbb{B} = \{0, 1\}$ die Menge der Bits oder Booleschen Werte. Für $n \in \mathbb{N}$ definieren wir durch

$$d(s_1 \dots s_n, t_1 \dots t_n) = |\{i : s_i \neq t_i\}|$$

eine Abstandsfunktion d auf $\mathbb{B}^n$.[3]

---

[1] Jiří Matoušek: Thirty-three Miniatures. Mathematical and Algorithmic Applications of Linear Algebra. American Mathematical Society, Providence (RI), 2010. [Da das vorliegende Skript dynamisch wächst, zitiere ich über Fußnoten. Beachten Sie aber bitte (etwa für Ihre Abschlussarbeit), dass in der Mathematik üblicherweise ein Literaturverzeichnis verwendet wird.]

[2] Hier folgt die erste Lücke (siehe die Bemerkung ‚zur Benutzung' auf der ersten Seite).

[3] Wir schreiben $\mathbb{N} = \{0, 1, 2 \dots\}$ für die Menge der natürlichen Zahlen mit der Null. Für Mengen $M_1, \dots, M_n$ bezeichnet $M_1 \times \dots \times M_n$ die Menge aller Tupel $(x_1, \dots, x_n)$ mit Einträgen $x_i \in M_i$ für alle $i = 1, \dots, n$. Dabei heißen zwei Tupel gleich, wenn für alle i jeweils die i-ten Einträge übereinstimmen. Man spricht statt von Tupeln auch von Folgen der Länge n oder für $n = 2$ und $n = 3$ von Paaren bzw. Tripeln. Im Kontext von Bits schreiben wir ausnahmsweise $s_1 \dots s_n$ statt $(s_1, \dots, s_n)$. Gilt $M_1 = \dots = M_n = M$, so schreibt man M^n anstelle von $M_1 \times \dots \times M_n$. Möchte man die Zahl der Grundbegriffe reduzieren, kann man M^n mit der Menge der Funktionen $x : \{1, \dots, n\} \to M$ identifizieren, wobei dann x dem Tupel $(x(1), \dots, x(n))$ entspricht (analog mit unterschiedlichen M_i). Ist M eine endliche Menge, so bezeichnet $|M| \in \mathbb{N}$ ihre Kardinalität, also die Zahl ihrer Elemente. Für eine gegebene Menge M schreiben wir $\{x \in M : P(x)\}$ oder auch

Es gilt etwa $d(1010, 1100) = |\{2, 3\}| = 2$, wobei die Menge $\{2, 3\}$ zum Ausdruck bringt, dass sich die beiden Folgen von Bits gerade an der zweiten und an der dritten Stelle unterscheiden.

Definition 1.1.2. Unter einer Kodierung verstehen wir eine Funktion $c : \mathbb{B}^m \to \mathbb{B}^n$ mit $m, n \in \mathbb{N}$.[4] Wir sagen, eine solche Kodierung korrigiert einen Fehler, wenn es eine Funktion $\bar{c} : \mathbb{B}^n \to \mathbb{B}^m$ mit folgender Eigenschaft gibt: Für alle $s \in \mathbb{B}^m$ und $t \in \mathbb{B}^n$ ist

$$d(c(s), t) \leq 1 \quad \Rightarrow \quad \bar{c}(t) = s.$$

Man nennt $\bar{c}$ dann eine zugehörige Dekodierung.

Beispiel 1.1.3. Die zu Beginn dieses Unterkapitels erwähnte Idee entspricht einer Kodierung $c : \mathbb{B}^4 \to \mathbb{B}^{12}$. Sie korrigiert einen Fehler mit Dekodierung

$$\bar{c}(u_1u_2u_3\, v_1v_2v_3\, w_1w_2w_3\, x_1x_2x_3) = uvwx,$$

wobei u die Mehrheit der u_i darstellt (also etwa $u = 1$ für $u_1u_2u_3 = 101$; analog für v, w, x). Bei dem oben angeführten Beispiel einer Fehlerkorrektur haben wir $s = 1010$ mit $c(s) = 111\,000\,111\,000$ und $t = 111\,000\,111\,010$, also $d(c(s), t) = 1$ und in der Tat $\bar{c}(t) = s$.

Wir nutzen die Gelegenheit, um die folgenden Begriffe einzuführen:

Definition 1.1.4. Eine Funktion[5] $f : M \to N$ zwischen zwei Mengen heißt

- *injektiv*, wenn für alle $x, y \in M$ aus $f(x) = f(y)$ schon $x = y$ folgt,
- *surjektiv*, wenn es zu jedem $z \in N$ ein $x \in M$ gibt mit $f(x) = z$,
- *bijektiv*, wenn sie injektiv und surjektiv ist.

Die Bedingung für die Injektivität ist dazu äquivalent, dass aus $x \neq y$ stets $f(x) \neq f(y)$ folgt,[6] dass f also keine Argumente ‚zusammenwirft'.

Kodierungen sollten stets injektiv sein, da man sonst selbst ohne Fehler verschiedene Nachrichten verwechseln würde. Der folgende Beweis macht dies formal.

Lemma 1.1.5. *Eine Kodierung $c : \mathbb{B}^m \to \mathbb{B}^n$, welche einen Fehler korrigiert, ist stets injektiv.*

Beweis. Wegen $d(c(s), c(s)) = 0 \leq 1$ erfüllt eine Dekodierung $\bar{c}$ stets $\bar{c}(c(s)) = s$. Aus $c(s) = c(t)$ folgt also $s = \bar{c}(c(s)) = \bar{c}(c(t)) = t$. □

Im vorigen Beweis haben wir implizit bereits einen Teil der folgenden Charakterisierung gezeigt, die wir ansonsten zur Übung überlassen.

Lemma 1.1.6. *Eine Funktion $f : M \to N$ mit $M \neq \emptyset$ ist*

$\{x \in M \mid P(x)\}$ für die Menge der Elemente von M, welche die Eigenschaft P erfüllen. Hierbei lassen wir M implizit, wenn der Kontext Klarheit schafft. In der obigen Definition versteht sich etwa $i \in \{1, \ldots, n\}$. [Hier und im Folgenden erklären wir grundlegende Notation in den Fußnoten. Dies erlaubt es uns, die Notation ohne Unterbrechung des Haupttextes einzuführen, wenn sie relevant wird. Die Notation ist dennoch ein wichtiger Lerninhalt.]

[4]Der Begriff ‚Kodierung' ist hier nicht unbedingt Standard und kann in anderen Kontexten anderes bedeuten.

[5]Der Begriff ‚Abbildung' wird synonym verwendet.

[6]Für Aussagen A, B ist die Implikation $A \Rightarrow B$ äquivalent zu ihrer sogenannten Kontraposition $\neg B \Rightarrow \neg A$, wobei $\neg$ die Negation oder Verneinung bezeichnet (jedenfalls in der ‚klassischen' Logik). Um $A \Rightarrow B$ per Kontraposition zu beweisen, kann man also $\neg B$ annehmen und daraus $\neg A$ herleiten. Hingegen folgt aus $A \Rightarrow B$ nicht auch $B \Rightarrow A$ (zu Beginn ein nicht seltener Fehler).

(a) *injektiv genau dann, wenn sie eine sogenannte Linksinverse* $g : N \to M$ *mit* $g \circ f = \mathrm{id}_M$ *besitzt,*[7]
(b) *surjektiv genau dann, wenn sie eine sogenannte Rechtsinverse* $h : N \to M$ *mit* $f \circ h = \mathrm{id}_N$ *besitzt,*
(c) *bijektiv genau dann, wenn sie eine sogenannte Umkehrfunktion* $g : N \to M$ *besitzt, welche zu ihr links- und rechtsinvers ist. Diese ist dann eindeutig.*

Angesichts der Eindeutigkeit in Teil (c) des vorigen Lemmas führen wir f^{-1} als Notation für die Umkehrfunktion einer Bijektion f ein.

Lemma 1.1.7. *Die Komposition von Funktionen ist assoziativ, d.h. man hat*

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

für beliebige Funktionen $f : K \to L$ *und* $g : L \to M$ *sowie* $h : M \to N$.[8]

Beweis. Für beliebiges $x \in K$ erhalten wir

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

aufgrund der Definition der Komposition. □

Wir können die folgende Eigenschaft ableiten. Man beachte hierbei die vertauschte Reihenfolge.

Lemma 1.1.8. *Sind* $f : M \to N$ *und* $g : N \to K$ *bijektiv, so ist es auch* $g \circ f$ *und es gilt* $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Beweis. Wegen Lemma 1.1.6 brauchen wir nur zu prüfen, dass $g \circ f$ und $f^{-1} \circ g^{-1}$ zueinander invers sind. Mit dem vorigen Lemma erhalten wir

$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = f^{-1} \circ (g^{-1} \circ g) \circ f = f^{-1} \circ \mathrm{id}_N \circ f = f^{-1} \circ f = \mathrm{id}_M .$$

Analog erhält man $(g \circ f) \circ (f^{-1} \circ g^{-1}) = \mathrm{id}_K$. □

Als nächstes erarbeiten wir ein Kriterium, mit dem wir entscheiden können, ob eine Kodierung einen Fehler korrigiert.

Definition 1.1.9. Für eine Kodierung $c : \mathbb{B}^m \to \mathbb{B}^n$ mit $m > 0$ definieren wir[9]

$$\#c = \min \{d(c(r), c(s)) : r, s \in \mathbb{B}^m \text{ mit } r \neq s\} .$$

Im folgenden Ergebnis besagen die Implikationen $\Leftarrow$ bzw. $\Rightarrow$, dass wir eine hinreichende und notwendige Bedingung für die Fehlerkorrektur gefunden haben.

Proposition 1.1.10. *Eine Kodierung* $c : \mathbb{B}^m \to \mathbb{B}^n$ *korrigiert einen Fehler genau dann, wenn* $\#c \geq 3$ *gilt.*

---

[7]Die Komposition zweier Funktionen $f : M \to N$ und $h : N \to K$ ist definiert als die Funktion $h \circ f : M \to K$ mit $(h \circ f)(x) = h(f(x))$. Wir schreiben $\mathrm{id}_M : M \to M$ für die Identitätsfunktion mit $\mathrm{id}_M(x) = x$ für alle $x \in M$.

[8]Zwei Abbildungen heißen gleich, wenn sie für jedes Argument gleiche Werte liefern (‚Extensionalität'). Werden Tupel wie in Fußnote 3 vorgeschlagen als Funktionen betrachtet, so erhält man die oben erwähnte Eigenschaft als Instanz der Extensionalität: Man hat $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ genau dann, wenn $x_i = y_i$ für alle i gilt.

[9]Gilt $\emptyset \neq M \subseteq \mathbb{N}$, so sei $\min M$ die kleinste Zahl in M. Hierbei ist $\emptyset$ die leere Menge, welche keine Elemente hat; durch $M \subseteq N$ wird angezeigt, dass jedes Element von M auch in N vorkommt. Für eine Funktion $f : M \to N$ bezeichnet $\{f(x) : x \in M\}$ oder $\{f(x) \,|\, x \in M\}$ die Menge aller Werte von f mit Argumenten aus M. (Dies ist zu unterscheiden von Ausdrücken der Form $\{x \in M : P(x)\}$ wie in Fußnote 3.)

Beweis. Die Implikation $\Rightarrow$ zeigen wir per Kontraposition (siehe Fußnote 6). Wir nehmen dazu an, dass $\#c < 3$ gilt, und müssen herleiten, dass c dann nicht fehlerkorrigierend ist. Es gibt also $r \neq t$ mit $d(c(r), c(t)) \leq 2$, sodass wir ein s finden mit

$$d(c(r), s \leq 1 \quad \text{und} \quad d(s, c(t)) \leq 1.$$

Wäre nun $\bar{c}$ eine Dekodierung, so erhielte man $\bar{c}(s) = r$ sowie $\bar{c}(s) = t$, was aber $r \neq t$ widerspricht.

Für die andere Implikation nehmen wir an, dass $\#c \geq 3$ gilt. Wir konstruieren dann eine Funktion $\bar{c} : \mathbb{B}^n \to \mathbb{B}^m$, indem wir Werte $\bar{c}(t)$ wählen mit

$$d(c(\bar{c}(t)), t) = \min\{d(c(s), t) : s \in \mathbb{B}^m\}.$$

Es bleibt zu zeigen, dass dies eine Dekodierung liefert. Angenommen also, es gilt die Ungleichung $d(c(s), t) \leq 1$. Um $\bar{c}(t) = s$ zu erhalten, zeigen wir, dass $d(c(r), t) > 1$ für alle $r \neq s$ gilt. Hätten wir ein Gegenbeispiel r, so erhielten wir

$$\#c \leq d(c(r), c(s)) \leq d(c(r), t) + d(t, c(s)) \leq 1 + 1 < 3,$$

im Widerspruch zur Annahme. □

Die zweite Ungleichung in der vorletzten Zeile des Beweises (welche man zur Übung verifizieren möge) fällt unter den Begriff der Dreiecksungleichung. Tatsächlich ist d eine sogenannte Metrik auf $\mathbb{B}^n$.

Bemerkung 1.1.11. Der vorige Beweis liefert auch ein Rezept, wie man die Dekodierung $\bar{c}$ bestimmen kann. Dieses ist allerdings umständlich, da es die Kenntnis aller Werte $d(c(s), t)$ mit $s \in \mathbb{B}^m$ und $t \in \mathbb{B}^n$ voraussetzt (und hiervon gibt es 2^{m+n} viele). Wir werden in der Folge eine Kodierung finden, die nicht nur einen Fehler korrigiert, sondern auch eine leicht zu berechnende Dekodierung besitzt.

Zum Abschluss dieses Unterkapitels bestimmen wir eine Schranke für die Länge von fehlerkorrigierenden Kodierungen.

Proposition 1.1.12. *Wenn $c : \mathbb{B}^m \to \mathbb{B}^n$ einen Fehler korrigiert, so gilt*

$$n < 2^{n-m}.$$

Beweis. Für $s \in \mathbb{B}^m$ betrachten wir

$$C_s = \{t \in \mathbb{B}^n : d(c(s), t) \leq 1\}.$$

Es gilt jeweils $|C_s| = 1 + n$ und die Mengen $C_s \subseteq \mathbb{B}^n$ sind paarweise disjunkt, d.h. man hat $C_r \cap C_s = \emptyset$ für $r \neq s$.[10] Da es 2^m viele $s \in \mathbb{B}^m$ gibt, erhält man

$$2^m \cdot (1 + n) \leq 2^n.$$

Das Resultat folgt durch Kürzen (beachte $m \leq n$ wegen Lemma 1.1.5). □

Wenn wir wie zu Beginn dieses Unterkapitels $m = 4$ wählen, so muss für eine fehlerkorrigierende Kodierung also $n \geq 7$ gelten (denn es ist etwa $6 \geq 4 = 2^{6-4}$ sowie andererseits $7 < 8 = 2^{7-4}$). Wie angekündigt werden wir sehen, dass $n = 7$ tatsächlich erreicht werden kann.

[10] Für Mengen M und N bezeichnet $M \cap N$ den Schnitt, welcher aus denjenigen Elementen besteht, die sowohl in M als auch in N vorkommen; hingegen bezeichnet $M \cup N$ die Vereinigung, welche aus denjenigen Elementen besteht, die in M oder N vorkommen (oder in beiden Mengen – das ist bei ‚oder' nie ausgeschlossen). Es gilt also $\{0, 1\} \cap \{1, 2\} = \{1\}$ und $\{0, 1\} \cup \{1, 2\} = \{0, 1, 2\}$.

1.2. **Gruppen, Körper und Vektorräume.** In diesem Unterkapitel führen wir einige zentrale Begriffe der linearen Algebra ein, wobei wir die Strukturen $\mathbb{B}^n$ aus dem vorigen Abschnitt als laufendes Beispiel im Blick behalten.

Wir erinnern uns, dass wir eine fehlerkorrigierende Kodierung $c : \mathbb{B}^4 \to \mathbb{B}^7$ finden wollen. Man mag hoffen, dass die folgende Struktur auf $\mathbb{B}$ dabei hilfreich ist.

Definition 1.2.1. Auf der Menge $\mathbb{B}$ der Booleschen Werte führen wir durch die folgenden Setzungen eine Addition und eine Multiplikation ein:

$$0+0=0, \quad 0+1=1, \quad 1+0=1, \quad 1+1=0,$$
$$0\cdot 0=0, \quad 0\cdot 1=0, \quad 1\cdot 0=0, \quad 1\cdot 1=1.$$

Wir diskutieren nun einige allgemeine Begriffe, durch welche ein wichtiger Aspekt der Struktur von $(\mathbb{B}, +, \cdot)$ erfasst wird.

Definition 1.2.2. Eine Gruppe besteht aus einer Menge G und einer Verknüpfung[11] $\cdot : G^2 \to G$, für die gilt:

(G1) Man hat $f \cdot (g \cdot h) = (f \cdot g) \cdot h$ für alle $f, g, h \in G$.
(G2) Es gibt ein $e \in G$ (‚neutrales Element') mit folgenden Eigenschaften:
  (i) Man hat $e \cdot g = g$ für alle $g \in G$.
  (ii) Zu jedem $g \in G$ gibt es ein $\bar{g} \in G$ (‚inverses Element') mit $\bar{g} \cdot g = e$.

Gilt zudem $g \cdot h = h \cdot g$ für alle $g, h \in G$, so heißt die Gruppe kommutativ oder auch abelsch.[12]

Die Teile (a) und (c) des folgenden Lemmas zeigen, dass eine natürliche Verstärkung von (G2) tatsächlich redundant wäre (selbst in nicht-abelschen Gruppen).

Lemma 1.2.3. *Sei $(G, \cdot)$ eine Gruppe.*

(a) Ist e neutral, so gilt auch $g \cdot e = g$ für alle $g \in G$.
(b) Es gibt genau ein neutrales Element.
(c) Gilt $\bar{g} \cdot g = e$ für das neutrale Element e, so gilt auch $g \cdot \bar{g} = e$.
(d) Zu jedem $g \in G$ gibt es genau ein inverses Element.

Beweis. Es erweist sich als zielführend, die Aussagen in einer anderen Reihenfolge zu zeigen.

(c) Da wir (b) noch nicht bewiesen haben, können wir zunächst nur voraussetzen, dass e *ein* neutrales Element ist. Wegen (G2) erhalten wir zu $\bar{g}$ wiederum ein (links-) inverses Element $\bar{\bar{g}}$ mit $\bar{\bar{g}} \cdot \bar{g} = e$. Unter Verwendung von (G1) rechnen wir

$$g \cdot \bar{g} = e \cdot (g \cdot \bar{g}) = (\bar{\bar{g}} \cdot \bar{g}) \cdot (g \cdot \bar{g}) = \bar{\bar{g}} \cdot ((\bar{g} \cdot g) \cdot \bar{g}) = \bar{\bar{g}} \cdot (e \cdot \bar{g}) = \bar{\bar{g}} \cdot \bar{g} = e.$$

(a) Zu gegebenem $g \in G$ erhalten wir wegen (G2) und dem bereits bewiesenen Teil (c) ein Element $\bar{g}$ mit $\bar{g} \cdot g = e = g \cdot \bar{g}$. Es ergibt sich

$$g \cdot e = g \cdot (\bar{g} \cdot g) = (g \cdot \bar{g}) \cdot g = e \cdot g = g.$$

(b) Angenommen, e und e' sind beide neutral. Teil (a) liefert $e' = e' \cdot e$. Mit (G2) für e' erhalten wir andererseits $e' \cdot e = e$, also insgesamt $e' = e$.

(d) Seien $\bar{g}$ und $\bar{g}'$ invers zu g. Es ergibt sich

$$\bar{g}' = e \cdot \bar{g}' = (\bar{g} \cdot g) \cdot \bar{g}' = \bar{g} \cdot (g \cdot \bar{g}') = \bar{g} \cdot e = \bar{g},$$

wobei wir das Ergebnis von (a) und (c) verwendet haben. □

[11] Das ist wieder ein anderes Wort für Funktion oder Abbildung. Allerdings sind Verknüpfungen meist zweistellig und man schreibst sie oft infix, d.h. man schreibt $g \cdot h$ statt $\cdot(g, h)$.
[12] Die Bezeichnung bezieht sich auf Niels Henrik Abel.

In Anbetracht der obigen Eindeutigkeitsaussagen schreiben wir oft e für das neutrale Element einer Gruppe sowie g^{-1} für das Inverse von g (und $\cdot$ wird manchmal weggelassen). In abelschen Gruppen schreibt man oft $+$ statt $\cdot$ sowie 0 und $-g$ anstelle von e und g^{-1}.

Beispiel 1.2.4. (a) Auf der Einermenge $\{e\}$ wird durch die (einzig mögliche) Verknüpfung $e \cdot e = e$ eine abelsche Gruppe definiert (‚triviale Gruppe').

(b) Die Struktur $(\mathbb{B}, +)$ ist eine abelsche Gruppe (wie man prüfen möge). Hingegen ist $(\mathbb{B}, \cdot)$ keine Gruppe (weil hier zwar 1 neutral ist aber 0 kein Inverses hat).

(c) Für jede Menge M bildet die Menge $\mathrm{Sym}(M)$ der bijektiven Funktionen $f : M \to M$ mit der Komposition als Verknüpfung eine Gruppe. Speziell kann man $S_3 := \mathrm{Sym}(\{1, 2, 3\})$ als Gruppe der Drehungen und Spiegelungen eines gleichseitigen Dreiecks veranschaulichen. Diese ist nicht abelsch, d.h. man kann Drehungen und Spiegelungen im Allgemeinen nicht vertauschen: Man zeige hierfür etwa $g \circ f \neq f \circ g$ mit

$$f(1) = 2, \quad f(2) = 3, \quad f(3) = 1, \qquad \text{(Drehung)}$$
$$g(1) = 1, \quad g(2) = 3, \quad g(3) = 2. \qquad \text{(Spiegelung)}$$

(d) Die ganzen Zahlen $\mathbb{Z}$ bilden mit der üblichen Addition eine abelsche Gruppe (und gleiches gilt für die rationalen und reellen Zahlen $\mathbb{Q}$ bzw. $\mathbb{R}$ aber nicht für die natürlichen Zahlen). Mit der üblichen Multiplikation bildet zwar nicht die ganze Menge $\mathbb{Q}$ eine Gruppe, wohl aber die Menge $\mathbb{Q}\backslash\{0\}$ (welche unter $\cdot$ abgeschlossen ist – aus $p, q \neq 0$ folgt hier $p \cdot q \neq 0$).[13]

Die folgenden Begriffe spiegeln wieder, dass wir in Definition 1.2.1 (wie auch in den Zahlbereichen $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$) zwei Verknüpfungen haben.

Definition 1.2.5. Ein Ring besteht aus einer Menge R und zwei Verknüpfungen $+, \cdot : R^2 \to R$, für die gilt:

(R1) Die Struktur $(R, +)$ ist eine abelsche Gruppe.
(R2) Die Verknüpfung $\cdot$ ist assoziativ.
(R3) Für alle $p, q, r \in R$ gelten die Distributivgesetze

$$p \cdot (q + r) = p \cdot q + p \cdot r \quad \text{und} \quad (p + q) \cdot r = p \cdot r + q \cdot r,$$

wobei wir gemäß ‚Punkt vor Strich' klammern.

Man spricht von einem Ring mit Eins, wenn es ein (dann eindeutig bestimmtes) Element $1 \in R$ gibt mit $1 \cdot p = p = p \cdot 1$ für alle $p \in R$.[14] Weiter nennt man den Ring kommutativ, wenn $p \cdot q = q \cdot p$ für alle $p, q \in R$ gilt.

Die folgende in konkreten Fällen bekannte Eigenschaft kann auch allgemein nachgewiesen werden.

Lemma 1.2.6. *In jedem Ring R gilt* $0 \cdot p = 0 = p \cdot 0$.[15]

[13] Für Mengen M, N ist $M\backslash N := \{x \in M \mid x \notin N\}$, also etwa $\{0, 1, 2\}\backslash\{2, 3\} = \{0, 1\}$.

[14] In manchen Quellen wird diese Eigenschaft für alle Ringe vorausgesetzt.

[15] Wenn nicht anders angegeben, schreiben wir die beiden Verknüpfungen in einem Ring stets als $+$ und $\cdot$, wobei wir von Addition und Multiplikation sprechen. Die neutralen und inversen Elemente bezüglich dieser beiden Verknüpfungen werden – soweit vorhanden – mit 0 bzw. 1 und $-a$ bzw. a^{-1} bezeichnet. Weiter schreibt man $a - b$ für $a + (-b)$ sowie a/b für $a \cdot b^{-1}$.

Beweis. Wegen der Neutralität der Null und dem Distributivgesetz erhält man

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

Hier können wir wie folgt kürzen: Weil $(R, +)$ eine Gruppe ist, haben wir ein inverses Element $q = -0 \cdot p$. Mit diesem rechnen wir

$$0 = 0 \cdot p + q = 0 \cdot p + 0 \cdot p + q = 0 \cdot p + 0 = 0 \cdot p.$$

Analog erhält man $0 = p \cdot 0$. □

Wir werden uns besonders für Ringe mit der folgenden Eigenschaft interessieren.

Definition 1.2.7. Ein Körper ist ein kommutativer Ring $K \neq \{0\}$ mit Eins, in dem es für jedes $a \in K \backslash \{0\}$ ein (dann auch hier eindeutig bestimmtes) multiplikativ Inverses a^{-1} gibt, welches also $a^{-1} \cdot a = 1$ erfüllt.

Wir merken an, dass $K = \{0\}$ gemäß unserer Definition als Ring mit Eins aber nicht als Körper zählt. In einem Körper gilt immer $0 \neq 1$ (weil die Null wegen Lemma 1.2.6 nicht das neutrale Element der Multiplikation sein kann).

Lemma 1.2.8. *In jedem Körper folgt aus* $a, b \neq 0$ *schon* $a \cdot b \neq 0$.[16]

Beweis. Wir argumentieren per Kontraposition und nehmen $a \cdot b = 0$ aber $a \neq 0$ an. Letzteres liefert das Inverse a^{-1} und damit

$$b = a^{-1} \cdot a \cdot b = a^{-1} \cdot 0 = 0,$$

wobei wir Lemma 1.2.6 verwendet haben. □

Wir sehen nun, dass $(K \backslash \{0\}, \cdot)$ für jeden Körper K eine abelsche Gruppe ist (wobei das vorige Lemma die Abgeschlossenheit unter der Multiplikation sichert).

Beispiel 1.2.9. (a) Die Struktur $(\mathbb{B}, +, \cdot)$ ist ein Körper[17] (wie man ausgehend von Beispiel 1.2.4 prüfen möge).

(b) Die Zahlbereiche $\mathbb{Q}$ und $\mathbb{R}$ sind Körper, aber $\mathbb{Z}$ ist nur ein Integritätsring.

(c) Wir betrachten die Menge der Funktionen $f : \mathbb{R} \to \mathbb{R}$ zusammen mit den durch $(f+g)(x) := f(x)+g(x)$ und $(f \cdot g)(x) := f(x) \cdot g(x)$ definierten Abbildungen. Diese ist ein kommutativer Ring mit Eins aber kein Körper (nämlich nicht einmal ein Integritätsring).

Die folgende Konstruktion sollte im Kontext des euklidischen Raumes $\mathbb{R}^3$ bekannt sein. Weiter ist sie im Fall $K = \mathbb{B}$ für die Kodierungen aus dem vorigen Unterkapitel relevant.

Definition 1.2.10. Für einen Körper K und $n \in \mathbb{N}$ definieren wir Operationen[18] $+ : K^n \times K^n \to K^n$ und $\cdot : K \times K^n \to K^n$ (‚Skalarmultiplikation') durch[19]

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) := (x_1 + y_1, \dots, x_n + y_n),$$
$$\lambda \cdot (x_1, \dots, x_n) := (\lambda \cdot x_1, \dots, \lambda \cdot x_n).$$

[16] Hat ein kommutativer Ring mit Eins diese Eigenschaft, so nennt man ihn Integritätsring.

[17] Dieser wird meist mit $\mathbb{F}_2$ bezeichnet.

[18] Dies ist noch einmal synonym zu ‚Verknüpfung'.

[19] Es mag hier zunächst irritieren, dass $+$ eine Operation auf K und eine auf K^n bezeichnet (analog für die Skalarmultiplikationn). Man muss diese sorgfältig unterscheiden – wofür man zur Übung gerne unterschiedliche Symbole wie $+$ und $\oplus$ verwenden kann. Nach einiger Zeit wird man aber feststellen, dass der Kontext eine Verwechslung verhindert, während man das zusätzliche Symbol $\oplus$ gerne für andere Zwecke zur Verfügung hat.

Wir merken an, dass $n = 0$ erlaubt ist. In diesem Fall hat man $K^n = \{0\}$, wobei das Element 0 dem leeren Tupel entspricht.

Die Skalarmultiplikation hat eine vertraute geometische Interpretation als Streckung (etwa im $\mathbb{R}^3$). Dennoch kann man natürlich auch eine Multiplikation mit Definitionsbereich $K^n \times K^n$ betrachten, was K^n zu einem kommutativen Ring mit Eins macht (vgl. Teil (c) von Beispiel 1.2.9). Es wird sich aber herausstellen, dass der folgende Begriff für die lineare Algebra von besonders zentraler Bedeutung ist.

Definition 1.2.11. Ein Vektorraum über einem Körper K (kurz K-Vektorraum) besteht aus einer abelschen Gruppe V und einer Operation $\cdot : K \times V \to V$, die mit den Verknüpfungen in K und V (vgl. Fußnote 19) in folgendem Sinn verträglich ist: Für alle $\lambda, \mu \in K$ und $v, w \in V$ gilt

$$(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v, \qquad \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w,$$
$$\lambda \cdot (\mu \cdot v) = (\lambda \cdot \mu) \cdot v, \qquad 1 \cdot v = v.$$

Analog zu den Lemmata 1.2.6 und 1.2.8 erhält man:

Lemma 1.2.12. *Für jeden K-Vektorraum V und alle $\lambda \in K$ sowie $v \in V$ gilt*

$$\lambda \cdot v = 0 \quad \Leftrightarrow \quad \lambda = 0 \text{ oder } v = 0.$$

Weiter hat man $(-1) \cdot v = -v$.

Das folgende Ergebnis ist äußerst wichtig, hat aber einen elementaren Beweis, den wir zur Übung überlassen.

Proposition 1.2.13. *Ist K ein Körper und $n \in \mathbb{N}$, so ist K^n ein K-Vektorraum.*

Die Vektorräume K^n sind nicht nur wichtige Beispiele, sondern schöpfen den allgemeinen Begriff des Vektorraums schon in erstaunlichem Maße aus. Dies werden wir später mithilfe des Begriffs der Basis zeigen. Dabei wird auch klar werden, dass sich das folgende Beispiel grundsätzlich von den K^n unterscheidet.

Beispiel 1.2.14. Die Menge der Funktionen $f : \mathbb{R} \to \mathbb{R}$ bildet mit der Addition aus Beispiel 1.2.9(c) und der durch $(\lambda \cdot f)(x) := \lambda \cdot f(x)$ für $\lambda \in \mathbb{R}$ gegebenen Skalarmultiplikation einen $\mathbb{R}$-Vektorraum.

1.3. **Lineare Abbildungen und Matrizen.** In diesem Abschnitt führen wir eine Klasse von Abbildungen ein, welche die Struktur von Vektorräumen respektieren, nämlich die linearen Abbildungen. Wir werden sehen, dass sich diese im Fall der Vektorräume K^n durch endliche Schemata von Körperelementen repräsentieren lassen – sogenannte Matrizen. Dies wird uns auch gute Kandidaten für die in Unterkapitel 1.1 diskutierten Kodierungen geben.

Definition 1.3.1. Seien V, W Vektorräume über einem gemeinsamen Körper K. Eine Abbildung $\varphi : V \to W$ heißt linear (oder auch ein Homomorphismus von K-Vektorräumen), wenn für alle $\lambda \in K$ und alle $v, w \in V$ gilt:

$$\varphi(\lambda \cdot v) = \lambda \cdot \varphi(v) \quad \text{und} \quad \varphi(v + w) = \varphi(v) + \varphi(w).$$

Wir schreiben $\mathrm{Hom}(V, W)$ für die Menge aller linearen Abbildungen von V nach W.

Man möge sich als Übung davon überzeugen, dass die beiden gegebenen Gleichungen zusammen äquivalent sind zu der einzelnen Gleichung

$$\varphi(\lambda \cdot v + \mu \cdot w) = \lambda \cdot \varphi(v) + \mu \cdot \varphi(w) \quad \text{für } \lambda, \mu \in K \text{ und } v, w \in V.$$

Hilfreich ist dabei die erste der folgenden Gleichungen oder mehr noch ihr Beweis.

Lemma 1.3.2. *Für jede lineare Abbildung $\varphi : V \to W$ und alle $v \in V$ gilt*

$$\varphi(0) = 0 \quad \textit{und} \quad \varphi(-v) = -\varphi(v).$$

Beweis. Die beiden Gleichungen ergeben sich mittels Lemma 1.2.12 aus

$$\varphi(0) = \varphi(0 \cdot 0) = 0 \cdot \varphi(0) = 0$$

und aus $\varphi(-v) = \varphi((-1) \cdot v) = (-1) \cdot \varphi(v) = -\varphi(v)$. □

Um eine Beziehung zu den Kodierungen aus Teilkapitel 1.1 herzustellen, zeigen wir zunächst die folgende Eigenschaft.

Lemma 1.3.3. *Die Abstandsfunktion auf $\mathbb{B}^n$ (siehe Definition 1.1.1) erfüllt*

$$d(r, s) = d(r + t, s + t) \quad \textit{für alle } r, s, t \in \mathbb{B}^n.$$

Beweis. Die Abstandsfunktion zählt, an wie vielen Positionen zwei Folgen von Bits sich unterscheiden. Da die Addition komponentenweise definiert ist (siehe Definition 1.2.10), lässt sich die Behauptung reduzieren auf

$$r_i = s_i \quad \Leftrightarrow \quad r_i + t_i = s_i + t_i \quad \text{für alle } r_i, s_i, t_i \in \mathbb{B}.$$

Die Richtung von links nach rechts gilt für jede Verknüpfung oder Funktion. Die Gegenrichtung ergibt sich, weil wir in Gruppen kürzen können (so wie im Beweis von Lemma 1.2.6). □

Wir sehen nun, dass sich das entscheidende Kriterium aus Proposition 1.1.10 im linearen Fall wie folgt vereinfacht.

Proposition 1.3.4. *Für jede lineare Kodierung $c : \mathbb{B}^m \to \mathbb{B}^n$ (also jede lineare Abbildung zwischen $\mathbb{B}$-Vektorräumen) gilt*

$$\#c = \min\big\{d(0, c(t)) : t \in \mathbb{B}^m \backslash \{0\}\big\}.$$

Beweis. Wir wählen zunächst $t \neq 0$ so, dass $d(0, c(t))$ minimal wird. Wegen des Minimums in Definition 1.1.9 und unter Verwendung von Lemma 1.3.2 erhalten wir dann

$$\#c \leq d(c(0), c(t)) = d(0, c(t)) = \min\big\{d(0, c(t)) : t \in \mathbb{B}^m \backslash \{0\}\big\}.$$

Für die entgegengesetzte Ungleichung wählen wir $r \neq s$ mit $d(c(r), c(s)) = \#c$. Wegen $s - r \neq 0$ und mit dem vorigen Lemma gilt dann

$$\begin{aligned}\min\big\{d(0, c(t)) : t \in \mathbb{B}^m \backslash \{0\}\big\} \leq d(0, c(s - r)) &= d(0 + c(r), c(s - r) + c(r)) = \\ &= d(c(r), c(s - r + r)) = d(c(r), c(s)) = \#c.\end{aligned}$$

Zusammen folgt die gewünschte Gleichung. □

Bevor wir die Anwendung auf Kodierungen im folgenden Unterkapitel abschließen, beschäftigen wir uns noch damit, wie lineare Abbildungen konkret dargestellt werden können. Wie wir später sehen, liefert die folgende Konstruktion eine sogenannte Basis.

Definition 1.3.5. Im Vektorraum K^n über einem Körper K sind die Einheitsvektoren $e_1, \dots, e_n$ definiert durch

$$e_i = (x_1, \dots, x_n) \quad \text{mit} \quad x_j = \begin{cases} 1 & \text{wenn } i = j, \\ 0 & \text{sonst.} \end{cases}$$

Um beliebige Vektoren als Linearkombinationen von Einheitsvektoren zu schreiben, führen wir die folgende Notation ein.

Definition 1.3.6. Wir betrachten eine Menge M, auf der eine Addition erklärt ist. Für $j, k \in \mathbb{Z}$ und $a_j, \ldots, a_k \in M$ schreiben wir

$$\sum_{i=j}^{k} a_i := a_j + a_{j+1} + \ldots + a_k,$$

wobei die Summe den Wert Null haben soll, falls $j > k$ gilt.

Das folgende Lemma gibt uns die Gelegenheit, eine äußerst wichtige Beweismethode einzusetzen, nämlich die (‚vollständige') Induktion.[20]

Lemma 1.3.7. *Für beliebige $\lambda_1, \ldots, \lambda_n \in K$ gilt im Vektorraum K^n die Gleichung*

$$(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i \cdot e_i.$$

Beweis. Per Induktion über j zeigen wir: Gilt $\lambda_i = 0$ für alle $i > j$, so ist

$$(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{j} \lambda_i \cdot e_i.$$

Für $j = n$ folgt dann die Behauptung.

Induktionsanfang: Hier ist $j = 0$ und damit nach Annahme $\lambda_i = 0$ für alle i, also $(\lambda_1, \ldots, \lambda_n) = 0$. Wegen $j < 1$ gilt gemäß der Konvention in Definition 1.3.6 auch $\sum_{i=1}^{j} \lambda_i \cdot e_i = 0$.

Induktionsannahme: Wir nehmen an, dass die obige Gleichung für ein beliebig gewähltes j gilt.

Induktionsschritt: Unser Ziel ist, die obige Gleichung mit $j+1$ an der Stelle von j herzuleiten. Gemäß der Voraussetzung in der Behauptung können wir dabei $\lambda_i = 0$ für $i > j + 1$ annehmen. Wir erhalten

$$(\lambda_1, \ldots, \lambda_n) = (\lambda_1, \ldots, \lambda_j, 0, \ldots, 0) + \lambda_{j+1} \cdot e_{j+1} = \\ \overset{\text{IA}}{=} \left(\sum_{i=1}^{j} \lambda_i \cdot e_i \right) + \lambda_{j+1} \cdot e_{j+1} = \sum_{i=1}^{j+1} \lambda_i \cdot e_i,$$

wobei die Anwendung der Induktionsannahme mit ‚IA' gekennzeichnet ist. □

[20]Das Beweisprinzip der Induktion kann man einsetzen, um eine Aussage der Form ‚für alle $n \in \mathbb{N}$ gilt $P(n)$' zu zeigen. Hierbei drückt $P(n)$ aus, dass der Zahl n eine gegebene Eigenschaft P zukommt. Gemäß dem Induktionsprinzip genügt es, wenn wir einerseits $P(0)$ (Induktionsanfang) sowie andererseits die Aussage ‚für alle $n \in \mathbb{N}$ mit $P(n)$ gilt auch $P(n+1)$' zeigen. Für letztere lässt man $n \in \mathbb{N}$ beliebig sein und setzt $P(n)$ voraus (Induktionsannahme oder -hypothese). Hieraus muss man dann $P(n+1)$ herleiten (Induktionsschritt oder -schluss). Da sich jede natürliche Zahl in endlich vielen Schritten von der Null aus erreichen lässt, ist die Eigenschaft P damit in der Tat für alle natürlichen Zahlen gesichert. Zu Anfang des Studiums empfiehlt es sich, Induktionsbeweise immer nach dem Schema aus dem Beweis von Lemma 1.3.7 aufzuschreiben. Später mag man zu kompakteren Darstellungen wie im Beweis von Lemma 1.3.8 übergehen. Weiter wird man verschiedene (letztlich äquivalente) Varianten des Induktionsprinzips verwenden: So kann man etwa den Induktionsanfang weglassen, wenn man die Induktionsannahme durch die Aussage ‚für alle $m < n$ gilt $P(m)$' ersetzt und daraus dann im Induktionsschritt $P(n)$ herleitet. (Hierbei liefert die Induktionsannahme für $n = 0$ keinerlei Information, sodass man $P(0)$ letztlich doch wieder ohne Annahme herleiten muss – wie sonst im Induktionsanfang.)

Die definierende Eigenschaft linearer Abbildungen verallgemeinert sich wie folgt auf mehr als zwei Summanden.

Lemma 1.3.8. *Ist $\varphi : V \to W$ eine lineare Abbildung zwischen K-Vektorräumen, so gilt für alle $\lambda_i \in K$ und $v_i \in V$ die Gleichung*

$$\varphi\left(\sum_{i=1}^{k} \lambda_i \cdot v_i\right) = \sum_{i=1}^{k} \lambda_i \cdot \varphi(v_i).$$

Beweis. Wir argumentieren per Induktion über k. Im Induktionsanfang betrachten wir $k = 0$ (wobei $k = 1$ auch eine natürliche Wahl wäre). Hier ist also $1 > k$, sodass die gewünschte Gleichnung wegen der Konvention aus Definition 1.3.6 auf $\varphi(0) = 0$ hinausläuft. Für den Induktionsschritt rechnen wir

$$\begin{aligned}\varphi\left(\sum_{i=1}^{k+1} \lambda_i \cdot v_i\right) &= \varphi\left(\left(\sum_{i=1}^{k} \lambda_i \cdot v_i\right) + \lambda_{k+1} \cdot v_{k+1}\right)\\ &= \varphi\left(\sum_{i=1}^{k} \lambda_i \cdot v_i\right) + \lambda_{k+1} \cdot \varphi(v_{k+1})\\ &\overset{\text{IA}}{=} \left(\sum_{i=1}^{k} \lambda_i \cdot \varphi(v_i)\right) + \lambda_{k+1} \cdot \varphi(v_{k+1}) = \sum_{i=1}^{k+1} \lambda_i \cdot \varphi(v_i),\end{aligned}$$

wobei die zweite Gleichung wegen der Linearität von φ gilt. □

Wir können folgern, dass eine lineare Abbildung bereits durch ihre Werte auf den Einheitsvektoren festgelegt ist:

Korollar 1.3.9. *Seien $\varphi, \psi : K^n \to V$ lineare Abbildungen in einen Vektorraum V über einem Körper K. Gilt $\varphi(e_i) = \psi(e_i)$ für alle $i = 1, \dots, n$, so hat man $\varphi = \psi$.*

Beweis. Wir betrachten einen beliebigen Vektor $v \in K^n$ und schreiben ihn gemäß Lemma 1.3.7 in der Form $v = \sum_{i=1}^{n} \lambda_i \cdot e_i$. Mit dem vorigen Lemma folgt

$$\varphi(v) = \sum_{i=1}^{n} \lambda_i \cdot \varphi(e_i) = \sum_{i=1}^{n} \lambda_i \cdot \psi(e_i) = \psi(v),$$

wie gewünscht. □

Für eine lineare Abbildung $\varphi : K^n \to K^m$ liefert Lemma 1.3.7 auch für die Bilder der Einheitsvektoren $e_1, \dots, e_n$ eine Darstellung der Form

$$\varphi(e_j) = \sum_{i=1}^{m} \lambda_{ij} \cdot e_i.$$

Aus dem obigen Korollar folgt, dass die Abbildung φ vollständig durch die Koeffizienten λ_{ij} für $i = 1, \dots, m$ und $j = 1, \dots, n$ beschrieben wird. Wegen Lemma 1.3.7 sind diese Koeffizienten eindeutig. Man hält sie in der folgenden Form fest.

Definition 1.3.10. Eine $m \times n$-Matrix über einem Körper K ist ein Schema[21] der Form

$$(\lambda_{ij})_{1\leq i\leq m, 1\leq j\leq n} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & x_{m2} & \dots & \lambda_{mn} \end{pmatrix}$$

mit $\lambda_{ij} \in K$. Man schreibt $K^{m\times n}$ für die Menge aller $m \times n$-Matrizen über K.

Um konkreter zu beschreiben, wie eine lineare Abbildung durch die erwähnten Koeffizienten bestimmt wird, benötigen wir den folgenden Hilfssatz, den wir als Übung in Induktion überlassen.[22]

Lemma 1.3.11. *Für das Summenzeichen gelten in Vektorräumen die verallgemeinerten Distributivgesetze*

$$\sum_{i=1}^{n} \lambda \cdot v_i = \lambda \cdot \sum_{i=1}^{n} v_i \quad \textit{und} \quad \sum_{i=1}^{n} \lambda_i \cdot v = \left(\sum_{i=1}^{n} \lambda_i\right) \cdot v.$$

Weiter gilt das verallgemeinerte Kommutativgesetz

$$\sum_{i=1}^{m}\sum_{j=1}^{n} v_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} v_{ij}.$$

Ist φ wie im Abschnitt vor Definition 1.3.10 mit den Koeffizienten λ_{ij} assoziiert, so erhält man nun mit Lemma 1.3.7 die Beziehung

$$\begin{aligned} \varphi\big((x_1,\dots,x_n)\big) &= \varphi\left(\sum_{j=1}^{n} x_j \cdot e_j\right) = \sum_{j=1}^{n} x_j \cdot \varphi(e_j) = \sum_{j=1}^{n} x_j \cdot \sum_{i=1}^{m} \lambda_{ij} \cdot e_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \lambda_{ij} \cdot x_j\right) \cdot e_i = \left(\sum_{j=1}^{n} \lambda_{1j} \cdot x_j, \dots, \sum_{j=1}^{n} \lambda_{mj} \cdot x_j\right) \in K^m. \end{aligned}$$

Wir haben Elemente von K^n bisher als Tupel (‚Zeilenvektoren') geschrieben. Im Kontext der Matrixmultiplikation schreibt man sie aber meist als Spaltenvektoren. Möchte man zwischen beiden Schreibweisen wechseln (etwa um Platz zu sparen), so kennzeichnet man dies durch ein Superskript[23] wie in

$$(x_1,\dots,x_n)^{\mathsf{T}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Die zuletzt betrachtete Beziehung für φ motiviert dann die Definition der Multiplikation zwischen Matrix und Vektor.

---

[21] Um die Zahl der Grundbegriffe zu reduzieren, kann man solch ein Schema formal als Tupel betrachten – etwa als Tupel aus den n Spaltenvektoren, wobei die Einträge dieses Tupels dann selbst wieder Tupel mit je m Einträgen aus K wären, sodass man also $K^{m\times n} = (K^m)^n$ hätte.

[22] Gemäß ‚Punkt vor Strich' schreiben wir dabei etwa $\sum_{i=1}^{n} \lambda \cdot v_i$ für $\sum_{i=1}^{n} (\lambda \cdot v_i)$.

[23] Es handelt sich hier um einen Spezialfall der Transposition von Matrizen, die wir später kennenlernen werden.

Definition 1.3.12. Wir setzen

$$\begin{pmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{m1} & \dots & \lambda_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \lambda_{1j} \cdot x_j \\ \vdots \\ \sum_{j=1}^n \lambda_{mj} \cdot x_j \end{pmatrix}.$$

Um die allgemeine Definition zu verstehen, ist es hilfreich, den Fall $m = n = 3$ zu betrachten:

Beispiel 1.3.13. Es gilt

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 \\ b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 \\ c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_3 \end{pmatrix}.$$

Wir haben angedeutet, wie sich ausgehend von einer linearen Abbildung eine zugehörige Matrix definieren lässt. Nun gehen wir in die andere Richtung:

Definition 1.3.14. Mit jeder $m \times n$-Matrix M über einem Körper K assoziieren wir eine Abbildung $\mathrm{Abb}(M) : K^n \to K^m$, die gegeben ist durch

$$\mathrm{Abb}(M)(x) := M \cdot x.$$

Wir prüfen die folgende Eigenschaft.

Proposition 1.3.15. *Für jede Matrix $M \in K^{m\times n}$ ist $\mathrm{Abb}(M) : K^n \to K^m$ eine lineare Abbildung.*

Beweis. Wir schreiben $M = (\lambda_{ij})$ sowie $x = (x_j)$ und $y = (y_j)$. Dann ist der i-te Eintrag von $\mathrm{Abb}(M)(\mu \cdot x) = M \cdot (\mu \cdot x)$ gegeben durch

$$\sum_{j=1}^n \lambda_{ij} \cdot \mu \cdot x_j = \mu \cdot \sum_{j=1}^n \lambda_{ij} \cdot x_j.$$

Dies entspricht dem μ-fachen des i-ten Eintrags von $\mathrm{Abb}(M)(x) = M \cdot x$. Weiter hat $\mathrm{Abb}(M)(x + y) = M \cdot (x + y)$ den i-ten Eintrag

$$\sum_{j=1}^n \lambda_{ij} \cdot (x_j + y_j) = \sum_{j=1}^n (\lambda_{ij} \cdot x_j + \lambda_{ij} \cdot y_j) = \sum_{j=1}^n \lambda_{ij} \cdot x_j + \sum_{j=1}^n \lambda_{ij} \cdot y_j,$$

was auch der i-te Eintrag von $\mathrm{Abb}(M)(x) + \mathrm{Abb}(M)(y) = M \cdot x + M \cdot y$ ist. Bei der letzten Rechnung haben wir einen Spezialfall des verallgemeinerten Kommutativgesetzes verwendet, in dem eine der Summen nur zwei Summanden hat. □

Das folgende Ergebnis drückt auf mathematisch präzise Weise aus, dass Matrizen und lineare Abbildungen zwei Seiten derselben Medaille sind. Wir rufen in Erinnerung, dass $\mathrm{Hom}(V, W)$ die Menge der linearen Abbildungen zwischen zwei K-Vektorräumen V und W bezeichnet.

Theorem 1.3.16. *Für jeden Körper K ist die Funktion*

$$\mathrm{Abb} : K^{m\times n} \to \mathrm{Hom}(K^n, K^m)$$

eine Bijektion.[24]

[24]Man beachte, wie m und n hier ‚vertauscht' werden, was auf die Schreibkonventionen für Matrizen zurückzuführen ist.

Beweis. Wie im Absatz vor Definition 1.3.10 gesehen, können wir eine Funktion

$$\mathrm{Mat} : \mathrm{Hom}(K^n, K^m) \to K^{m\times n}$$

definieren durch

$$\mathrm{Mat}(\varphi) = (\lambda_{ij}^{\varphi})_{1\le i\le m, 1\le j\le n} \quad \text{für} \quad \varphi(e_j) = \sum_{i=1}^{m} \lambda_{ij}^{\varphi} \cdot e_i = \begin{pmatrix} \lambda_{1j}^{\varphi} \\ \vdots \\ \lambda_{mj}^{\varphi} \end{pmatrix}.$$

Mit Blick auf Lemma 1.1.6 zeigen wir, dass die Funktionen Abb und Mat zueinander invers sind. Um zu zeigen, dass $\mathrm{Mat} \circ \mathrm{Abb}$ die Identität ist, betrachten wir eine beliebige Matrix $M = (\lambda_{ij}) \in K^{m\times n}$ und rechnen

$$\mathrm{Abb}(M)(e_j) = M \cdot e_j = \begin{pmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{mj} \end{pmatrix}.$$

Aufgrund der Definition von Mat liefert das in der Tat

$$\mathrm{Mat}(\mathrm{Abb}(M)) = (\lambda_{ij}) = M.$$

Um zu zeigen, dass auch $\mathrm{Abb} \circ \mathrm{Mat}$ die Identität ist, betrachten wir eine beliebige lineare Abbildung $\varphi : K^n \to K^m$. Wir müssen zeigen, dass diese mit der Abbildung $\mathrm{Abb}(\mathrm{Mat}(\varphi))$ übereinstimmt. Da letztere linear ist, brauchen wir nach Korollar 1.3.9 nur die Werte auf den Einheitsvektoren zu betrachten. Hier gilt

$$\mathrm{Abb}(\mathrm{Mat}(\varphi))(e_j) = \mathrm{Mat}(\varphi) \cdot e_j = (\lambda_{ij}^{\varphi}) \cdot e_j = \begin{pmatrix} \lambda_{1j}^{\varphi} \\ \vdots \\ \lambda_{mj}^{\varphi} \end{pmatrix} = \varphi(e_j),$$

wie gewünscht. □

Zum Abschluss dieses Unterkapitels wollen wir die Verknüpfung linearer Abbildungen noch zu einer geeigneten Matrixmultiplikation in Beziehung setzen. Das folgende Ergebnis sei zur Übung überlassen.

Lemma 1.3.17. *Für lineare Abbildungen $\varphi : U \to V$ und $\psi : V \to W$ zwischen Vektorräumen über einem gemeinsamen Körper ist auch $\psi \circ \varphi : U \to W$ linear.*

Da lineare Abbildungen und Matrizen gemäß Theorem 1.3.16 in Bijektion stehen, ist die folgende Definition also zulässig.

Definition 1.3.18. Für Matrizen $M \in K^{l\times m}$ und $N \in K^{m\times n}$ über einem Körper K definieren wir $M \cdot N \in K^{l\times n}$ durch die Setzung

$$\mathrm{Abb}(M \cdot N) = \mathrm{Abb}(M) \circ \mathrm{Abb}(N).$$

Aus dem folgenden Ergebnis geht unter anderem hervor, dass die vorige Definition eine Verallgemeinerung der Multiplikation von Matrix und Vektor darstellt, wie sie in Definition 1.3.12 eingeführt wurde (wenn wir die Vektoren aus K^n mit den $n \times 1$-Matrizen identifizieren – woraus auch noch einmal ersichtlich wird, weswegen wir die Elemente des K^n nun als Spaltenvektoren schreiben). Deswegen ist es auch harmlos, die Verknüpfung mit demselben Symbol zu bezeichnen.

Proposition 1.3.19. *Für $M = (\mu_{ij})_{1\leq i\leq l, 1\leq j\leq m}$ und $N = (\nu_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ ist*

$$M \cdot N = (\lambda_{ik})_{1\leq i\leq l, 1\leq k\leq n} \quad \textit{mit} \quad \lambda_{ik} = \sum_{j=1}^{m} \mu_{ij} \cdot \nu_{jk}.$$

Beweis. Wir schreiben temporär $M * N := (\lambda_{ik})$ mit λ_{ik} wie in der Proposition. Für jedes $k = 1, \dots n$ gilt

$$\begin{aligned}\operatorname{Abb}(M \cdot N)(e_k) = \operatorname{Abb}(M) \circ \operatorname{Abb}(N)(e_k) = M \cdot (N \cdot e_k) = M \cdot \begin{pmatrix} \nu_{1k} \\ \vdots \\ \nu_{mk} \end{pmatrix} \\ = \begin{pmatrix} \sum_{j=1}^{m} \mu_{1j} \cdot \nu_{jk} \\ \vdots \\ \sum_{j=1}^{m} \mu_{lj} \cdot \nu_{jk} \end{pmatrix} = \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{lk} \end{pmatrix} = (M * N) \cdot e_k = \operatorname{Abb}(M * N)(e_k).\end{aligned}$$

Mit Korollar 1.3.9 folgt $\operatorname{Abb}(M \cdot N) = \operatorname{Abb}(M * N)$ und also $M \cdot N = M * N$. □

Ein Spezialfall in suggestiver Notation ist sicher erneut hilfreich:

Beispiel 1.3.20. Es gilt

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \cdot \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix} = \begin{pmatrix} a_1x_1 + a_2x_2 & a_1y_1 + a_2y_2 & a_1z_1 + a_2z_2 \\ b_1x_1 + b_2x_2 & b_1y_1 + b_2y_2 & b_1z_1 + b_2z_2 \end{pmatrix}.$$

Es ist natürlich genauso möglich (und wohl sogar weitaus üblicher), das Ergebnis von Proposition 1.3.19 zur Definition zu machen und daraus die Gleichung in unserer Definition 1.3.18 abzuleiten – wobei man zur Motivation wohl dennoch mit der Verknüpfung von Abbildungen beginnen und unseren Beweis von Proposition 1.3.19 quasi als informelle Vorüberlegung vorausschicken würde. In jedem Fall hat die abstraktere Charakterisierung aus Definition 1.3.18 den Vorteil, dass wir das folgende Ergebnis ohne mühsame Rechnungen mit Koeffizienten erhalten.

Korollar 1.3.21. *Die Multiplikation von Matrizen ist assoziativ.*[25]

Beweis. Für beliebige Matrizen L, M, N mit geeigneter Zeilen- und Spaltenzahl gibt Lemma 1.1.7 uns

$$\begin{aligned}\operatorname{Abb}\big(L \cdot (M \cdot N)\big) &= \operatorname{Abb}(L) \circ \operatorname{Abb}(M \cdot N) = \operatorname{Abb}(L) \circ \big(\operatorname{Abb}(M) \circ \operatorname{Abb}(N)\big) \\ &= \big(\operatorname{Abb}(L) \circ \operatorname{Abb}(M)\big) \circ \operatorname{Abb}(N) = \operatorname{Abb}\big((L \cdot M) \cdot N\big).\end{aligned}$$

Wegen der Injektivität von Abb folgt $L \cdot (M \cdot N) = (L \cdot M) \cdot N$. □

Man kann sehen, dass die Menge $K^{m\times n}$ von Matrizen zum Vektorraum $K^{m\cdot n}$ in Bijektion steht (‚Zeilen nebeneinander schreiben'). Damit erbt auch die Menge der linearen Abbildungen eine Vektorraumstruktur, welche wir allerdings erst später thematisieren werden.

[25]Gemäß der Bemerkung nach Definition 1.3.18 erhält man als Spezialfall auch die Gleichung $M \cdot (N \cdot x) = (M \cdot N) \cdot x$ für einen Vektor x.

1.4. **Der Hamming-Code.** Am Ende von Unterkapitel 1.1 hatten wir uns das Ziel gesetzt, eine fehlerkorrigierende Kodierung $c : \mathbb{B}^4 \to \mathbb{B}^7$ zu finden. In Anbetracht der Propositionen 1.1.10 und 1.3.4 scheint es einen Versuch wert, diese linear zu wählen. Da lineare Abbildungen zu Matrizen korrespondieren, hätten wir dann

$$c(x) = H \cdot x \quad \text{für eine Matrix } H \in \mathbb{B}^{7\times 4}.$$

Aus den beiden eben zitierten Propositionen haben wir das entscheidende Kriterium

$$d(0, H \cdot x) \geq 3 \quad \text{für alle } x \neq 0.$$

Es ist hierfür nicht unmittelbar zwingend aber doch naheliegend, auf

$$d(0, H \cdot x) \geq d(0, x)$$

zu zielen, weil wir dann nur noch Vektoren $x \neq 0$ mit $d(0, x) < 3$ bedenken müssen. Wir versuchen es aus diesem Grund mit einer Matrix der Form

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{pmatrix},$$

weil wir dann nämlich

$$H \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ ? \\ ? \\ ? \end{pmatrix}$$

erhalten. Ein weiterer wichtiger Grund für diesen Ansatz ist, dass wir so x leicht aus $H \cdot x$ zurückgewinnen können. Nun rechnen wir

$$H \cdot e_i = e_i + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ a_i \\ b_i \\ c_i \end{pmatrix} \quad \text{und} \quad H \cdot (e_i + e_j) = e_i + e_j + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ a_i + a_j \\ b_i + b_j \\ c_i + c_j \end{pmatrix}.$$

Die erste Gleichung macht es nötig, dass in jeder Spalte von H zumindest zwei weitere Einser vorkommen. Aus der zweiten Gleichung ergibt sich, dass für beliebige $i \neq j$ zumindest eine der Summen $a_i + a_j$ und $b_i + b_j$ sowie $c_i + c_j$ den Wert 1 annehmen muss. Im Körper mit zwei Elementen läuft das darauf hinaus, dass zumindest eine der Ungleichungen $a_i \neq a_j$ und $b_i \neq b_j$ sowie $c_i \neq c_j$ zu erfüllen ist – sprich, dass die Spalten $(a_i, b_i, c_i)^\mathsf{T}$ paarweise verschieden sein müssen. Bis auf Vertauschung dieser (Teil-)Spalten ergibt sich als einzige Möglichkeit:

Definition 1.4.1. Die Hamming-Matrix ist gegeben durch

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Die Abbildung $\mathrm{Abb}(H) : \mathbb{B}^4 \to \mathbb{B}^7$ nennen wir Hamming-Kodierung.

Aus unseren vorigen Überlegungen ergibt sich das folgende Ergebnis.

Proposition 1.4.2. *Die Hamming-Kodierung korrigiert einen Fehler.*

Wir wollen zum Abschluss noch überlegen, wie aus einer ggf. leicht fehlerhaft übertragenen Nachricht y mit $d(H \cdot x, y) \leq 1$ die intendierte Nachricht x rekonstruiert werden kann. Hierzu führen wir den folgenden Begriff ein.

Definition 1.4.3. Eine lineare Korrektur[26] für eine Kodierung $c : \mathbb{B}^m \to \mathbb{B}^n$ ist eine lineare Abbildung $\tilde{c} : \mathbb{B}^n \to \mathbb{B}^k$ mit $k \in \mathbb{N}$, die folgende Eigenschaften hat:

(i) Es gilt $\tilde{c} \circ c(x) = 0$ für alle $x \in \mathbb{B}^m$.
(ii) Die Bilder $\tilde{c}(e_i)$ der Einheitsvektoren $e_1, \dots, e_n$ sind alle ungleich Null und außerdem paarweise verschieden.

Die Linearität der Korrektur ist insbesondere für das Folgende wichtig.

Proposition 1.4.4. *Für $c : \mathbb{B}^m \to \mathbb{B}^n$ mit linearer Korrektur $\tilde{c} : \mathbb{B}^n \to \mathbb{B}^k$ gilt: Hat man $d(c(x), y) \leq 1$, so ist $\tilde{c}(y) \in \{0\} \cup \{\tilde{c}(e_i) : i = 1, \dots, n\}$ und*

$$c(x) = \begin{cases} y & \text{wenn } \tilde{c}(y) = 0, \\ y + e_i & \text{wenn } \tilde{c}(y) = \tilde{c}(e_i). \end{cases}$$

Beweis. Ist $d(c(x), y) \leq 1$, so haben wir $y = c(x) + b \cdot e_j$ für geeignete $b \in \mathbb{B}$ und j (wegen $1 + 1 = 0$ in $\mathbb{B}$). Mit der Linearität von $\tilde{c}$ folgt

$$\tilde{c}(y) = \tilde{c} \circ c(x) + b \cdot \tilde{c}(e_j) = b \cdot \tilde{c}(e_j).$$

Hierbei gilt die zweite Gleichheit wegen Punkt (i) aus der vorigen Definition. Hat man nun $\tilde{c}(y) = 0$, so liefert Punkt (ii) aus der Definition $b = 0$ und also $c(x) = y$. Gilt hingegen $\tilde{c}(y) = \tilde{c}(e_i) \neq 0$ für beliebiges i, so ergibt sich $b = 1$ und damit auch $\tilde{c}(y) = \tilde{c}(e_j)$. Hieraus folgt $i = j$ und also $c(x) = c(x) + e_j + e_j = y + e_i$. □

Wie wir oben gesehen haben, besteht x im Fall der Hamming-Kodierung aus den ersten vier Einträgen von $c(x)$, kann also leicht aus letzterem berechnet werden. Wegen der vorigen Proposition kann man aus einer linearen Korrektur hier also leicht eine Dekodierung im Sinn von Definition 1.1.2 bestimmen.

Die Motivation für die folgende Definition wird im Beweis von Proposition 1.4.6 klar werden.

[26] Dieser Begriff von Korrektur ist in der Literatur nicht unbedingt standard aber aufgrund der folgenden Ergebnisse gerechtfertigt. Es besteht eine enge Verbindung zur sogenannten Paritätsprüfung (‚Parity checks').

Definition 1.4.5. Die Korrekturmatrix für die Hamming-Kodierung ist gegeben durch

$$\tilde{H} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Wir merken an, dass drei die optimale Zeilenzahl ist, da der Zielraum $\mathbb{B}^3$ dann sieben von Null verschiedene Vektoren enthält, was angesichts von Definition 1.4.3(ii) gerade ausreicht.

Proposition 1.4.6. *Die Abbildung* $\mathrm{Abb}(\tilde{H}) : \mathbb{B}^7 \to \mathbb{B}^3$ *ist eine lineare Korrektur für die Hamming-Kodierung.*

Beweis. Die Matrizen H und $\tilde{H}$ sind von der Form

$$\tilde{H} = \left(\begin{array}{c|ccc} & 1 & 0 & 0 \\ M & 0 & 1 & 0 \\ & 0 & 0 & 1 \end{array}\right) \quad \text{und} \quad H = \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline & & & \\ & & M & \\ & & & \end{array}\right)$$

mit gemeinsamer Matrix M.[27] Für $M = (\lambda_{ij})$ schreiben wir $M \cdot 2 = (\lambda_{ij} + \lambda_{ij})$. Man sieht dann

$$\tilde{H} \circ H = M \cdot 2 = 0,$$

wobei die letzte Gleichung auf $1+1 = 0$ in $\mathbb{B}$ zurückzuführen ist. Hierbei bezeichnet 0 die Matrix, in der alle Einträge gleich Null sind (hier mit drei Zeilen und vier Spalten). Auf Ebene der linearen Abbildungen gilt also

$$\mathrm{Abb}(\tilde{H}) \circ \mathrm{Abb}(H)(x) = \mathrm{Abb}(\tilde{H} \cdot H)(x) = \tilde{H} \cdot H \cdot x = 0.$$

Das bedeutet, dass die Abbildung $\mathrm{Abb}(\tilde{H})$ in Bezug auf die Hamming-Kodierung Bedingung (i) aus Definition 1.4.3 erfüllt. Bedingung (ii) ist erfüllt, weil $\mathrm{Abb}(\tilde{H})(e_i)$ die i-te Spalte von $\tilde{H}$ ist, wobei die Spalten ungleich Null und außerdem paarweise verschieden sind. □

Es ist nun an der Zeit, die Kodierung und Dekodierung konkret auszuprobieren:

Beispiel 1.4.7. Die folgenden Schritte kann man am Besten im Austausch mit einer anderen Person nachvollziehen:

(1) Suchen Sie sich eine Folge von vier Bits aus und notieren Sie diese als Spaltenvektor

$$x =$$

[27] Die quadratischen Matrizen mit Einsern auf der Diagonale und Nullern abseits davon heißen Einheitsmatrizen. Sie werden später noch einmal Thema sein. Man kann sich aber bereits jetzt überzeugen, dass sie als neutrales Element der Multiplikation fungieren und also der Identitätsabbildung entsprechen. Auch die Addition von Matrizen werden wir später allgemein betrachten.

(2) Berechnen Sie mittels der Hamming-Kodierung die siebenstellige Folge

$$c(x) := \mathrm{Abb}(H)(x) = H \cdot x =$$

(3) Wählen Sie eine siebenstellige Folge y von Bits, die sich an höchstens einer Stelle von $c(x)$ unterscheidet. Notieren Sie diese als Spaltenvektor

$$y =$$

(4) Lassen Sie sich von einer anderen Person die Folge nennen, welche diese Person in Schritt (3) gewählt hat. Notieren Sie diese als

$$y' =$$

(5) Verwenden Sie die oben besprochene Korrektur $\tilde{c} := \mathrm{Abb}(\tilde{H})$ und berechnen Sie

$$\tilde{c}(y') = \tilde{H} \cdot y' =$$

(6) Sei x' die (ihnen noch unbekannte) Folge, welche sich die andere Person in Schritt (1) ausgesucht hat. Gemäß Proposition 1.4.4 ist $c(x') = y'$ im Fall von $\tilde{c}(y') = 0$ sowie $c(x') = y' + e_i$ falls $\tilde{c}(y')$ mit der i-ten Spalte von $\tilde{H}$ übereinstimmt (siehe dazu auch den letzten Satz im Beweis von Proposition 1.4.6). Bestimmen Sie so

$$c(x') =$$

(7) Im Abschnitt vor Definition 1.4.1 haben wir gesehen, dass x' aus den ersten vier Einträgen von $c(x') = H \cdot x'$ besteht. Notieren Sie

$$x' =$$

(8) Fragen Sie jeweils die andere Person, ob Sie die von dieser gewählte Nachricht korrekt wiedergewonnen haben.

Zum Abschluss dieses Kapitels weisen wir auf eine Verallgemeinerung hin:

Bemerkung 1.4.8. Die Hamming-Kodierung funktioniert nicht nur mit ‚Codewörtern‘ der Länge $7 = 2^3 - 1$ sondern analog auch mit Codewörtern der Länge $2^k - 1$

für beliebiges $k \geq 2$, wobei die Menge der kodierten Nachrichten für jedes k so groß wie möglich ist.[28]

[28]Siehe hierzu Matoušek: Thirty-three Miniatures (vgl. Fußnote 1).

2. Klassifikation von Vektorräumen

In diesem Kapitel werden wir sehen, dass die Vektorräume über einem gegebenen Körper im Wesentlichen[29] durch ihre sogenannte Dimension bestimmt sind. Das bedeutet insbesondere, dass die K^n aus dem vorigen Kapitel im Wesentlichen die einzigen Vektorräume von endlicher Dimension sind. Hieraus wird auch ersichtlich, dass alle linearen Abbildungen zwischen endlichdimensionalen Vektorräumen durch Matrizen darstellbar sind.

2.1. Erzeugendensysteme, lineare Unabhängigkeit und Basen.

In diesem Unterkapitel untersuchen wir, wie sich Vektoren als Linearkombinationen $\sum_{i=1}^n \lambda_i v_i$ von festgelegten v_i schreiben lassen. Unser Ziel ist gewissermaßen eine Verallgemeinerung von Lemma 1.3.7 auf beliebige Vektorräume. In einem gewissen Sinn wird dies auch eine verallgemeinerte Matrixdarstellung liefern.

Definition 2.1.1. Sei V ein Vektorraum über einem Körper K. Der Spann[30] einer Menge $M \subseteq V$ ist gegeben durch

$$\operatorname{Span}(M) := \left\{ \sum_{i=1}^n \lambda_i v_i : n \in \mathbb{N} \text{ und } \lambda_1, \dots, \lambda_n \in K \text{ und } v_1, \dots, v_n \in M \right\}.$$

Gilt $\operatorname{Span}(M) = V$, so nennt man M ein Erzeugendensystem.

Man beachte, dass $0 \in \operatorname{Span}(M)$ selbst für $M = \emptyset$ gilt (mit $n = 0$).

Beispiel 2.1.2. Im $\mathbb{R}^3$ ist

$$\operatorname{Span}(e_1, e_2) = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right) = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x, y \in \mathbb{R} \right\}$$

die sogenannte xy-Ebene.[31] Insbesondere ist die Menge $\{e_1, e_2\}$ kein Erzeugendensystem aber die Menge $\{e_1, e_2, e_3\}$ schon. Allgemein bilden Vektoren im $\mathbb{R}^3$ genau dann ein Erzeugendensystem, wenn sie nicht in einer gemeinsamen Ebene durch den Ursprung liegen.

Die folgende Eigenschaft wird später wichtig werden.

Lemma 2.1.3. *Die Abbildung $M \mapsto \operatorname{Span}(M)$ ist ein Hüllenoperator, d.h. es gilt*

$$M \subseteq \operatorname{Span}(N) \quad \Leftrightarrow \quad \operatorname{Span}(M) \subseteq \operatorname{Span}(N).$$

Beweis. Die Richtung von rechts nach links ist äquivalent zu $M \subseteq \operatorname{Span}(M)$, wovon man sich überzeugen möge. Für die andere Richtung betrachten wir ein beliebiges Element

$$v = \sum_{i=1}^n \lambda_i \cdot v_i \in \operatorname{Span}(M)$$

[29] Das hat hier die präzise Bedeutung ‚bis auf Isomorphie', wobei wir den Begriff des Isomorphismus erst später klären werden.

[30] Man spricht auch von der linearen Hülle oder dem Erzeugnis.

[31] Für endliches $M = \{u_1, \dots, u_m\}$ schreibt man oft $\operatorname{Span}(u_1, \dots, u_m)$ statt $\operatorname{Span}(M)$, lässt also die Mengenklammern weg.

mit $v_i \in M$. Gilt $M \subseteq \mathrm{Span}(N)$, so können wir jeweils

$$v_i = \sum_{j=1}^{m(i)} \mu_{ij} \cdot w_{ij}$$

mit $w_{ij} \in N$ schreiben. Die rechte Seite ergibt sich wegen

$$v = \sum_{i=1}^{n} \sum_{j=1}^{m(i)} \lambda_i \mu_{ij} \cdot w_{ij} \in \mathrm{Span}(N),$$

wobei man die Doppelsumme als eine einzelne Summe mit $m(1) + \ldots + m(n)$ Summanden auffasst. □

Wegen $M \subseteq \mathrm{Span}(M)$ ist V selbst ein Erzeugendensystem. Wir sind an möglichst kleinen Erzeugendensystemen interessiert, was sich durch den folgenden Begriff präzisieren lässt.

Definition 2.1.4. Eine Teilmenge M eines K-Vektorraums V heißt linear abhängig, wenn es $\lambda_1, \ldots, \lambda_n \in K$ und paarweise verschiedene $v_1, \ldots, v_n \in M$ mit

$$\sum_{i=1}^{n} \lambda_i v_i = 0$$

gibt, wobei $\lambda_i \neq 0$ für mindestens ein $i \in \{1, \ldots, n\}$ gilt (insbesondere also $n > 0$). Ansonsten heißt M linear unabhängig.

Man beachte, dass $\emptyset$ linear unabhängig und $\{0\}$ linear abhängig ist. Würde man nicht auf paarweise verschiedenen Vektoren bestehen, so wäre jede Menge $M \neq \emptyset$ linear abhängig, denn für beliebiges $v \in M$ ist $1 \cdot v + (-1) \cdot v = 0$ trotz $1 \neq 0$.

Beispiel 2.1.5. Im $\mathbb{R}^3$ sind $\{e_1, e_2\}$ und $\{e_1, e_2, e_3\}$ linear unabhängig, denn

$$x \cdot e_1 + y \cdot e_2 + z \cdot e_3 = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

gilt nur für $x = y = z = 0$. Hingegen ist

$$2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = 0,$$

sodass e_1 und $2 \cdot e_1$ linear abhängig sind. Allgemein sind zwei (bzw. drei) Vektoren im $\mathbb{R}^3$ genau dann linear abhängig, wenn sie in einer gemeinsamen Geraden (bzw. Ebene) durch den Ursprung liegen.

Aus den Definitionen ist unmittelbar ersichtlich, dass M genau dann ein Erzeugendensystem ist, wenn sich jeder Vektor auf mindestens eine Weise als Linearkombination von Vektoren aus M schreiben lässt. Im Fall einer linear unabhängigen Menge kann man ihn auf höchstens eine Weise so schreiben:

Lemma 2.1.6. *Eine Menge $M \subseteq V$ ist genau dann linear unabhängig, wenn für paarweise verschiedene $v_1, \ldots, v_n \in M$ stets gilt: Hat man*

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

so ist $\lambda_i = \mu_i$ für alle $i = 1, \ldots, n$.

Beweis. Die gegebene Gleichung ist äquivalent zu

$$\sum_{i=1}^{n}(\lambda_i - \mu_i)\cdot v_i = 0.$$

Ist M linear unabhängig, so folgt $\lambda_i - \mu_i = 0$ und somit $\lambda_i = \mu_i$ für alle i. Ist hingegen M linear abhängig, so findet man v_i mit

$$\sum_{i=1}^{n}\lambda_i \cdot v_i = 0 = \sum_{i=1}^{n} 0 \cdot v_i,$$

wobei $\lambda_i \neq 0$ für ein i gilt. Wählt man $\mu_1 = \ldots = \mu_n = 0$, so widerspricht dies der Folgerung des Lemmas. □

Wir merken an, dass man das Lemma auch anwenden kann, wenn auf beiden Seiten einer Gleichung verschiedene Vektoren aus M stehen. Gilt etwa

$$\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 = \lambda_2' \cdot v_2 + \lambda_3 \cdot v_3$$

für paarweise verschiedene v_1, v_2, v_3 aus einer linear unabhängigen Menge M, so ergänzt man zu

$$\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 + 0 \cdot v_3 = 0 \cdot v_1 + \lambda_2' \cdot v_2 + \lambda_3 \cdot v_3,$$

um mit dem Lemma auf $\lambda_1 = 0 = \lambda_3$ und $\lambda_2 = \lambda_2'$ zu schließen.

Definition 2.1.7. Sei V ein K-Vektorraum. Eine Teilmenge $M \subseteq V$ heißt Basis, wenn M sowohl ein Erzeugendensystem als auch linear unabhängig ist.

Aus dem letzten Lemma und der Bemerkung davor erhält man:

Korollar 2.1.8. *Ist $B \subseteq V$ eine Basis, so lässt sich jedes $v \in V$ als*

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

mit paarweise verschiedenen $v_i \in B$ und eindeutig bestimmten λ_i schreiben.

Analog zur Bemerkung nach dem Lemma ist dabei impliziert, dass für $\lambda_i \neq 0$ auch die v_i bis auf ihre Reihenfolge eindeutig bestimmt sind.

Beispiel 2.1.9. Im $\mathbb{R}^3$ ist $\{e_1, e_2, e_3\}$ eine Basis, die sogenannte Standardbasis. Auch die Menge

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right\} \subseteq \mathbb{R}^3$$

ist ein Beispiel für eine Basis, wie man prüfen möge.

Zur Konstruktion einer Basis ist die folgende Charakterisierung hilfreich.

Proposition 2.1.10. *Für einen K-Vektorraum V und $M \subseteq V$ sind äquivalent:*

(i) M ist eine Basis,
(ii) M ist ein minimales Erzeugendensystem,[32]
(iii) M ist maximal linear unabhängig.[33]

[32]Minimalität bedeutet hier, dass keine echte Teilmenge $M' \subsetneq M$ noch Erzeugendensystem ist.
[33]Eine echte Obermenge $M'' \supsetneq M$ ist also stets linear abhängig.

Beweis. Wir nehmen zunächst (i) an und müssen die Minimalität bzw. Maximalität in (ii) und (iii) beweisen. Wäre $M' \subsetneq M$ im Widerspruch zu (ii) ein Erzeugendensystem, so könnte man $v \in M \backslash M'$ schreiben als

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

mit paarweise verschiedenen $v_1, \ldots, v_n \in M'$. Insbesondere sind die v_i von v verschieden. Wegen Lemma 2.1.6 kann also M nicht linear unabhängig und somit keine Basis sein. Dieser Widerspruch zu (i) etabliert (ii). Da M gemäß (i) selbst ein Erzeugendensystem ist, folgt genauso, dass $M'' \supsetneq M$ nicht linear unabhängig sein kann. Dies liefert (iii).

Wir nehmen nun (ii) an und zeigen für (i), dass M linear unabhängig ist. Wäre dies nicht so, dann hätte man

$$\sum_{i=1}^{n} \lambda_i v_i = 0$$

für paarweise verschiedene $v_i \in M$, wobei man nach Umordnen der Summanden von $\lambda_n \neq 0$ ausgehen kann. Es ergibt sich

$$v_n = \sum_{i=1}^{n-1} -\frac{\lambda_i}{\lambda_n} \cdot v_i \in \operatorname{Span}(M \backslash \{v_n\})$$

und also $M \subseteq \operatorname{Span}(M \backslash \{v_n\})$. Mit Lemma 2.1.3 folgt $\operatorname{Span}(M) = \operatorname{Span}(M \backslash \{v_n\})$, im Widerspruch zu (ii).

Schließlich nehmen wir (iii) an und zeigen für (i), dass M ein Erzeugendensystem ist. Hätte man $v \notin \operatorname{Span}(M)$ und also $v \notin M$, so wäre $M \cup \{v\}$ nach (iii) linear abhängig. Dies liefert

$$\lambda_{n+1} \cdot v + \sum_{i=1}^{n} \lambda_i \cdot v_i = 0$$

mit paarweise verschiedenen $v_1, \ldots, v_n \in M$ und $\lambda_i \neq 0$ für ein $i \in \{1, \ldots, n+1\}$. Dabei kann nicht $\lambda_{n+1} = 0$ gelten, weil sonst schon M linear abhängig wäre. Wir erhalten also

$$v = \sum_{i=1}^{n} -\frac{\lambda_i}{\lambda_{n+1}} \cdot v_i \in \operatorname{Span}(M),$$

anders als angenommen. □

Hat man in einem Vektorraum ein endliches Erzeugendensystem zur Verfügung, so sieht man unmittelbar, dass es ein minimales Erzeugendensystem und also eine Basis gibt. Im nächsten Unterkapitel werden wir ein technisches Hilfsmittel beweisen, mit dem wir im unendlichen Fall ganz ähnlich schließen können. Zuvor untersuchen wir, wie sich lineare Abbildungen mit Hilfe von Basen charakterisieren lassen. Die folgende Konstruktion ist wohldefiniert wegen Korollar 2.1.8.

Definition 2.1.11. Betrachte K-Vektorräume V und W sowie eine Basis $B \subseteq V$. Für eine Funktion $f : B \to W$ definieren wir dann $\operatorname{Abb}(f) : V \to W$ durch

$$\operatorname{Abb}(f)(v) = \sum_{i=1}^{n} \lambda_i f(v_i) \quad \text{für } v = \sum_{i=1}^{n} \lambda_i v_i \text{ mit } \lambda_i \in K \text{ und } v_i \in B.$$

Hat man $V = K^n$ und $W = K^m$ und ist B die Standardbasis, so stimmt $\mathrm{Abb}(f)$ mit der Abbildung $\mathrm{Abb}(M)$ aus Definition 1.3.14 überein, wenn M die Matrix mit i-ter Spalte $f(e_i)$ ist. Dies folgt mit Korollar 1.3.9 aus

$$\mathrm{Abb}(f)(e_i) = f(e_i) = M \cdot e_i = \mathrm{Abb}(M)(e_i).$$

In diesem Sinn handelt es sich um eine Verallgemeinerung der vorigen Schreibweise (weswegen wir auch dieselbe Notation verwenden). Hat man im allgemeinen Fall noch eine Basis $C \subseteq W$ gegeben, so kann man die Werte $f(v)$ mit $v \in B$ wieder als endliche Linearkombinationen von Vektoren aus C schreiben. Man könnte $\mathrm{Abb}(f)$ also durch eine ggf. unendliche Matrix darstellen, wobei in jeder Spalte nur endlich viele Einträge von Null verschieden wären. Diesen Standpunkt werden wir nicht weiterverfolgen. Nützen wird uns hingegen die folgende Verallgemeinerung von Theorem 1.3.16, die genauso bewiesen wird wie im endlichen Fall.

Proposition 2.1.12. *Für eine fest gewählte Basis $B \subseteq V$ ist $f \mapsto \mathrm{Abb}(f)$ eine Bijektion zwischen den Funktionen $B \to W$ und den linearen Abbildungen $V \to W$.*

Das folgende Ergebnis stellt eine Verbindung mit den Begriffen aus dem vorliegenden Unterkapitel her.

Proposition 2.1.13. *Betrachte $f : B \to W$ wie oben. Die Abbildung $\mathrm{Abb}(f)$ ist*

(a) injektiv genau dann, wenn f injektiv und $\mathrm{img}(f)$ linear unabhängig ist,[34]
(b) surjektiv genau dann, wenn $\mathrm{img}(f)$ ein Erzeugendensystem von W ist,
(c) bijektiv genau dann, wenn f injektiv und $\mathrm{img}(f)$ eine Basis von W ist.

Beweis. Wir überlassen (b) zur Übung und bemerken, dass (c) sich unmittelbar aus (a) und (b) ergibt. Im Folgenden beweisen wir Teil (a).

Um die Injektivität von $\mathrm{Abb}(f)$ nachzuweisen, zeigen wir, dass $\mathrm{Abb}(f)(v) = 0$ nur für $v = 0$ gelten kann.[35] Wir schreiben dazu $v = \sum_{i=1}^n \lambda_i v_i$ mit paarweise verschiedenen Vektoren $v_i \in B$. Angenommen, wir haben

$$\mathrm{Abb}(f)(v) = \sum_{i=1}^n \lambda_i f(v_i) = 0.$$

Ist f injektiv, so sind auch die Vektoren $f(v_i) \in \mathrm{img}(f)$ paarweise verschieden. Wenn also $\mathrm{img}(f)$ linear unabhängig ist, muss $\lambda_1 = \ldots = \lambda_n = 0$ gelten, was in der Tat $v = 0$ ergibt. Für die Injektivität nehmen wir $\mathrm{Abb}(f)(v) = \mathrm{Abb}(f)(w)$ an. Mit der Linearität folgt $\mathrm{Abb}(f)(v - w) = 0$. Wie eben gesehen ergibt das $v - w = 0$ und also $v = w$, wie gewünscht.

Ist umgekehrt $\mathrm{Abb}(f)$ injektiv, so ist es auch f, da $\mathrm{Abb}(f)(v) = f(v)$ für alle $v \in B$ gilt. Für die lineare Unabhängigkeit betrachten wir eine Gleichung

$$\sum_{i=1}^n \lambda_i f(v_i) = 0$$

mit paarweise verschiedenen Vektoren $f(v_i) \in \mathrm{img}(f)$. Man hat dann also

$$\mathrm{Abb}(f)(v) = 0 \quad \text{für } v = \sum_{i=1}^n \lambda_i v_i.$$

[34]Mit $\mathrm{img}(g) = \{g(x) : x \in M\}$ wird das Bild einer Funktion $g : M \to N$ bezeignet.

Ist $\mathrm{Abb}(f)$ injektiv, so folgt daraus $v = 0$. Da die v_i paarweise verschieden sind und aus der Basis B stammen, muss $\lambda_1 = \ldots = \lambda_n = 0$ gelten. □

Man kann das vorige Ergebnis in die Sprache der Matrizen übersetzen, was für Anwendungen besonders wichtig ist:

Bemerkung 2.1.14. Mit den Überlegungen, die wir nach Definition 2.1.11 angestellt haben, folgt: Für eine Matrix $M \in K^{m\times n}$ ist die durch $\mathrm{Abb}(M)(x) = M \cdot x$ gegebene Abbildung $\mathrm{Abb}(M) : K^n \to K^m$ genau dann

(a) injektiv, wenn die Spalten von M paarweise verschieden sowie linear unabhängig sind,
(b) surjektiv, wenn die Spalten von M ein Erzeugendensystem von K^m bilden,
(c) bijektiv, wenn die Spalten von M paarweise verschieden sind und eine Basis von K^m bilden.

Zum Abschluss dieses Kapitels widmen wir uns kurz den Umkehrabbildungen von linearen Abbildungen.

Lemma 2.1.15. *Ist eine lineare Abbildung $\Phi : V \to W$ bijektiv, so ist auch ihre Umkehrabbildung $\Phi^{-1} : W \to V$ linear.*

Beweis. Für gegebenes $w \in W$ schreiben wir $w = \Phi(v)$. Wir rechnen dann

$$\Phi^{-1}(\lambda w) = \Phi^{-1}(\lambda\Phi(v)) = \Phi^{-1}(\Phi(\lambda v)) = \lambda v = \lambda\Phi^{-1}(w).$$

Die Verträglichkeit mit der Addition überlassen wir als Übung. □

In Anbetracht des Lemmas führen wir die folgende Terminologie ein.[36]

Definition 2.1.16. Eine lineare Abbildung $\Phi : V \to W$ heißt Isomorphismus, wenn sie bijektiv ist. Gibt es einen solchen Isomorphismus, so schreiben wir $V \cong W$.

2.2. **Zornsches Lemma und Kardinalität.** In diesem Unterkapitel beweisen wir ein technisches Resultat, mit dem wir Basen in beliebigen Vektorräumen erhalten werden. Wir untersuchen auch, wie unendliche Mengen anhand ihrer Kardinalität – also gewissermaßen der Anzahl ihrer Elemente – verglichen werden können.[37]

Definition 2.2.1. Sei $\mathcal{Z}$ eine Menge, deren Elemente selbst wieder Mengen sind. Wir nennen $\mathcal{L} \subseteq \mathcal{Z}$ eine Kette, wenn man für $M, N \in \mathcal{L}$ stets $M \subseteq N$ oder $N \subseteq M$ hat. Man sagt, $\mathcal{Z}$ ist induktiv, wenn für eine Kette $\mathcal{L} \subseteq \mathcal{Z}$ stets $\bigcup \mathcal{L} \in \mathcal{Z}$ gilt.[38]

---

[35] Wie der Rest des Beweises zeigt, ist dies allgemein ausreichend, um die Injektivität einer linearen Abbildung nachzuweisen (vgl. Korollar 2.4.16).

[36] Wenn in einem Kontext der Begriff des Morphismus erklärt ist, so ist ein Isomorphismus definiert als ein Morphismus mit einer Umkehrabbildung, die wieder ein Morphismus ist. Das Lemma zeigt, dass diese Bedingung in unserem Fall mit der Bijektivität zusammenfällt. Das ist aber nicht immer der Fall: Für manche Fragen der Analysis möchte man nur stetige Funktionen als Morphismen zulassen. Betrachte nun etwa die Funktion $f : [0, 1) \cup [2, 3] \to [0, 2]$ mit $f(x) = x$ für $x \in [0, 1)$ und $f(x) = x - 1$ für $x \in [2, 3]$. Diese ist stetig und bijektiv, aber ihre Umkehrfunktion ist nicht stetig. In so einer Situation würde man f nicht als Isomorphismus bezeichnen. In einem gewissen Sinn teilen isomorphe Objekte alle wesentlichen Eigenschaften. Gleichzeitig werden wir sehen (z.B. beim Thema Basiswechsel), dass Isomorphismen einen nicht-trivialen Effekt haben können.

[37] Ein anderes Wort für Kardinalität ist Mächtigkeit.

[38] Ist $\mathcal{Y}$ eine Menge von Mengen, so bezeichnet $\bigcup \mathcal{Y} = \bigcup_{M\in\mathcal{Y}} M$ die Menge der Elemente, die in mindestens einer Menge $M \in \mathcal{Y}$ enthalten sind. Insbesondere gilt $\bigcup \mathcal{Y} = M \cup N$ für $\mathcal{Y} = \{M, N\}$. Analog besteht $\bigcap \mathcal{Y} = \bigcap_{M\in\mathcal{Y}} M$ aus den Elementen, die in allen Mengen $M \in \mathcal{Y}$ liegen. Das Adjektiv ‚induktiv' ist mit einiger Vorsicht zu verwenden, um Verwechslungen z.B. mit der – letztlich verwandten aber doch klar zu unterscheidenden – Induktion über $\mathbb{N}$ zu vermeiden.

Man beachte, dass für jedes induktive $\mathcal{Z}$ wegen $\emptyset \subseteq \mathcal{Z}$ auch $\emptyset = \bigcup \emptyset \in \mathcal{Z}$ und also insbesondere $\mathcal{Z} \neq \emptyset$ gilt. Das folgende Resultat liefert eine wichtige Veranschaulichung und Motivation.

Lemma 2.2.2. *Die Menge aller linear unabhängigen Teilmengen eines Vektorraums ist induktiv.*

Beweis. Betrachte eine Kette $\mathcal{L}$ aus linear unabhängigen Teilmengen. Wir argumentieren per Widerspruch und nehmen dazu an, dass $\bigcup \mathcal{L}$ linear abhängig ist. Man hat dann also

$$\sum_{i=1}^{n} \lambda_i v_i = 0$$

für paarweise verschiedene $v_1, \dots, v_n \in \bigcup \mathcal{L}$, wobei mindestens ein λ_i ungleich Null ist. Wir zeigen durch Induktion über i, dass die Vektoren $v_1, \dots, v_i$ in einer gemeinsamen Menge $M_i \in \mathcal{L}$ liegen. Für $i = n$ erhält man einen Widerspruch zur Annahme, dass $\mathcal{L}$ aus linear unabhängigen Mengen besteht. Im Induktionsschritt hat man $v_1, \dots, v_i \in M_i \in \mathcal{L}$ sowie außerdem $v_{i+1} \in N$ für eine Menge $N \in \mathcal{L}$ (wegen der Definition von $\bigcup \mathcal{L}$). Da $\mathcal{L}$ eine Kette ist, erhält man $M_i \subseteq N$ oder $N \subseteq M_i$. Man kann also $M_{i+1} := N$ beziehungsweise $M_{i+1} := M_i$ setzen. □

Folgendes Prinzip wird in verschiedenen mathematischen Gebieten angewendet.

Theorem 2.2.3 (Zornsches Lemma[39])**.** *Ist eine Menge $\mathcal{Z}$ induktiv, so hat sie ein maximales Element, d.h. es gibt ein $M \in \mathcal{Z}$ mit $N \notin \mathcal{Z}$ für alle $N \supsetneq M$.*

Beweis. Wir beweisen das Ergebnis zunächst unter der zusätzlichen Annahme, dass aus $M \subseteq N \in \mathcal{Z}$ stets $M \in \mathcal{Z}$ folgt. Für $M \in \mathcal{Z}$ sei

$$\bar{M} = \left\{ x \in \bigcup \mathcal{Z} : M \cup \{x\} \in \mathcal{Z} \right\}.$$

Man hat stets $M \subseteq \bar{M}$. Für jede Menge $M \in \mathcal{Z}$ mit $M \neq \bar{M}$ wählen wir ein Element $f(M) \in \bar{M} \backslash M$.[40] Wir setzen nun

$$g(M) := \begin{cases} M & \text{wenn } M = \bar{M}, \\ M \cup \{f(M)\} & \text{sonst.} \end{cases}$$

Unser Ziel ist, ein $M \in \mathcal{Z}$ mit $g(M) = M$ zu finden. Ein solches M ist nämlich maximal: Hätte man $M \subsetneq N \in \mathcal{Z}$, so könnte man ein $x \in N \backslash M$ betrachten. Mit der Zusatzannahme vom Beginn dieses Beweises würde dann $M \cup \{x\} \in \mathcal{Z}$ und also $M \neq \bar{M}$ folgen.

Unter einem Gerüst[41] verstehen wir eine Menge $\mathcal{G} \subseteq \mathcal{Z}$, sodass gilt:

(I) man hat $\emptyset \in \mathcal{G}$,
(II) aus $M \in \mathcal{G}$ folgt stets $g(M) \in \mathcal{G}$,
(III) für jede Kette $\mathcal{L} \subseteq \mathcal{G}$ gilt $\bigcup \mathcal{L} \in \mathcal{G}$.

Sei nun

$$\mathcal{G}_0 := \bigcap \{\mathcal{G} : \mathcal{G} \text{ ist ein Gerüst}\}$$

die Menge aller $M \in \mathcal{Z}$, welche in allen Gerüsten enthalten sind. Der Schnitt ist dabei nicht leer, weil $\mathcal{Z}$ selbst ein Gerüst ist (wegen $g(M) \in \mathcal{Z}$ und der Induktivität).

[39] Ergebnisse dieser Art haben neben Max Zorn auch Felix Hausdorff und Kazimierz Kuratowski gezeigt. Unser Beweis geht auf Ernst Zermelo zurück und ist übernommen aus Paul Halmos, Naive Set Theory. Van Nostrand, Princeton (NJ), 1960. Tatsächlich wird dort mit ein wenig zusätzlichem Aufwand eine noch etwas allgemeinere Version des Zornschen Lemmas bewiesen.

Man möge sich überzeugen, dass dann $\mathcal{G}_0$ auch ein Gerüst ist. Wegen der Definition des Durchschnitts ist es das kleinste Gerüst, d.h. es gilt $\mathcal{G}_0 \subseteq \mathcal{G}$ für jedes Gerüst $\mathcal{G}$.

Behauptung. Das Gerüst $\mathcal{G}_0$ ist eine Kette.

Beweis der Behauptung. Wir werden zeigen, dass

$$\mathcal{C} = \{M \in \mathcal{G}_0 : \text{für alle } N \in \mathcal{G}_0 \text{ ist } M \subseteq N \text{ oder } N \subseteq M\}$$

ein Gerüst ist. Wegen der Minimalität von $\mathcal{G}_0$ folgt daraus $\mathcal{C} = \mathcal{G}_0$, was die Behauptung liefert. Da stets $\emptyset \subseteq N$ gilt, haben wir zunächst $\emptyset \in \mathcal{C}$, sodass Eigenschaft (I) für $\mathcal{C}$ erfüllt ist.

Mit Blick auf Eigenschaft (III) betrachten wir eine Kette $\mathcal{L} \subseteq \mathcal{C}$. Für beliebiges $N \in \mathcal{G}_0$ tritt einer von zwei Fällen ein: Im ersten Fall gilt $M \subseteq N$ für alle $M \in \mathcal{L}$, woraus $\bigcup \mathcal{L} \subseteq N$ folgt. Im zweiten Fall gibt es ein $M \in \mathcal{L}$ mit $N \subseteq M \subseteq \bigcup \mathcal{L}$. In jedem Fall ist die Bedingung für $\bigcup L \in \mathcal{C}$ erfüllt.

Für Eigenschaft (II) fixieren wir ein $M \in \mathcal{C}$ und bilden

$$\mathcal{C}_M = \{N \in \mathcal{G}_0 : g(M) \subseteq N \text{ oder } N \subseteq M\}.$$

Wir werden zeigen, dass $\mathcal{C}_M$ ein Gerüst ist. Damit wird $\mathcal{C}_M = \mathcal{G}_0$ und dann $g(M) \in \mathcal{C}$ folgen: Für beliebiges $N \in \mathcal{G}_0$ erhält man nämlich $g(M) \subseteq N$ oder $N \subseteq M \subseteq g(M)$. Ähnlich wie oben sieht man, dass $\mathcal{C}_M$ die Eigenschaften (I) und (III) erfüllt. Für (II) nehmen wir $N \in \mathcal{C}_M$ an und müssen $g(N) \in \mathcal{C}_M$ nachweisen. Wir unterscheiden drei Fälle: Im ersten Fall ist $g(M) \subseteq N$. Wir erhalten $g(M) \subseteq g(N)$ und also $g(N) \in \mathcal{C}_M$. Im zweiten Fall ist $N = M$. Wir erhalten $g(M) \subseteq g(N)$ mit Gleichheit und also wieder $g(N) \in \mathcal{C}_M$. Im letzten Fall ist $N \subsetneq M$ eine echte Teilmenge. Wegen $M \in \mathcal{C}$ können wir zwei Unterfälle betrachten: Im ersten ist $g(N) \subseteq M$ und also abermals $g(N) \in \mathcal{C}_M$. Im verbliebenen Fall ist $M \subsetneq g(N)$ eine echte Teilmenge. Dann hat aber N mindestens zwei Elemente weniger als $g(N)$, was der Definition von g widerspricht. Dieser letzte Fall kann also nicht eintreten. ◇

Aus der eben gezeigten Behauptung folgt $\bigcup \mathcal{G}_0 \in \mathcal{G}_0$ und dann $g(\bigcup \mathcal{G}_0) \in \mathcal{G}_0$. Wegen der Definition der Vereinigung gilt also $g(\bigcup \mathcal{G}_0) \subseteq \bigcup \mathcal{G}_0$. Für $M = \bigcup \mathcal{G}_0$ hat man somit $g(M) = M$, wie gewünscht.

Zum Abschluss leiten wir den allgemeinen Fall ab, in dem aus $M \subseteq N \in \mathcal{Z}$ nicht unbedingt $M \in \mathcal{Z}$ folgt. Sei $\hat{\mathcal{Z}}$ die Menge aller Ketten in $\mathcal{Z}$. Um zu prüfen, dass $\hat{\mathcal{Z}}$ wiederum induktiv ist, betrachten wir eine Kette $\mathcal{L} \subseteq \hat{\mathcal{Z}}$. Für beliebige $M_0, M_1 \in \bigcup \mathcal{L}$ findet man $L_i \in \mathcal{L}$ mit $M_i \in L_i$. Da $\mathcal{L}$ ein Kette ist, können

[40] Dass man all diese Elemente als Werte einer Funktion f betrachten kann garantiert das sogenannte Auswahlaxiom, das wir später genauer diskutieren.

[41] Dieser Begriff ist nicht unbedingt standard und in der Tat eher eine temporäre Hilfskonstruktion. Er liefert uns einen Ersatz für die unendlichen Zahlen (sogenannte Ordinalzahlen) aus der Mengenlehre. Die Eigenschaften (I) und (II) entsprechen den Tatsachen, dass wir eine erste Ordinalzahl Null haben und für jede Zahl α den Nachfolger $\alpha + 1$ bilden können. Eigenschaft (III) besagt, dass wir auch nach einer unendlichen Kette von Zahlen (man denke an $0 < 1 < 2 < \ldots$) noch weiterzählen können. Die Idee unseres Beweises ist, dass wir – beginnend mit $M = \emptyset$ – so lange Elemente $f(M)$ hinzufügen, bis es keine geeigneten Elemente mehr gibt – sodass dann eben M maximal ist. Mit dem Hinzufügen sind wir dabei nicht unbedingt fertig, nachdem wir entland von $\mathbb{N}$ iteriert haben – weswegen wir über die natürlichen Zahlen hinaus eben noch die transfiniten Ordinalzahlen benötigen. Mag diese Erklärung den Beweis (hoffentlich) transparenter machen, so ist doch auch zu betonen, dass wir unabhängig von unserer Intuition über unendliche Zahlen einen formal korrekten Beweis erhalten.

wir $L_0 \subseteq L_1$ annehmen,[42] sodass $M_0, M_1 \in L_1$ gilt. Da nun L_i eine Kette ist, hat man $M_0 \subseteq M_1$ oder $M_1 \subseteq M_0$. Daher ist $\bigcup \mathcal{L}$ eine Kette und also ein Element von $\hat{\mathcal{Z}}$, wie gewünscht. Weiter sieht man, dass aus $\mathcal{L} \subseteq \mathcal{L}' \in \hat{\mathcal{Z}}$ stets $\mathcal{L} \in \hat{\mathcal{Z}}$ folgt, sodass die Zusatzannahme vom Anfang des Beweises für $\hat{\mathcal{Z}}$ erfüllt ist. Wir bekommen nun ein maximales Element L von $\hat{\mathcal{Z}}$. Dann ist aber $\bigcup L$ ein maximales Element von $\mathcal{Z}$. Hätte man nämlich ein $N \in \mathcal{Z}$ mit $\bigcup L \subsetneq M$, so erhielte man $L \subsetneq L \cup \{M\} \in \hat{\mathcal{Z}}$. □

Im Rest dieses Kapitels diskutieren wir Kardinalitäten, also Größenvergleiche zwischen Mengen.

Definition 2.2.4. Für Mengen M und N schreiben wir $|M| \leq |N|$ bzw. $|M| = |N|$, wenn es eine injektive bzw. bijektive Funktion von M nach N gibt.

Für endliche M hatten wir $|M|$ bereits zuvor durch einen intuitiven Begriff von Anzahl erklärt. Dieser Begriff lässt sich präzisieren, indem man $|M|$ als die eindeutige Zahl $n \in \mathbb{N}$ charakterisiert, welche eine Bijektion zwischen M und $\{0, \dots, n-1\}$ zulässt. Analog werden in der Mengenlehre auch unendlichen Mengen M sogenannte Kardinalzahlen $|M|$ zugeordnet. Hingegen ist in unserer Darstellung der alleinstehende Ausdruck $|M|$ für unendliche M nicht erklärt. Er ist für uns nur als Teil von Ausdrücken wie $|M| \leq |N|$ sinnvoll. Sind M und N endlich, so haben wir nun allerdings zwei unterschiedliche Interpretationen von Ausdrücken wie $|M| \leq |N|$. Diese sind aber äquivalent, weil eine Injektion $\{0, \dots, m\} \to \{0, \dots, n\}$ genau für $m \leq n$ existiert (sogenanntes Taubenschlagprinzip). Durch das folgende Ergebnis überzeugen wir uns insbesondere, dass die Notation aus der vorigen Definition sinnvoll ist. Hierbei ist $|M| \geq |N|$ synonym zu $|N| \leq |M|$.

Theorem 2.2.5 (Satz von Schröder-Bernstein[43])**.** *Aus* $|M| \leq |N|$ *und* $|M| \geq |N|$ *folgt zusammen* $|M| = |N|$.

Beweis. Betrachte Injektionen $f : M \to N$ und $g : N \to M$. Wir werden Mengen $A \subseteq M$ und $B \subseteq N$ konstruieren mit[44]

$$f[A] = B \quad \text{und} \quad g[N \backslash B] = M \backslash A.$$

Eine Bijektion $h : M \to N$ ergibt sich – wie man prüfen möge – durch

$$h(x) = \begin{cases} f(x) & \text{für } x \in A, \\ y & \text{für } x = g(y) \in M \backslash A. \end{cases}$$

Um die Mengen A und B zu finden, betrachten wir $\Phi : \mathcal{P}(M) \to \mathcal{P}(M)$ mit

$$\Phi(X) = M \backslash g\big[N \backslash f[X]\big].$$

Wir werden zeigen, dass es einen Fixpunkt $A = \Phi(A)$ gibt. Setzen wir für diesen dann $B := f[A]$, so erhalten wir

$$A = \Phi(A) = M \backslash g[N \backslash B].$$

[42]Der Fall $L_1 \subseteq L_0$ ist nämlich komplett analog. In manchen Texten heißt es in so einem Fall, dass $L_0 \subseteq L_1$ ohne Beschränkung der Allgemeinheit (o.B.d.A.) gilt.

[43]Dieses Ergebnis war von Georg Cantor formuliert worden, bevor dann verschiedene Forschende – insbesondere Richard Dedekind, Felix Bernstein und Ernst Schröder – teils unabhängig voneinander Beweise vorgeschlagen haben.

Indem wir auf beiden Seiten das Komplement nehmen, sehen wir $M\backslash A = g[N\backslash B]$, wie oben gewünscht. Um einen Fixpunkt zu finden, folgern wir aus $X \subseteq Y$ zunächst $f[X] \subseteq f[Y]$ und dann $N\backslash f[X] \supseteq N\backslash f[Y]$, was uns schließlich auf

$$X \subseteq Y \quad \Rightarrow \quad \Phi(X) \subseteq \Phi(Y)$$

führt.[45] Man betrachte

$$A := \bigcap \{X \subseteq M : \Phi(X) \subseteq X\},$$

wobei der Schnitt wegen $\Phi(M) \subseteq M$ durch eine nicht-leere Menge indiziert ist. Für beliebiges $X \subseteq M$ mit $\Phi(X) \subseteq X$ gilt $A \subseteq X$ (wegen der Definition des Schnitts) und somit $\Phi(A) \subseteq \Phi(X) \subseteq X$ (wegen der Monotonie von Φ). Man erhält daher $\Phi(A) \subseteq A$ (wieder mit der Definition des Schnitts). Hiermit folgt aber auch $\Phi(\Phi(A)) \subseteq \Phi(A)$ (nochmal Monotonie) und damit $A \subseteq \Phi(A)$ (nochmal Definition des Schnitts), also insgesamt $A = \Phi(A)$. □

Insbesondere erhalten wir die in der folgenden Definition angeführte Äquivalenz.

Definition 2.2.6. Wir schreiben $|M| < |N|$, wenn $|M| \leq |N|$ gilt aber $|M| = |N|$ (oder nach Schröder-Bernstein äquivalent auch $|M| \geq |N|$) falsch ist.

Das folgende Ergebnis besagt, dass Kardinalitäten linear geordnet sind.

Proposition 2.2.7. *Es gilt stets* $|M| \leq |N|$ *oder* $|M| \geq |N|$.

Beweis. Sei $\mathcal{Z}$ die Menge aller Injektionen $f : M_f \to N$ mit $M_f \subseteq M$, wobei wir eine solche Funktion f mit ihrem Graph

$$\{(x, f(x)) : x \in M_f\}$$

identifizieren. Jede Teilmenge $g \subseteq f$ ist dann wiederum der Graph einer Funktion $g : M_g \to N$ mit $M_g \subseteq M_f$ und $g(x) = f(x)$ für alle $x \in M_g$. Man kann folgern, dass $\mathcal{Z}$ induktiv ist.[46] Das Zornsche Lemma liefert also ein maximales Element f von $\mathcal{Z}$. Aus der Maximalität folgt, dass einer der folgenden beiden Fälle eintritt: Im ersten Fall ist $M_f = M$. Dann ist f ein Zeuge für $|M| \leq |N|$. Im zweiten Fall ist f surjektiv. Dann gibt es also eine Rechtsinverse $g : N \to M_f \subseteq M$. Diese ist selbst injektiv mit Linksinverser f, bezeugt also $|N| \leq |M|$. □

Aus der vorigen Proposition und dem Satz von Schröder-Bernstein ergibt sich, dass für beliebige M und N entweder $|M| < |N|$ oder $|M| = |N|$ oder $|M| > |N|$ gilt.[47] Die folgenden Größenvergleiche sind besonders wichtig.

---

[44]Wir schreiben $f[A] = \{f(x) : x \in A\}$ für das Bild von $A \subseteq M$ unter $f : M \to N$. Später werden wir auch $f^{-1}[B] = \{x \in A : f(x) \in B\}$ für das Urbild von $B \subseteq N$ schreiben. Dies ist nicht mit der Notation für die Umkehrfunktion zu verwechseln, weswegen wir hier eckige statt runde Klammern verwenden (wobei viele andere Texte für beide Fälle runde Klammern wählen).

[45]Eine Funktion $\Phi : \mathcal{P}(M) \to \mathcal{P}(M)$ mit dieser Eigenschaft nennt man auch monotonen Operator auf M. Der Satz von Knaster-Tarski (benannt nach Bronisław Knaster und Alfred Tarski) besagt, dass jeder solche Operator einen (kleinsten und größten) Fixpunkt hat. In der Tat funktioniert der Rest unseres Beweises für allgemeines Φ. Die Anwendung auf den Satz von Schröder-Bernstein geht ebenfalls zurück auf B. Knaster, Un théorème sur les fonctions d'ensembles, Annales de la Société Polonaise de Mathématique 6 (1928) 133-134.

[46]Hier möge man sich insbesondere überzeugen, dass die Vereinigung über eine Kette wieder einen Funktionsgraphen ergibt. Die entscheidende Eigenschaft einer Funktion ist dabei, dass für Tupel (x, y) und (x, y') im Graph stets $y = y'$ gelten muss, dass ein und demselben Argument x also nur ein Wert zugeordnet wird.

[47]Wir hatten zuvor angemerkt, dass man ‚oder' im Normalfall inklusiv zu verstehen hat, d.h. dass auch mehrere der genannten Optionen gleichzeitig eintreten können. Hier und in vielen anderen

Definition 2.2.8. Eine Menge M ist endlich bzw. abzählbar unendlich bzw. überabzählbar, wenn $|M| < |\mathbb{N}|$ bzw. $|M| = |\mathbb{N}|$ bzw. $|M| > |\mathbb{N}|$ gilt.[48]

Das nächste Ergebnis versöhnt zwei intuitive Begriffe von Endlichkeit.

Lemma 2.2.9. *Ist M endlich, so gilt $|M| = \{0, \dots, n-1\}$ für ein $n \in \mathbb{N}$.*

Beweis. Wegen der Endlichkeit gibt es insbesondere eine Injektion $f : M \to \mathbb{N}$. Definiere $c : \operatorname{img}(f) \to \mathbb{N}$ rekursiv[49] durch

$$c(m) = \min\{n \in \mathbb{N} : c(k) \neq n \text{ für alle } k < m\}.$$

Diese Funktion kann nicht surjektiv sein: Sonst wäre es auch $c \circ f : M \to \mathbb{N}$ und man hätte $|M| \geq |\mathbb{N}|$, wie im vorigen Beweis. Wir können also das minimale $n \in \mathbb{N}$ mit $n \notin \operatorname{img}(c)$ betrachten.[50] Per Konstruktion ist c injektiv mit $c(l) < n$ für alle $l \in \operatorname{img}(f)$. Dann ist also $c \circ f$ eine Bijektion zwischen M und $\{0, \dots, n-1\}$. □

Das Folgende liefert insbesondere die Existenz überabzählbarer Mengen.

Proposition 2.2.10 (Cantor)**.** *Es gilt stets $|M| < |\mathcal{P}(M)|$.*[51]

Beweis. Eine Injektion $\iota : M \to \mathcal{P}(M)$ ist durch $\iota(x) = \{x\}$ gegeben. Für einen Widerspruch nehmen wir an, dass es auch eine Injektion $f : \mathcal{P}(M) \to M$ gibt. Man bilde nun die Menge

$$D := \{f(N) \,|\, N \subseteq M \text{ und } f(N) \notin N\} \subseteq M.$$

Wir erhalten die widersprüchliche Äquivalenz

$$f(D) \notin D \quad \Leftrightarrow \quad f(D) \in D.$$

Die Richtung von links nach rechts ergibt sich hier direkt aus der Definition von D. Für die andere Richtung nehmen wir $f(D) \in D$ an. Dann ist also $f(D) = f(N)$ für ein $N \subseteq M$ mit $f(N) \notin N$. Die Injektivität liefert $D = N$ und also $f(D) \notin D$. □

Wir haben bereits gesehen, dass Kardinalitäten linear geordnet sind: Das ist gerade der Inhalt von Theorem 2.2.5 und Proposition 2.2.7. Zum Abschluss erwähnen wir, dass die Kardinalitäten eine Wohlordnung bilden:

Proposition 2.2.11. *Es gibt keine unendliche Folge von Mengen M_i mit*

$$|M_0| > |M_1| > |M_2| > \dots.$$

Texten wird der Zusatz ‚entweder' verwendet, um eine Ausnahme von dieser Regel zu markieren, sodass dann also genau eine der Optionen gelten soll.

[48]Man verwendet ‚abzählbar' oft für ‚endlich oder abzählbar unendlich', teils aber auch synonym zu ‚abzählbar unendlich'.

[49]Dies funktioniert analog zur (vollständigen) Induktion: Um $g : N \to Y$ mit $N \subseteq \mathbb{N}$ zu definieren, braucht man nur anzugeben, wie $g(n)$ jeweils aus den vorigen Werten $g(m)$ mit $m < n$ hervorgeht. Oft hat man $N = \mathbb{N}$ und spezifiziert zunächst $g(0)$ sowie dann $g(n+1)$ in Abhängigkeit von $g(n)$. Beispielsweise kann die Funktion $g : \mathbb{N} \to \mathbb{N}$ mit $g(n) = k + n$ für festes $k \in \mathbb{N}$ durch $g(0) = k$ und $g(n+1) = g(n) + 1$ gegeben werden. Formal kann man das Rekursionsprinzip rechtfertigen, indem man die Existenz und Eindeutigkeit der Mengen $\{(m, g(m)) : m < n\}$ per Induktion über n beweist. Für Details (auch zur später verwendeten Rekursion über beliebige Wohlordnungen) verweisen wir auf Kapitel 18 in Halmos, Naive Set Theory (vgl. Fußnote 39).

[50]Die Existent solch eines minimalen Elements ist tatsächlich äquivalent zum Induktionsprinzip: Gäbe es kein minimales Element, so würde aus $\{0, \dots, m-1\} \subseteq \operatorname{img}(f)$ stets auch $m \in \operatorname{img}(f)$ folgen. Per Induktion erhielte man also $m \in \operatorname{img}(f)$ für alle $m \in \mathbb{N}$.

[51]Mit $\mathcal{P}(M)$ bezeichnet man die Potenzmenge – also die Menge aller Teilmengen – einer Menge M.

Beweisskizze.[52] Mit einer etwas allgemeineren Version des Zornschen Lemmas (vgl. Fußnote 39) erhält man eine Wohlordnung $\prec$ auf M_0. Haben wir $|M_0| \geq |M_n|$, so erhalten wir Injektionen $f_n : M_n \to M_0$, deren Bilder jeweils Anfangssegmente von $\prec$ sind (ähnlich wie im Beweis von Lemma 2.2.9 aber nun mit der transfiniten Erweiterung des Rekursionsprinzips, vgl. Fußnote 49). Gilt sogar $|M_0| > |M_n|$, so müssen diese von M_0 verschieden und also von der Form

$$\operatorname{img}(f_n) = \{x \in M_0 : x \prec y_n\}$$

für Elemente $y_n \in M$ sein. Da $\prec$ eine Wohlordnung ist, kann nicht $y_0 \succ y_1 \succ \dots$ gelten. Es muss also eine Zahl $n \in \mathbb{N}$ mit $y_n \preceq y_{n+1}$ geben. Für diese hat man dann $\operatorname{img}(f_n) \subseteq \operatorname{img}(f_{n+1})$, woraus wir $|M_n| \leq |M_{n+1}|$ erhalten. □

2.3. **Dimension.** In diesem Unterkapitel zeigen wir, dass alle Basen desselben Vektorraums dieselbe Kardinalität haben, welche man dann als Dimension des Vektorraums bezeichnet. Die beiden vorigen Abschnitte zusammen liefern unmittelbar:

Korollar 2.3.1. *Jeder Vektorraum hat eine Basis.*

Beweis. Wegen Lemma 2.2.2 liefert das Zornsche Lemma eine maximal linear unabhängige Menge, die dann nach Proposition 2.1.10 eine Basis ist. □

Man benötigt manchmal auch das folgende allgemeinere Ergebnise. Das vorige Korollar ist der Spezialfall mit $M = \emptyset$ und $N = V$.

Proposition 2.3.2 (Austauschsatz von Steinitz[53])**.** *Wir betrachten einen Vektorraum V sowie eine linear unabhängige Menge $M \subseteq V$ und ein Erzeugendensystem $N \subseteq V$. Es gibt dann ein $N' \subseteq N$, sodass $M \cup N'$ eine Basis von V bildet.*

Beweis. Man betrachte die Menge

$$\mathcal{Z} = \{L \subseteq N : M \cup L \text{ ist linear unabhängig}\}.$$

Wie im Beweis von Lemma 2.2.2 ist diese induktiv (wobei man für den Fall der leeren Kette hier noch die lineare Unabhängigkeit von M nutzen muss). Das Zornsche Lemma liefert also ein maximales Element N' von $\mathcal{Z}$. Für einen Widerspruch nehmen wir an, dass $M \cup N'$ keine Basis und hier also kein Erzeugendensystem ist. Es gilt somit

$$\operatorname{Span}(N) = V \not\subseteq \operatorname{Span}(M \cup N').$$

Gemäß Lemma 2.1.3 folgt $N \not\subseteq \operatorname{Span}(M \cup N')$, d.h. wir finden einen Vektor $v \in N$ mit $v \notin \operatorname{Span}(M \cup N')$. Wie im Beweis von Proposition 2.1.10 schließen wir, dass $M \cup N' \cup \{v\}$ linear unabhängig ist. Dann hat man also $N' \subsetneq N' \cup \{v\} \in \mathcal{Z}$, entgegen der Maximalität von N'. □

Das folgende scheinbar technische Lemma impliziert bereits, dass in endlichdimensionalen Vektorräumen alle Basen die gleiche Kardinalität haben (betrachte hierfür zwei Basen $M = A$ und $N = B$).[54]

---

[52] Um diese Skizze vollständig zu verstehen, muss ggf. weitere Literatur konsultiert werden. Der Autor schließt die Skizze daher vom Stoff für seine Prüfungen zur Linearen Algebra I und II aus.

[53] Dieser Satz ist benannt nach Ernst Steinitz.

[54] Tatsächlich ist das Lemma äquivalent zu dieser Aussage. Wie wir später sehen werden, korrespondieren A und B nämlich zu Basen des sogenannten Faktorraumes $\operatorname{Span}(M)/\operatorname{Span}(M \setminus A)$.

Lemma 2.3.3. *Betrachte linear unabhängige Mengen M und N sowie Teilmengen $A \subseteq M$ und $B \subseteq N$ mit*

$$\operatorname{Span}(M\backslash A) = \operatorname{Span}(N\backslash B) \quad \textit{und} \quad \operatorname{Span}(M) = \operatorname{Span}(N).$$

Ist A endlich, so gilt $|A| = |B|$.

Beweis. Wir argumentieren zunächst unter der zusätzlichen Annahme, dass die Mengen A und B disjunkt sind. Wähle dann ein maximales $A_0 \subseteq A$, für das es ein $B_0 \subseteq B$ mit folgenden Eigenschaften gibt:

(i) man hat $|A_0| = |B_0|$,
(ii) die Menge $(N\backslash B_0) \cup A_0$ ist linear unabhängig,
(iii) es gilt $\operatorname{Span}(M) = \operatorname{Span}\big((N\backslash B_0) \cup A_0\big)$.

Man beachte, dass dies von $A_0 = \emptyset = B_0$ erfüllt wird. Ein maximales Beispiel existiert wegen der Endlichkeit von A (genau genommen per Induktion), ohne den Einsatz des Zornschen Lemmas. Wegen der linearen Unabhängigkeit von M hat man $A \cap \operatorname{Span}(M\backslash A) = \emptyset$. Mit der Zusatzannahme folgt $(N\backslash B_0) \cap A_0 = \emptyset$.

Wir nehmen zunächst $A_0 = A$ an. Mit Lemma 2.1.3 sieht man

$$\operatorname{Span}(M) = \operatorname{Span}\big((N\backslash B) \cup A\big).$$

Wegen (iii) und der linearen Unabhängigkeit aus (ii) muss $B_0 = B$ gelten. Mit (i) folgt also $|A| = |A_0| = |B_0| = |B|$, wie gewünscht.

Für einen Widerspruch nehmen wir nun $A_0 \subsetneq A$ an. Wähle ein $v \in A\backslash A_0$. Wegen der linearen Unabhängigkeit von M gilt

$$v \notin \operatorname{Span}\big((M\backslash A) \cup A_0\big) = \operatorname{Span}\big((N\backslash B) \cup A_0\big).$$

Mit (iii) hat man also eine Darstellung

$$v = \sum_{i=1}^{n} \lambda_i v_i \quad \text{mit } v_i \in (N\backslash B_0) \cup A_0,$$

wobei man $v_n \in B\backslash B_0$ und $\lambda_n \neq 0$ annehmen kann. Für die Mengen $A_1 = A_0 \cup \{v\}$ und $B_1 = B_0 \cup \{v_n\}$ erhält man

$$v_n = \frac{1}{v_n} \cdot v + \sum_{i=1}^{n-1} -\frac{\lambda_i}{\lambda_n} \cdot v_i \in \operatorname{Span}\big((N\backslash B_1) \cup A_1\big).$$

Es ergibt sich

$$\operatorname{Span}\big((N\backslash B_0) \cup A_0\big) = \operatorname{Span}\big((N\backslash B_1) \cup A_1\big),$$

sodass Eigenschaft (iii) auch für A_1 und B_1 erfüllt ist. Gleiches gilt für (i) und (ii), was dann aber der Maximalität von A_0 widerspricht.

Wir leiten nun den Fall ab, in dem $C := A \cap B$ nicht leer ist. Hier betrachten wir $A' := A\backslash C$ und $B' := B\backslash C$, sodass also $A' \cap B' = \emptyset$ gilt. Wegen $M\backslash A' = (M\backslash A) \cup C$ und $N\backslash B' = (N\backslash B) \cup C$ haben wir $\operatorname{Span}(M\backslash A') = \operatorname{Span}(M\backslash B')$. Mit dem bereits gezeigten Fall ist also $|A'| = |B'|$ und somit $|A| = |A'| + |C| = |B'| + |C| = |B|$. □

Wir leiten nun das Resultat für allgemeine Dimension ab.

Theorem 2.3.4. *Sind A und B Basen desselben Vektorraums, so gilt $|A| = |B|$.*

Beweis.[55] Man betrachte die Menge

$$\mathcal{Z} := \{f : A_f \to B \mid A_f \subseteq A \text{ und } f \text{ ist injektiv und } \operatorname{Span}(A_f) = \operatorname{Span}(f[A_f])\},$$

wobei jede Funktion f mit ihrem Graph identifiziert wird (wie im Beweis von Proposition 2.2.7). Um zu zeigen, dass $\mathcal{Z}$ induktiv ist, betrachten wir eine Kette $\mathcal{L} \subseteq \mathcal{Z}$. Wie im eben zitierten Beweis liefert $g := \bigcup \mathcal{L}$ eine Injektion

$$g : A_g = \bigcup\{A_f : f \in \mathcal{L}\} \to B,$$

welche $g(v) = f(v)$ für alle $f \in \mathcal{L}$ und $v \in A_f$ erfüllt. Insbesondere hat man also jeweils $g[A_f] = f[A_f]$. Wir rechnen

$$\begin{aligned} A_g \subseteq \bigcup\{\operatorname{Span}(A_f) : f \in \mathcal{L}\} &= \bigcup\{\operatorname{Span}(f[A_f]) : f \in \mathcal{L}\} \\ &\subseteq \operatorname{Span}\left(\bigcup\{g[A_f] : f \in \mathcal{L}\}\right) = \operatorname{Span}(g[A_g]) \end{aligned}$$

sowie außerdem

$$g[A_g] \subseteq \bigcup\{\operatorname{Span}(g[A_f]) : f \in \mathcal{L}\} = \bigcup\{\operatorname{Span}(A_f) : f \in \mathcal{L}\} \subseteq \operatorname{Span}(A_g).$$

Mit Lemma 2.1.3 folgt $\operatorname{Span}(A_g) = \operatorname{Span}(g[A_g])$ und also $g = \bigcup \mathcal{L} \in \mathcal{Z}$.

Das Zornsche Lemma liefert nun ein maximales Element $f : A_f \to B$ von $\mathcal{Z}$. Gilt $A_f = A$, so ist $\operatorname{Span}(A) = \operatorname{Span}(f[A])$ der ganze Vektorraum. Mit Proposition 2.1.10 folgt $f[A] = B$, sodass f also $|A| = |B|$ bezeugt. Um den Beweis abzuschließen, nehmen wir $A_f \neq A$ an und leiten einen Widerspruch her. Wähle Teilmengen $A_f = A_0 \subsetneq A_1 \subseteq A_2 \subseteq \ldots$ von A und Teilmengen $f[A_f] = B_0 \subseteq B_1 \subseteq \ldots$ von B so, dass jeweils

$$\operatorname{Span}(A_n) \subseteq \operatorname{Span}(B_n) \subseteq \operatorname{Span}(A_{n+1})$$

gilt und die Mengen $A_{n+1} \backslash A_n$ und $B_{n+1} \backslash B_n$ endlich sind. Man hat dann

$$\operatorname{Span}(A_\omega) = \operatorname{Span}(B_\omega) \quad \text{für } A_\omega = \textstyle\bigcup_{n\in\mathbb{N}} A_n \text{ und } B_\omega = \bigcup_{n\in\mathbb{N}} B_n.$$

Ist $A_\omega \backslash A_0$ oder $B_\omega \backslash B_0$ endlich, so liefert das vorige Lemma

$$|A_\omega \backslash A_0| = |B_\omega \backslash B_0|.$$

Anderenfalls gilt diese Gleichung ebenso, weil dann nämlich $A_\omega \backslash A_0$ und $B_\omega \backslash B_0$ abzählbar unendlich sind. Man erhält daher eine Funktion $g : A_\omega \to B$ mit $g \in \mathcal{Z}$ und $f \subsetneq g$, was der Maximalität von f widerspricht. □

Das vorige Theorem und Proposition 2.3.1 garantieren die Wohldefiniertheit des folgenden zentralen Begriffs.

Definition 2.3.5. Die Dimension $\dim(V)$ eines Vektorraums V ist die Mächtigkeit einer beliebigen Basis von V.[56]

Laut dem folgenden Beispiel stimmt die übliche Intuition (man denke insbesondere an den Euklidischen Raum $\mathbb{R}^3$) mit dem formalen Begriff überein.

---

[55] Wir orientieren uns an einem Beweis von Justin Tatch Moore, A Zorn's Lemma Proof of the Dimension Theorem for Vector Spaces, The American Mathematical Monthly 121:3 (2014) 260-262.

[56] Sind die Basen von V endlich, so liefert uns dies eine Zahl $\dim(V) \in \mathbb{N}$. Im unendlichen Fall haben wir $\dim(V)$ gemäß dem Absatz nach Definition 2.2.4 nicht als Objekt definiert, können aber Ausdrücken wie $\dim(V) = \dim(W)$ einen präzisen Sinn abgewinnen: Die gegebene Gleichung besagt, dass es für Basen $B \subseteq V$ und $C \subseteq W$ eine Bijektion $f : B \to C$ gibt.

Beispiel 2.3.6. Man hat $\dim(K^n) = n$. Die Einheitsvektoren aus Definition 1.3.5 bilden nämlich eine Basis („Standardbasis"), wie aus Lemma 1.3.7 ersichtlich ist.

Das folgende Resultat ist die im Titel des Kapitels versprochene Klassifikation.

Theorem 2.3.7. *Für Vektorraüme V und W über einem gemeinsamen Körper gilt*

$$V \cong W \quad \Leftrightarrow \quad \dim(V) = \dim(W).$$

Beweis. Sei zunächst ein Isomorphismus $\Phi : V \to W$ gegeben. Mit Korollar 2.3.1 wählen wir eine Basis $B \subseteq V$. Proposition 2.1.12 liefert eine Funktion $f : B \to W$ mit $\mathrm{Abb}(f) = \Phi$. Laut Proposition 2.1.13 ist f injektiv und hat eine Basis von W zum Bild. Dies liefert $\dim(V) = |B| = |\operatorname{img}(f)| = \dim(W)$.

Gilt umgekehrt $\dim(V) = \dim(W)$, so hat man eine Bijektion $f : B \to C$ zwischen Basen $B \subseteq V$ und $C \subseteq W$. Man kann f auch als Funktion mit Wertebereich W und Bild C auffassen. Wieder mit Proposition 2.1.13 ist dann also die Abbildung $\mathrm{Abb}(f) : V \to W$ ein Isomorphismus. □

Wir betrachten auch das folgende verwandte Ergebnis.

Proposition 2.3.8. *Es gilt* $\dim(V) \leq \dim(W)$ *genau dann, wenn es eine injektive lineare Abbildung von V nach W gibt.*

Beweis. Man argumentiert wie im vorigen Beweis, wobei ein zusätzlicher Schritt zu beachten ist: Hat man eine Injektion $f : B \to W$ für eine Basis $B \subseteq V$ und linear unabhängiges Bild $\operatorname{img}(f) \subseteq W$, so findet man wegen des Austauschsatzes von Steinitz eine Basis $C \subseteq W$ mit $\operatorname{img}(f) \subseteq C$, sodass also $|B| \leq |C|$ gilt. □

Wegen der beiden vorigen Resultate übertragen sich unsere Ergebnisse über Kardinalitäten direkt auf Vektorräume:

Bemerkung 2.3.9. Hat man eine injektive lineare Abbildung $V \to W$, so ist V isomorph zu einem sogenannten Untervektorraum von W, wie wir im nächsten Teilkapitel sehen werden. Mit Proposition 2.2.7 folgt dann, dass von zwei beliebigen Vektorräumen über demselben Körper stets einer isomorph zu einem Untervektorraum des anderen ist. Der Satz von Schröder-Bernstein impliziert, dass zwei Vektorräume isomorph sind, wenn jeder isomorph zu einem Untervektorraum des anderen ist.[57]

Für endliche Dimension haben wir mit den Vektorräumen K^n sehr konkrete Repräsentanten. Diese lassen sich wie folgt verallgemeinern.

Definition 2.3.10. Für einen Körper K und eine Menge M definieren wir $\bigoplus_M K$ als die Menge der Funktionen $f : M \to K$, für die der Träger

$$\operatorname{supp}(f) := \{x \in M \mid f(x) \neq 0\}$$

endlich ist. Wir definieren $\cdot : K \times \bigoplus_M K \to \bigoplus_M K$ und $+ : \bigoplus_M K \times \bigoplus_M K \to \bigoplus_M K$ durch $(\lambda \cdot f)(x) = \lambda \cdot f(x)$ und $(f + g)(x) = f(x) + g(x)$.

[57]Beide Eigenschaften sind in anderen Kontexten nicht erfüllt: So ist keiner der Körper $\mathbb{Q}$ und $\mathbb{F}_2$ (vgl. Fußnote 17) isomorph zu einem Teilkörper des anderen. Dies sollte auch ohne eine Definition von Körperhomomorphismen intuitiv sein, weil $1 + 1 = 0$ in $\mathbb{F}_2$ aber nicht in $\mathbb{Q}$ gilt. Weiter ist jede der geordneten Mengen $(0, \infty)$ und $[0, \infty)$ isomorph zu einer Teilmenge der anderen (etwa durch $f : [0, \infty) \to [1, \infty)$ mit $f(x) = 1 + x$). Hingegen sind die beiden nicht isomorph (d.h. es gibt keine strikt monotone Bijektion zwischen ihnen), da nur eine ein minimales Element hat.

Man überzeuge sich (analog zu Beispiel 1.2.14), dass $\bigoplus_M K$ mit den angegebenen Operationen ein K-Vektorraum ist. Fasst man n-Tupel als Funktionen mit Definitionsbereich $\{0,\dots,n-1\}$ auf (vgl. Fußnote 3), so stimmt $\bigoplus_{\{0,\dots,n-1\}} K$ mit dem vertrauten Vektorraum K^n überein, d.h. die neue Konstruktion verallgemeinert in der Tat die vorige. Das folgende Ergebnis zeigt insbesondere auch, dass es zu jeder gegebenen Kardinalität einen Vektorraum mit entsprechender Dimension gibt.

Proposition 2.3.11. *Für jeden Körper K gilt:*

(a) Es ist $\dim\left(\bigoplus_M K\right) = |M|$ für jede Menge M. Insbesondere gilt

$$\bigoplus_M K \cong \bigoplus_N K \quad \Leftrightarrow \quad |M| = |N|.$$

(b) Jeder K-Vektorraum ist isomorph zu einem von der Form $\bigoplus_M K$.

Beweis. (a) Für $x \in M$ betrachten wir

$$\delta_x : M \to K \quad \text{mit} \quad \delta_x(y) = \begin{cases} 1 & \text{wenn } y = x, \\ 0 & \text{sonst.} \end{cases}$$

Sind $x(1),\dots,x(n) \in M$ paarweise verschieden und hat man $\lambda_1,\dots,\lambda_n \in K\backslash\{0\}$, so ist dann $f := \lambda_1 \cdot \delta_{x(1)} + \ldots + \lambda_n \cdot \delta_{x(n)}$ die Funktion mit

$$\operatorname{supp}(f) = \{x(1),\dots,x(n)\} \quad \text{und} \quad f(x(i)) = \lambda_i.$$

Daraus ist ersichtlich, dass $\{\delta_x \,|\, x \in M\}$ eine Basis von $\bigoplus_M K$ ist. Es gilt also

$$\dim\left(\bigoplus_M K\right) = |\{\delta_x \,|\, x \in M\}| = |M|.$$

Mit Theorem 2.3.7 folgt auch die angegebene Äquivalenz.

(b) Für einen gegebenen K-Vektorraum V wähle man mit Korollar 2.3.1 eine Basis B. Dank Teil (a) hat man $\dim(V) = |B| = \dim(\bigoplus_B K)$. Theorem 2.3.7 liefert dann $V \cong \bigoplus_B K$. □

Die vorangegangenen Betrachtungen könnten einen wohl zu der Ansicht verleiten, die Theorie der Vektorräume gehe nicht wesentlich über die Theorie der Mächtigkeiten hinaus. Dies wäre freilich ein Trugschluss: Wir werden in den nächsten Kapiteln sehen, dass insbesondere die Strukturen der Untervektorräume und der linearen Abbildungen zwischen Vektorräumen wesentlich reicher sind als die der Teilmengen bzw. der Funktionen zwischen Mengen.

Die nächsten Teilkapitel werden ebenfalls offenbaren, dass der Begriff der Dimension für verschiedene Anwedungen von zentraler Bedeutung ist – wobei wir uns meist auf den endlichdimensionalen Fall konzentrieren werden. Zum Abschluss dieses Teilkapitels skizzieren wir eine (innermathematische) Anwendung unserer Klassifikation, bei der die relevanten Dimensionen unendlich sind:

Bemerkung 2.3.12. Die $\mathbb{R}$-Vektorräume $\mathbb{R} = \mathbb{R}^1$ und $\mathbb{R}^2$ sind wegen

$$\dim(\mathbb{R}) = 1 \neq 2 = \dim(\mathbb{R}^2)$$

nicht isomorph. Dabei bleibt aber offen, ob $(\mathbb{R},+)$ und $(\mathbb{R}^2,+)$ als Gruppen isomorph sind (ob es also eine Bijektion $h : \mathbb{R} \to \mathbb{R}^2$ mit $h(x+y) = h(x)+h(y)$ gibt). Dies ist insbesondere deshalb interessant, weil $(\mathbb{R}^2,+)$ mit der additiven Struktur der komplexen Zahlen $\mathbb{C}$ übereinstimmt (wobei wir $\mathbb{C}$ erst später in diesem Kurs behandeln werden). Hierfür ist es hilfreich, $\mathbb{R}$ und $\mathbb{R}^2$ als Vektorräume über $\mathbb{Q}$ zu betrachten. Wie wir im folgenden kurz skizzieren, haben sie dann nämlich beide die

Dimension $|\mathbb{R}|$.[58] Wegen Theorem 2.3.7 sind $\mathbb{R}$ und $\mathbb{R}^2$ (oder eben $\mathbb{C}$) also tatsächlich als $\mathbb{Q}$-Vektorräume und insbesondere als additive Gruppen isomorph.[59]

Um die Behauptung aus dem vorigen Absatz einzusehen, beachte man zunächst, dass $\mathbb{R}$ überabzählbar ist. Dies folgt im Wesentlichen aus Proposition 2.2.10, weil die Dezimaldarstellung reeller Zahlen eine Injektion $f : \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ liefert (wobei die i-te Nachkommastelle von $f(M)$ etwa für $i \in M$ gleich Eins und sonst gleich Null gesetzt werden kann). Weiter zeigt man, dass die Funktion $\pi : \mathbb{N}^2 \to \mathbb{N}$ mit

$$\pi(m,n) = m + \sum_{i=0}^{m+n} i = \frac{(m+n)\cdot(m+n+1)}{2}$$

bijektiv ist, sodass also $|\mathbb{N}^2| = |\mathbb{N}|$ gilt. Insbesondere ist damit $\mathbb{Q}$ abzählbar, da die Darstellung von rationalen Zahlen als Brüche eine Injektion von $\mathbb{Q}$ nach $\mathbb{N}^2$ liefert. Allgemeiner schließt man für unendliches M auf $|M^2| = |M|$, wobei das Zornsche Lemma zum Tragen kommen.[60] Induktiv ergibt sich damit $|M^n| = |M|$ für jedes unendliche M und beliebiges $n \in \mathbb{N}\backslash\{0\}$. Schreibt man $M^{<\mathbb{N}} = \bigcup\{M^n : n \in \mathbb{N}\}$ für die Menge aller Tupel aus M, so erhält man $|M^{<\mathbb{N}}| = |M|$ für unendliches M (nämlich via $|\bigcup\{M^n : n \in \mathbb{N}\}| \leq |\mathbb{N} \times M| = |M|$). Sei nun B eine Basis von $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum. Da dann jedes Element von $\mathbb{R}$ als Linearkombination $\sum_{i=1}^{n} \lambda_i b_i$ mit $\lambda_i \in \mathbb{Q}$ und $b_i \in B$ dargestellt werden kann, erhält man eine Surjektion von $(\mathbb{Q} \times B)^{<\mathbb{N}}$ nach $\mathbb{R}$, also eine Injektion in die andere Richtung. Dies ergibt

$$|\mathbb{R}| \leq |(\mathbb{Q} \times B)^{<\mathbb{N}}| = |\mathbb{N} \times B| = |B| \leq |\mathbb{R}|.$$

Hierbei wurde schon verwendet, dass B unendlich sein muss, was durch eine ähnliche Rechnung aus $|\mathbb{R}| > |\mathbb{Q}|$ folgt. In der Tat ist dann also $|B| = |\mathbb{R}| = |\mathbb{R}^2|$ die Dimension der $\mathbb{Q}$-Vektorräume $\mathbb{R}$ und $\mathbb{R}^2$.

2.4. **Untervektorräume, Dimensionsformel und Rang linearer Abbildungen.** In diesem Teilkapitel beziehen wir den Begriff der Dimension auf Untervektorräume und lineare Abbildungen.

Definition 2.4.1. Sei V ein K-Vektorraum. Eine nicht-leere Menge $U \subseteq V$ heißt Untervektorraum, wenn für $v, w \in U$ und $\lambda \in K$ stets $v + w \in U$ und $\lambda \cdot v \in U$ gilt.

Ist die Bedingung aus der Definition erfüllt, so lassen sich die Addition und Skalarmultiplikation von V zu Operationen $+ : U^2 \to U$ und $\cdot : K \times U \to U$ einschränken. Man überzeuge sich, dass U damit selbst zum Vektorraum wird (wobei Lemma 1.2.12 additiv Inverse in U liefert). Bevor wir ein Beispiel geben, überzeugen wir uns von den folgenden Zusammenhängen.

Lemma 2.4.2. *Sei $U \subseteq V$ ein Untervektorraum.*

(a) Es gilt $\dim(U) \leq \dim(V)$.
(b) Ist $\dim(U) = \dim(V)$ *endlich, so hat man* $U = V$.

[58] Die im nächsten Absatz gegebene Begründung für diesen Sachverhalt schließt der Autor vom Stoff für seine Prüfungen zur Linearen Algebra I und II aus.

[59] Es ist allerdings nicht möglich, einen Isomorphismus konkret anzugeben. Die Isomorphie kann nämlich nicht ohne eine Form des Auswahlaxioms nachgewiesen werden, wie gezeigt von Cristopher J. Ash, A consequence of the axiom of choice, Journal of the Australian Mathematical Society, Series A, 19 (1975) 306-308. In einem späteren Teilkapitel werden wir das Auswahlaxiom detaillierter betrachten.

[60] Detaillierte Lösungshinweise findet man bei Aufgabe 6.1.17 in Anton Freund, An Introduction to Mathematical Logic, 2023, `https://arxiv.org/abs/2310.09921`.

Beweis. (a) Wähle mit Korollar 2.3.1 eine Basis B von U. Ergänze diese gemäß dem Austauschsatz von Steinitz zu einer Basis $B' \supseteq B$ von V. Die Injektion $\iota : B \to B'$ mit $\iota(v) = v$ bezeugt $\dim(U) = |B| \leq |B'| = \dim(V)$.

(b) Man gehe zunächst wie in (a) vor. Ist $\dim(U) = |B| = |B'| = \dim(V)$ endlich, so folgt aus $B \subseteq B'$ schon $B = B'$ und also $U = \mathrm{Span}(B) = \mathrm{Span}(B') = V$.[61] □

Teil (b) des Lemmas ist für unendliche Dimension falsch, wie etwa die $\mathbb{Q}$-Vektorräume $\mathbb{R} \subsetneq \mathbb{R}^2$ aus Bemerkung 2.3.12 bezeugen (wobei sich auch einfachere Gegenbeispiele finden lassen).

Beispiel 2.4.3. Im euklidischen Raum $\mathbb{R}^3$ hat man zunächst die ‚trivialen' Untervektorräume $\{0\}$ und $\mathbb{R}^3$ mit Dimension Null bzw. Drei. Alle anderen Untervektorräume haben Dimension Eins oder Zwei, werden also von einem Vektor oder von zwei linear unabhängigen Vektoren aufgespannt. Es handelt sich hierbei also genau um die Geraden und Ebenen durch den Ursprung.

Untervektorräume lassen sich auch mittels des Spanns charakterisieren.

Lemma 2.4.4. *Sei V ein Vektorraum.*

(i) Ist $U \subseteq V$ ein Untervektorraum, so gilt $\mathrm{Span}(U) = U$.
(ii) Jede Menge $\mathrm{Span}(M) \subseteq V$ *mit $M \subseteq V$ ist ein Untervektorraum, nämlich der kleinste Untervektorraum $U \subseteq V$ mit $M \subseteq U$.*

Beweis. Wir überlassen (a) und den ersten Teil von (b) als Übung. Die Minimalität in (b) ergibt sich, weil man für einen Untervektorraum $U \supseteq M$ auf

$$\mathrm{Span}(M) \subseteq \mathrm{Span}(U) = U$$

schließen kann. □

Diese Charakterisierung ist etwa für den nächsten Beweis nützlich.

Lemma 2.4.5. *Ist $\mathcal{U}$ eine Menge von Untervektorräumen eines Vektorraums V, so ist auch $\bigcap \mathcal{U}$ ein Untervektorraum von V.*

Beweis. Für jedes $U \in \mathcal{U}$ gilt $\mathrm{Span}(\bigcap \mathcal{U}) \subseteq \mathrm{Span}(U) = U$. Man erhält dann also $\mathrm{Span}(\bigcap \mathcal{U}) = \bigcap \mathcal{U}$, und das Resultat folgt mit dem vorigen Lemma. □

Die Vereinigung von Untervektorräumen ist im Allgemeinen kein Untervektorraum. Man hat sogar das folgende Ergebnis.

Lemma 2.4.6. *Für Untervektorräume U_0 und U_1 eines Vektorraums ist $U_0 \cup U_1$ genau dann ein Untervektorraum, wenn $U_0 \subseteq U_1$ oder $U_1 \subseteq U_0$ gilt.*

Beweis. Hat man etwa $U_0 \subseteq U_1$, so ist $U_0 \cup U_1 = U_1$ nach Annahme ein Untervektorraum. Die andere Richtung beweisen wir per Kontraposition: Gilt $U_0 \not\subseteq U_1$ und $U_1 \not\subseteq U_0$, so wählen wir $v \in U_0 \backslash U_1$ und $w \in U_1 \backslash U_0$. Es kann nicht $v + w \in U_0$ gelten, da man sonst mit $-v \in U_0$ auch $w = -v + v + w \in U_0$ hätte. Ebenso wenig kann $v + w \in U_1$ gelten. Also ist $U_0 \cup U_1$ nicht unter Summen abgeschlossen und somit kein Untervektorraum. □

[61] Streng genommen wird hier $\mathrm{Span}(B)$ und U aber $\mathrm{Span}(B')$ in V gebildet, was jedoch der Gleichheit keinen Abbruch tut. Wir merken an, dass eine Menge M genau dann endlich ist, wenn sie zu keiner echten Teilmenge in Bijektion steht. Ist nämlich M unendlich, mag man $M = \mathbb{N} \cup M'$ als disjunkte Vereinigung schreiben (vgl. den Absatz vor Definition 2.2.8). Nun steht aber $\mathbb{N}$ vermöge $n \mapsto 2n$ in Bijektion mit den geraden Zahlen $\{0, 2, 4, \ldots\} \subsetneq \mathbb{N}$. Dies liefert also eine äquivalente Definition von Endlichkeit, die insbesondere mit Richard Dedekind assoziiert wird.

Wegen Lemma 2.4.4 erscheint das Folgende als bester Ersatz für die Vereinigung.

Definition 2.4.7. Sei $\mathcal{U}$ eine Menge von Untervektorräumen eines Vektorraums V. Wir setzen

$$\bigoplus\nolimits^V \mathcal{U} := \operatorname{Span}\left(\bigcup \mathcal{U}\right).$$

Im Fall von zwei Untervektorräumen schreiben wir auch $U \oplus^V U'$ für $\bigoplus^V \{U, U'\}$. Weiter schreiben wir $U \oplus U'$ statt $U \oplus^V U'$, wenn dabei $U \cap U' = \{0\}$ gilt.[62]

Als Übung sei überlassen:

Lemma 2.4.8. *Es gilt* $U \oplus^V U' = \{u + u' : u \in U \text{ und } u' \in U'\}$.

Im folgenden wichtigen Satz wird unsere Konstruktion auf Untervektorräumen zum Begriff der Dimension in Bezug gesetzt.

Proposition 2.4.9 (Dimensionsformel)**.** *Für Untervektorräume U, W eines Vektorraums V gilt*[63]

$$\dim(U) + \dim(W) = \dim(U \oplus^V W) + \dim(U \cap W).$$

Beweis. Wähle eine Basis B von $U \cap W$ und ergänze diese dann zu Basen $B \cup C_U$ und $B \cup C_W$ von U bzw. W, wobei die Mengen B, C_U, C_W paarweise disjunkt sind. Wegen $U \cup W \subseteq \operatorname{Span}(B \cup C_U \cup C_W)$ ist die Menge $B \cup C_U \cup C_W$ ein Erzeugendensystem von $U \oplus^V W = \operatorname{Span}(U \cup W)$.

Wir zeigen, dass $B \cup C_U \cup C_W$ auch linear unabhängig ist. Betrachte hierfür

$$u + v + w = 0 \quad \text{mit} \quad \begin{cases} u = \sum_{i=1}^{l} \lambda_i u_i & \text{für paarweise verschiedene } u_i \in C_U, \\ v = \sum_{i=1}^{m} \mu_i v_i & \text{für paarweise verschiedene } v_i \in B, \\ w = \sum_{i=1}^{n} \nu_i w_i & \text{für paarweise verschiedene } w_i \in C_W. \end{cases}$$

Es ist nun $w = -u - v \in U \cap W = \operatorname{Span}(B)$. Da die Darstellung von w bezüglich der Basis $B \cup C_W$ eindeutig ist (siehe Lemma 2.1.8 und die auf dieses folgende Bemerkung), muss $w = 0$ und somit $\nu = 1 = \ldots = \nu_n = 0$ gelten. Analog erhält man $u = 0$ und damit auch $v = 0$. Also sind alle Koeffizienten gleich Null.

Wie haben gesehen, dass $B \cup C_U \cup C_W$ eine Basis von $U \oplus^V W$ ist. Es gilt also

$$\dim(U) = |B| + |C_U|, \qquad \dim(U \cap W) = |B|,$$
$$\dim(W) = |B| + |C_W|, \qquad \dim(U \oplus^V W) = |B| + |C_U| + |C_W|.$$

Hieraus ergibt sich die Behauptung (auch im unendlichdimensionalen Fall durch Angabe einer geeigneten Bijektion). □

Beispiel 2.4.10. Seien $U, W \subseteq \mathbb{R}^3$ zwei Untervektorräume der Dimension zwei (also Ebenen durch den Ursprung im Raum). Gilt $U \neq W$ und also $U \subsetneq U \oplus^{\mathbb{R}^3} W$, so muss wegen Lemma 2.4.2 jedenfalls

$$2 = \dim(U) < \dim(U \oplus^{\mathbb{R}^3} W) \leq \dim(\mathbb{R}^3) = 3$$

[62] Diese Notation ist sinnvoll, weil aus $U \cong W$ und $U' \cong W'$ mit $U \cap U' = \{0\} = W \cap W'$ schon $U \oplus U' \cong W \otimes W'$ folgt, wie man prüfen möge. Bei trivialem Schnitt (und nur dann) hängt die Konstruktion also nicht wesentlich vom umgebenden Raum V ab. Tatsächlich werden wir später eine allgemeinere Operation $\oplus$ definieren, die sich auf beliebige Vektorräume anwenden lässt (also nicht nur auf Untervektorräume).

[63] Im unendlichen Fall lese man $|M| + |N| = |M \sqcup N|$, wobei $M \sqcup N := \{(0, x) : x \in M\} \cup \{(1, y) : y \in N\}$ die disjunkte Summe bezeichnet.

und also dann $\mathrm{Span}(U \cup W) = U \oplus^{\mathbb{R}^3} W = \mathbb{R}^3$ gelten. Weiter liefert die Dimensionsformel nun $\dim(U \cap W) = 1$. Mit anderen Worten: Aus unseren allgemeinen Überlegungen zur Dimension folgt, dass zwei verschiedene Ebenen durch den Ursprung stets den Raum aufspannen und sich in einer Geraden schneiden (was wir uns in diesem Fall freilich auch direkter hätten klarmachen können).

Eine weitere Charakterisierung von Untervektorräumen werden wir über den folgenden Begriff erhalten.

Definition 2.4.11. Der Kern einer linearen Abbildung $\varphi : V \to W$ zwischen Vektorräumen ist definiert als

$$\ker(\varphi) = \varphi^{-1}[\{0\}] = \{v \in V : \varphi(v) = 0\}.$$

Die zuvor angesprochene Charakterisierung lautet wie folgt.

Lemma 2.4.12. *Sei V ein Vektorraum. Für eine Menge $M \subseteq V$ sind äquivalent:*

(i) Die Menge M ist ein Untervektorraum von V.
(ii) Man hat $M = \ker(\varphi)$ für eine lineare Abbildung $\varphi : V \to W$.
(iii) Man hat $M = \mathrm{img}(\psi)$ für eine lineare Abbildung $\psi : U \to V$.

Beweis. Um zu zeigen, dass (i) aus (ii) wie auch aus (iii) folgt, müssen wir prüfen, dass $\ker(\varphi)$ und $\mathrm{img}(\psi)$ stets Untervektorräume sind. Wegen $\varphi(0) = 0$ ist jedenfalls $\ker(\varphi)$ und ebenso $\mathrm{img}(\psi)$ nicht leer. Für $v, w \in \ker(\varphi)$ gilt

$$\varphi(v + w) = \varphi(v) + \varphi(w) = 0 + 0 = 0$$

und somit $v + w \in \ker(\varphi)$. Ist weiter λ ein Skalar, so erhält man auch

$$\varphi(\lambda \cdot v) = \lambda \cdot \varphi(v) = \lambda \cdot 0 = 0$$

und also $\lambda \cdot v \in \ker(\varphi)$. Hat man andererseits $v, w \in \mathrm{img}(\psi)$ gegeben, so schreibt man $v = \psi(v')$ und $w = \psi(w')$, um dann $v + w = \psi(v' + w') \in \mathrm{img}(\psi)$ sowie auch $\lambda \cdot v = \psi(\lambda \cdot v') \in \mathrm{img}(\psi)$ für jeden Skalar λ zu erhalten.

Um zu zeigen, dass jeder Untervektorraum $M \subseteq V$ als $\mathrm{img}(\psi)$ für ein $\psi : U \to V$ geschrieben werden kann, setzt man schlicht $U := M$ und $\psi(v) := v$. Wir suchen nun noch ein $\varphi : V \to W$ mit $M = \ker(\varphi)$. Dazu setzen wir zunächst $W = V$. Sodann wählen wir eine Basis B von M und ergänzen diese zu einer Basis C von V. Mittels Definition 2.1.11 setzen wir $\varphi := \mathrm{Abb}(f)$ für

$$f : C \to V \quad \text{mit} \quad f(v) = \begin{cases} 0 & \text{wenn } v \in B, \\ v & \text{wenn } v \in C \backslash B. \end{cases}$$

Man überzeuge sich zunächst von $\varphi(w) = 0$ für $w \in M$. Wir betrachten nun einen beliebigen Vektor $w \in V \backslash M$ und schreiben diesen als

$$w = w_0 + w_1 \quad \text{mit } w_0 \in M \text{ und } 0 \neq w_1 = \sum_{i=1}^{n} \lambda_i v_i \text{ mit } v_i \in C \backslash B.$$

Hier ergibt sich $\varphi(w) = w_1 \neq 0$. Man hat also $\ker(\varphi) = M$, wie gewünscht. □

Den Kern von $\varphi : V \to W$ kann man als Urbild des Untervektorraums $\{0\} \subseteq W$ auffassen. Allgemeiner und mit einem ähnlichen Beweis erhält man:

Korollar 2.4.13. *Sei $\varphi : V \to W$ eine lineare Abbildung.*

(a) Ist U ein Untervektorraum von W, so ist das Urbild $\varphi^{-1}[U]$ ein Untervektorraum von V.

(b) *Ist U' ein Untervektorraum von V und B eine Basis von U', so ist das Bild der Untervektorraum*

$$\varphi[U'] = \mathrm{Span}(\{\varphi(v) : v \in B\}).$$

Beweis. (a) Wir haben im vorigen Lemma gezeigt, dass $\ker(\varphi) = \varphi^{-1}[\{0\}]$ ein Untervektorraum ist. Für einen allgemeinen Untervektorraum U argumentiert man ähnlich wie zuvor für $U = \{0\}$, wovon man sich als Übung überzeugen möge.

(b) Man schränke φ zu einer Abbildung $\varphi' : U' \to W$ mit $\varphi'(v) = \varphi(v)$ ein. Aus dem Lemma folgt dann, dass $\varphi[U'] = \mathrm{img}(\varphi')$ ein Untervektorraum ist. Die Beschreibung durch den Spann überlassen wir ebenfalls zur Übung. □

Punkt (iii) aus Lemma 2.4.12 rechtfertigt den folgenden Begriff.

Definition 2.4.14. Der Rang einer linearen Abbildung $\varphi : V \to W$ ist die Dimension ihres Bildes, also

$$\mathrm{rank}(\varphi) = \dim(\mathrm{img}(\varphi)).$$

Für eine Matrix $M \in K^{m\times n}$ schreibt man auch $\mathrm{rank}(M)$ statt $\mathrm{rank}(\mathrm{Abb}(M))$ und spricht vom Rang von M (beachte $\mathrm{Abb}(M) : K^n \to K^m$).

Um den Rang einer Matrix zu bestimmen, ist die folgende Beobachtung hilfreich:

Bemerkung 2.4.15. Wir betrachten eine Matrix $M \in K^{m\times n}$ und die induzierte Abbildung $\mathrm{Abb}(M) : K^n \to K^m$. Das vorige Korollar liefert

$$\mathrm{rank}(M) = \dim\big(\mathrm{Span}(E)\big) \quad \text{mit} \quad E = \{M \cdot e_i : 1 \le i \le n\},$$

wobei e_i der i-te Einheitsvektor ist. Betrachte eine maximal linear unabhängige Teilmenge L von E. Nach dem Austauschsatz von Steinitz können wir L zu einer Basis $B \subseteq E$ von $\mathrm{Span}(E)$ ergänzen. Wegen der Maximalität muss aber dann bereits $L = B$ und also $\mathrm{rank}(M) = |L|$ gelten. Nun ist $M \cdot e_i$ die i-te Spalte von M. Der Rang einer Matrix ist also die Kardinalität einer maximal linear unabhängigen Menge von Spaltenvektoren – oder von einem minimalen Erzeugendensystem.[64]

Das folgende Ergebnis war bereits im Beweis von Proposition 2.1.13 implizit (wie man sich klarmachen möge).

Korollar 2.4.16. *Eine lineare Abbildung $\varphi : V \to W$ ist genau dann injektiv, wenn man $\ker(\varphi) = \{0\}$ hat.*

Unser nächstes Ergebnis kann zu einem gewissen Grad als Verallgemeinerung des vorigen Korollars gesehen werden.

Proposition 2.4.17 (Rangsatz)**.** *Für jede lineare Abbildung $\varphi : V \to W$ gilt*

$$\dim(V) = \dim(\ker(\varphi)) + \mathrm{rank}(\varphi).$$

Beweis.[65] Wähle eine Basis B von $\ker(\varphi)$ und ergänze sie zu einer Basis C von V. Wir schränken φ zu einer Abbildung $\varphi_0 : \mathrm{Span}(C\backslash B) \to W$ ein. Diese hat dasselbe

[64] Man spricht auch vom Spaltenrang. Dieser ist tatsächlich gleich dem Zeilenrang, also der Kardinalität einer maximal linear unabhängigen Menge von Zeilenvektoren. Wir werden dies erst deutlich später im Zusammenhang mit dem sogenannten Dualraum beweisen.

[65] Aus Sicht des Autors nutzt der ‚kanonische' Beweis die Darstellung $\mathrm{img}(\varphi) \cong V/\ker(\varphi)$ des Bildes als Faktorraum. Diese werden wir allerdings erst später diskutieren. Wir geben hier also einen alternativen Beweis.

Bild wie φ, denn jedes $v \in V$ lässt sich als $v = u + w$ mit $u \in \mathrm{Span}(B) = \ker(\varphi)$ und $w \in \mathrm{Span}(C \backslash B)$ schreiben, sodass man

$$\varphi(v) = \varphi(u) + \varphi(w) = 0 + \varphi(w) = \varphi_0(w)$$

erhält. Weiter gilt $\ker(\varphi_0) = \{0\}$, sodass φ injektiv ist: Hat man nämlich $\varphi_0(v) = 0$ mit $v \in \mathrm{Span}(C \backslash B)$, so gilt auch $\varphi(v) = 0$ und also $v \in \mathrm{Span}(B)$. Wegen der Eindeutigkeit der Darstellung bezüglich der Basis C kann dann aber nur $v = 0$ sein. Insgesamt bezeugt φ_0 also

$$\mathrm{Span}(C \backslash B) \cong \mathrm{img}(\varphi).$$

Es folgt $\dim(V) = |C| = |B| + |C \backslash B| = \dim(\ker(\varphi)) + \mathrm{rank}(\varphi)$. □

Im endlichdimensionalen Fall liefert Lemma 2.4.2(b) direkt das Folgende.

Korollar 2.4.18. *Ist* $\dim(W)$ *endlich, so ist eine lineare Abbildung* $\varphi : V \to W$ *genau dann surjektiv, wenn* $\mathrm{rank}(\varphi) = \dim(W)$ *gilt.*

Mit der folgenden Proposition können wir leichter testen, ob eine Abbildung ein Isomorphismus ist.[66]

Proposition 2.4.19. *Wenn* $\dim(V) = \dim(W)$ *endlich ist, so ist eine lineare Abbildung* $\varphi : V \to W$ *genau dann injektiv, wenn sie surjektiv ist.*

Beweis. Mit der Dimensionsformel erhält man

$$\ker(f) = \{0\} \quad \Leftrightarrow \quad \dim(\ker(\varphi)) = 0 \quad \Leftrightarrow \quad \mathrm{rank}(\varphi) = \dim(W).$$

Das Ergebnis folgt mit dem vorigen Korollar und Korollar 2.4.16. □

Es ist instruktiv, für das vorige Ergebnis noch einen direkteren Beweis zu erarbeiten, bei dem man $\varphi = \mathrm{Abb}(M)$ schreibt und Bemerkung 2.1.14 verwendet. Umgekehr liefern diese Bemerkung und unser jetziges Korollar zusammen das Folgende.

Bemerkung 2.4.20. Für $M \in K^{n \times n}$ ist $\mathrm{Abb}(M) : K^n \to K^n$ genau dann ein Isomorphismus, wenn eine der folgenden äquivalenten Eigenschaften erfüllt ist:

(i) Die Spalten von M sind paarweise verschieden und linear unabhängig.
(ii) Die Spalten von M bilden ein Erzeugendensystem.
(iii) Die Spalten von M bilden eine Basis.

2.5. **Exkurs: Das Auswahlaxiom.** In diesem Teilkapitel diskutieren wir, inwieweit das Auswahlaxiom essentiell ist, um zu beweisen, dass jeder Vektorraum eine Basis hat. Wir begeben uns dabei auf einen kleinen Ausflug in die mathematische Logik. Letztere ist ein vielfältiges Gebiet mit Anwendungen von Algebra bis Optimierung und Bezügen zur Informatik und Philosophie.[67]

Eine wichtige Grundfrage der Logik lautet: Welche Axiome oder Grundannahmen sind zwingend erforderlich, um ein gegebenes mathematisches Ergebnis zu beweisen? Um diese Frage mathematisch präzise zu machen, modelliert man Beweise selbst als mathematische Objekte (‚Metamathematik'): Ein Beweis ist eine endliche Folge von Zeichen, die nach klar definierten Regeln gebaut ist. Es mag

[66] Das Ergebnis gilt tatsächlich nur im endlichdimensionalen Fall (vgl. Fußnote 61).

[67] Das Skript zu einer Einführungsvorlesung des Autors ist in Fußnote 60 verlinkt. Der Autor schließt das ganze Teilkapitel 2.5 des vorliegenden Skripts vom Stoff für seine Prüfungen zur Linearen Algebra I und II aus.

hilfreich sein, hier an eine Programmiersprache zu denken, die flexibel eingesetzt werden kann, dabei aber mit maschineller Genauigkeit ausgewertet wird.[68]

Es ist weithin akzeptiert, dass die heutige Mathematik größtenteils auf Basis der Zermelo-Fraenkelschen Mengenlehre mit dem Auswahlaxiom (ZFC) entwickelt werden kann.[69] Dies gilt auch für das Material in diesem Skript. Die zuvor informell gestellte Frage, wann das Auswahlaxiom eine essentielle Rolle spielt, ist nun auf folgende Weise mathematisch präzise zu fassen: Welche Ergebnisse können auch in der Zermelo-Fraenkelschen Mengenlehre ohne Auswahlaxiom (ZF) bewiesen werden? Im Folgenden werden wir dies exemplarisch untersuchen, wobei unsere Argumente insofern lückenhaft bleiben, als wir den logischen Formalismus und die Definition von ZF nicht weiter klären werden. Hingegen ist es sicher an der Zeit, unser Untersuchungsobjekt explizit zu machen:

Definition 2.5.1. Das Auswahlaxiom ist das folgende Prinzip: Ist A eine Menge von nicht-leeren Mengen, so gibt es eine Funktion $f : A \to \bigcup A$ mit der Eigenschaft, dass $f(X) \in X$ für alle $X \in A$ gilt.

Manchen wird das Auswahlaxiom selbstverständlich erscheinen, da man in nichtleeren Mengen sicher Elemente findet. Andere mögen einwenden, dass man den Akt des Wählens nicht einfach zu unendlich vielen parallelen ‚Entscheidungen' erweitern könne. Jedenfalls lässt sich konstatieren, dass Beweise ihren konstruktiven Charakter verlieren, wenn die Auswahlfunktion $f : A \to \bigcup A$ nicht näher anzugeben ist. Freilich sind auch viele andere Beweise – etwa solche durch Widerspruch – nicht konstruktiv, indem sie die Existenz eines Objektes nachweisen, ohne zumindest implizit seine Konstruktion möglich zu machen. In der Praxis scheint es heute (anders als um 1900) weitgehend Konsens zu sein, dass das Auswahlaxiom bedenkenlos verwendet werden kann.

In Bezug auf die mathematische Logik sei angemerkt, dass das Auswahlaxiom nur ein Forschungsthema unter vielen ist. Insbesondere hat die Forschung ergeben, dass bereits ZF für signifikante Teile der Mathematik viel zu stark ist, sodass die Suche nach den ‚angemessenen' Axiomen für ein mathematisches Gebiet mit gleicher Berechtigung an ganz anderer Stelle ansetzen kann.

Als Voraussetzung für den eher technischen Teil unserer Diskussion benötigen wir das folgende Resultat.

Lemma 2.5.2. *Für jede Menge M gibt es einen Körper K mit $|M| \leq |K\backslash\{0\}|$.*

Beweisskizze. Wir setzen an dieser Stelle die Konstruktion von $\mathbb{Q}$ aus $\mathbb{Z}$ als bekannt voraus. Analog lässt sich jeder Integritätsring in einen Körper einbetten. Es reicht also aus, wenn man Integritätsringe mit beliebig großer Kardinalität konstruiert. In erster Näherung betrachten wir hierfür Polynome in einer Variablen X, also Ausdrücke der Form $a_{n-1} \cdot X^{n-1} + \ldots + a_0 \cdot X^0$ mit Koeffizienten a_i beispielsweise aus $\mathbb{Z}$. Diese können wie erwartet addiert und multipliziert werden,[70] wodurch

[68] Tatsächlich gibt es ‚interactive theorem prover', mit deren Hilfe man Beweise am Computer formalisieren und auf ihre Richtigkeit prüfen lassen kann. Selbst aufwändige Beweise wie der des Vier-Farben-Satzes wurden inzwischen so verifiziert (in diesem Fall 2005 durch eine Gruppe um Georges Gonthier mit dem Programm Rocq).

[69] Diese ist benannt nach Ernst Zermelo und Abraham Fraenkel, wobei unter anderem auch Thoralf Skolem einen wichtigen Beitrag geleistet hat.

[70] Es gilt etwa $(3X^2 + X) + (X + 2) = 3X^2 + 2X + 2$ und $(3X^2 + X) \cdot (X + 2) = 3X^3 + 7X^2 + 2X$. Wir werden Polynome später noch genauer studieren.

sie zu einem mit $\mathbb{Z}[X]$ bezeichneten Integritätsring werden. Genauso erhält man einen Integritätsring, wenn man Polynome in mehreren Variablen betrachtet (so wie $4X^2Y + 3Y^2$). Verwendet man die $|M|$-vielen Variablen X_i für $i \in M$, so ergibt sich das Ergebnis.[71] □

Wir können nun das Hauptresultat dieses Teilkapitels ableiten.

Theorem 2.5.3 (Halpern[72])**.** *Über* ZF *sind äquivalent:*

(i) Das Auswahlaxiom ist gültig.

(ii) Für ein Erzeugendensystem E eines Vektorraums V gibt es stets eine Basis $B \subseteq E$ von V.

Beweis. Die Richtung von (i) nach (ii) ist der Austauschsatz von Steinitz (für den wir das Zornsche Lemma und darüber das Auswahlaxiom verwendet haben). Wir nehmen nun (ii) an und leiten (i) her. Sei dazu A eine Menge von nicht-leeren Mengen. Wir können davon ausgehen, dass die Elemente von A paarweise disjunkt sind (indem wir $X \in A$ jeweils durch $\{(X, y) : y \in X\}$ ersetzen). Wegen des vorigen Lemmas (welches über ZF gültig ist) finden wir einen Körper K mit einer Injektion $\iota : \bigcup A \to K\backslash\{0\}$. Gemäß der Konstruktion in Definition 2.3.10 setzen wir $V = \bigoplus_A K$. Für $y \in \bigcup A$ sei $\eta_y : A \to K$ gegeben durch

$$\eta_y(X) = \begin{cases} \iota(y) & \text{wenn } y \in X, \\ 0 & \text{sonst.} \end{cases}$$

Da die Elemente von A disjunkt sind, gibt es je genau ein X mit $y \in X$, für das man dann $\operatorname{supp}(\eta_y) = \{X\}$ hat. Insbesondere ist also $\eta_y \in V$. Nun ist

$$E := \left\{\eta_x : x \in \bigcup A\right\}$$

ein Erzeugendensystem. Ist nämlich $f : A \to K$ mit $\operatorname{supp}(f) = \{X_1, \ldots, X_n\}$ beliebig, so findet man per Induktion (und also ohne das Auswahlaxiom) eine Folge von Elementen $y(i) \in X_i$. Damit erhält man

$$f = \sum_{i=1}^{n} \frac{f(X_i)}{\iota(y(i))} \cdot \eta_{y(i)}.$$

Mit (ii) erhält man nun eine Basis

$$B = \{\eta_y : y \in Z\} \subseteq E,$$

wobei also $Z \subseteq \bigcup A$ gilt. Damit B ein Erzeugendensystem ist, muss es für $X \in A$ mindestens ein $y \in X$ mit $y \in Z$ geben. Wegen der linearen Unabhängigkeit kann es aber auch nur ein solches y geben. Hat man nämlich $y, z \in X \cap Z$, so ist

$$\frac{1}{\iota(y)} \cdot \eta_y - \frac{1}{\iota(z)} \cdot \eta_z = 0,$$

sodass $\eta_y = \eta_z$ und mit der Injektivität von ι dann $y = z$ gelten muss. Die gewünschte Auswahlfunktion $f : A \to \bigcup A$ kann also explizit (und damit ohne

[71] Es mag helfen, sich an die zumindest intuitiv verwandte Konstruktion der Vektorräume $\bigoplus_M K$ aus Definition 2.3.10 zu erinnern.

[72] Das Ergebnis stammt von James Halpern, Bases in vector spaces and the axiom of choice, Proceedings of the American Mathematical Society 17 (1966) 670-673. Der hier angegebene Beweis findet sich ebenfalls in diesem Artikel, wird dort aber Hans Läuchli zugeschrieben.

Verwendung des Auswahlaxioms) durch die Bedingung $X \cap Z = \{f(X)\}$ charakterisiert werden. □

Aussage (ii) aus dem Theorem wird schwächer, wenn wenn wir nur die Existenz einer Basis behaupten (ohne Referenz auf ein gegebenes Erzeugendensystem). Auch diese schächere Form von (ii) impliziert aber noch das Auswahlaxiom, wie Andreas Blass gezeigt hat.[73]

Im obigen Theorem ist die Äquivalenz mit dem Auswahlaxiom nur deshalb informativ, weil dieses nicht in ZF enthalten ist. Stünde uns nämlich das Auswahlaxiom uneingeschränkt zur Verfügung, so könnten wir schlicht argumentieren, dass (i) und (ii) beide wahr und damit wie alle wahren Aussagen äquivalent sind. Im Fall des Auswahlaxioms gilt tatsächlich:

Theorem 2.5.4. *Das Auswahlaxiom ist von* ZF *unabhängig. Man kann es dort also weder widerlegen (Gödel) noch beweisen (Cohen).*[74]

Die Beweise müssen einer fortgeschrittenen Vorlesung über Mengenlehre vorbehalten bleiben. Sowohl Kurt Gödel als auch Paul Cohen (dem dafür die Fields-Medaille zugesprochen wurde) haben für sie umfangreiche neue Methoden entwickelt, welche die Mengenlehre wie kaum ein anderer Beitrag geprägt haben. Zum Abschluss dieses Kurses ziehen wir die folgende Schlussfolgerung.

Korollar 2.5.5. *In ZF kann nicht bewiesen werden, dass es für jedes Erzeugendensystem E eines Vektorraums V eine Basis $B \subseteq E$ von V gibt.*

Beweis. Wäre die Behauptung falsch, so könnte man wegen Theorem 2.5.3 auch das Auswahlaxiom in ZF beweisen, was aber dem oben zitierten Ergebnis von Cohen widerspricht. □

[73] Andreas Blass, Existence of bases implies the axiom of choice, in: James Baumgartner, Donald Martin und Saharon Shelah (Hg.), Axiomatic Set Theory, Contemporary Mathematics 31, American Mathematical Society, Providence (RI), 1984.

[74] Die ursprünglichen Beweise sind zu finden in Kurt Gödel, The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory, Annals of Mathematics Studies 3, Princeton University Press, 1940, sowie in Paul Cohen, The Independence of the Continuum Hypothesis I/II, Proceedings of the U.S. National Academy of Sciences 50/51 (1964/64) 1143-48/105-110.

3. Invertierbarkeit

In diesem Kapitel beschäftigen wir uns damit, wie man Lösungen von linearen Gleichungssystemen und inverse Matrizen explizit bestimmen kann. Weiter diskutieren wir die sogenannte Determinante, welche unter anderem als Test auf Invertierbarkeit eingesetzt werden kann, aber auch darüber hinaus eine reiche Theorie und vielfältige Anwendungen bietet.

3.1. **Lineare Gleichungssysteme.** Ein lineares Gleichungssystem wie

$$\begin{aligned} x + 2y - z &= -1, \\ x + 2y + z &= 3, \\ -2x + y + z &= 0, \end{aligned}$$

lässt sich auch schreiben als Matrixgleichung

$$M \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} \quad \text{mit} \quad M = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 1 \\ -2 & 1 & 1 \end{pmatrix}.$$

In Anbetracht von $M \cdot v = \mathrm{Abb}(M)(v)$ beschäftigen wir uns noch allgemeiner mit Gleichungen $\varphi(v) = w$ für lineare Abbildungen φ. Für deren Lösungsmengen führen wir die folgende Notation ein.

Definition 3.1.1. Sei $\varphi : V \to W$ eine lineare Abbildung zwischen Vektorräumen. Für $w \in W$ schreiben wir

$$\mathrm{Lsg}(\varphi, w) = \varphi^{-1}[\{w\}] = \{v \in V : \varphi(v) = w\}.$$

Ist $M \in K^{m\times n}$ eine Matrix und $w \in K^m$, so schreiben wir auch $\mathrm{Lsg}(M, w)$ anstelle von $\mathrm{Lsg}(\mathrm{Abb}(M), w)$ mit $\mathrm{Abb}(M) : K^n \to K^m$.

Explizit hat man also $\mathrm{Lsg}(M, w) = \{v \in K^n : M \cdot v = w\}$. Wir werden sehen, dass Lösungsmengen die folgende Struktur haben.

Definition 3.1.2. Für einen Vektorraum V heißt $A \subseteq V$ ein affiner Teilraum, wenn es ein $v \in V$ und einen Untervektorraum $U \subseteq V$ gibt mit

$$A = v + U = \{v + w : w \in U\}.$$

Ist $U \subseteq \mathbb{R}^2$ eine Gerade durch den Ursprung, so ist $v + U$ die parallele Gerade mit Aufpunkt v. Hieraus sollte das folgende Ergebnis plausibel sein.

Lemma 3.1.3. *Für jeden Vektorraum V und beliebige $x, x' \in V$ sowie Untervektorräume $U, U' \subseteq V$ gilt*

$$\begin{aligned} x + U = x' + U \quad &\Leftrightarrow \quad x' \in x + U, \\ x + U = x + U' \quad &\Leftrightarrow \quad U = U'. \end{aligned}$$

Aus $x + U = x' + U'$ folgt damit $U = U'$.

Beweis. Gilt $x + U = x' + U$, so liefert $x' = x' + 0 \in x' + U$ auch $x' \in x + U$. Gilt andererseits $x' \in x + U$ und also $x' = x + u$ für ein $u \in U$, so erhält man für beliebiges $x' + \tilde{u} \in x' + U$ mit $\tilde{u} \in U$ auch $x' + \tilde{u} = x + u + \tilde{u} \in x + U$, sodass also $x' + U \subseteq x + U$ gilt. Weiter ist $x = x' - u \in x' + U$ und symmetrisch $x + U \subseteq x' + U$.

Für die zweite Äquivalenz genügt es, wenn wir zeigen, dass aus $x + U \subseteq x + U'$ schon $U \subseteq U'$ folgt. Sei dazu $u \in U$ beliebig. Dann erhalten wir $x + u \in x + U'$

und können also $x + u = x + u'$ mit $u' \in U'$ schreiben. Kürzen in der Gruppe V liefert $u = u' \in U'$, wie gewünscht.

Gilt schließlich $x + U = x' + U'$, so erhält man zunächst $x' \in x + U$ und damit

$$x' + U = x + U = x' + U',$$

woraus dann $U = U'$ folgt. □

Das obige Lemma stellt sicher, dass der folgende Begriff wohldefiniert ist.

Definition 3.1.4. Die Dimension eines affinen Teilraums $x + U$ ist gegeben durch

$$\dim(x + U) = \dim(U),$$

wobei die rechte Seite die Dimension des Untervektorraums U bezeichnet.

Das folgende Lemma liefert den versprochenen Bezug zu linearen Gleichungen. Es besagt insbesondere, dass die Lösungen der ‚inhomogenen' Gleichung $\varphi(v) = w$ sich als Summen $v + v'$ einer festen Lösung v mit den Lösungen v' der ‚homogenen' Gleichung $\varphi(v') = 0$ ergeben.

Lemma 3.1.5. *Für jede lineare Abbildung $\varphi : V \to W$ gilt*

$$\varphi(v) = w \quad \Rightarrow \quad \mathrm{Lsg}(\varphi, w) = v + \ker(\varphi) = v + \mathrm{Lsg}(\varphi, 0).$$

Insbesondere ist $\mathrm{Lsg}(\varphi, w)$ entweder leer oder ein affiner Teilraum von V.

Beweis. Gilt $\varphi(v) = w$, so hat man

$$v' \in \mathrm{Lsg}(\varphi, w) \Leftrightarrow \varphi(v') = \varphi(v) \Leftrightarrow \varphi(v' - v) = 0 \Leftrightarrow v' - v \in \ker(\varphi).$$

Hat man $v' - v \in \ker(\varphi)$, so ergibt sich $v' = v + v' - v \in v + \ker(\varphi)$. Gilt umgekehrt die Beziehung $v' = v + w \in v + \ker(\varphi)$, so folgt $v' - v = w \in \ker(\varphi)$. □

Es ergibt sich direkt:

Korollar 3.1.6. *Wir betrachten eine lineare Abbildung $\varphi : V \to W$ sowie $w \in W$. Hat man $\mathrm{Lsg}(\varphi, w) \neq \emptyset$, so gilt*

$$\dim(\mathrm{Lsg}(\varphi, w)) = \dim(\ker(\varphi)).$$

Eine Matrix $M \in K^{m \times n}$ mit $m < n$ entspricht einem linearen Gleichungssystem, in dem es weniger Gleichungen als Unbekannte gibt. Das nächste Ergebnis besagt insbesondere, dass ein solches System nicht eindeutig lösbar sein kann.

Korollar 3.1.7. *Betrachte eine Matrix $M \in K^{m \times n}$ mit $m < n$ sowie ein $w \in K^m$. Gibt es ein $v \in K^n$ mit $M \cdot v = w$, so gilt*

$$\dim(\mathrm{Lsg}(M, w)) \geq n - m.$$

Beweis. Mit dem vorigen Ergebnis und der Dimensionsformel erhält man

$$\dim(\mathrm{Lsg}(M, w)) = \dim(K^n) - \mathrm{rank}(M) \geq n - m.$$

Hierbei gilt die Ungleichung wegen $\mathrm{rank}(M) \leq m$, was sich mit Bemerkung 2.4.15 begründen lässt. □

Ganz konkret lassen sich die Lösungen eines linearen Gleichungssystems ermitteln, indem man

- je eine Zeile mit einem Skalar ungleich Null multipliziert,
- eine Zeile von einer anderen abzieht,

- Zeilen vertauscht.[75]

Man beachte, dass man durch Kombination der ersten beiden Operationen auch das λ-fache einer Zeile von einer anderen Zeile abziehen kann: Dazu multipliziert man die erste Zeile mit λ und zieht diese dann von der zweite Zeile ab, bevor man die erste Zeile wieder mit $1/\lambda$ multipliziert. In der Praxis wird man dies wohl in einem Schritt ausführen.

Im Gleichungssystem vom Beginn dieses Teilkapitels mag man etwa die dritte Zeile mit $-1/2$ multiplizieren und dann die erste Zeile von der zweiten sowie von der dritten abziehen. Dies führt auf das Gleichungssystem

$$\begin{aligned} x + \qquad 2y - \qquad z &= -1, \\ 2z &= \quad 4, \\ -5/2 \cdot y + 1/2 \cdot z &= \quad 1, \end{aligned}$$

welches (wie wir auch noch zeigen werden) dieselbe Lösungsmenge hat. Nun multiplizieren wir die dritte Zeile mit $-2/5$ und die zweite Zeile mit $1/2$. Dann vertauschen wir die zweite und dritte Zeile. Dies ergibt

$$\begin{aligned} x + 2y - \qquad z &= \quad -1, \\ y - 1/5 \cdot z &= -2/5, \\ z &= \qquad 2. \end{aligned}$$

Es lässt sich nun von unten nach oben ablesen, dass die einzige Lösung durch $z = 2$, $y = 0$ und $x = 1$ gegeben ist. Wir werden sehen, dass sich die angegebenen Umformungen durch Multiplikation mit geeigneten Matrizen realisieren lassen. Hierfür ist es nützlich, zunächst eine Ringstruktur auf $\mathbb{K}^{n\times n}$ zu etablieren. Wir werden diese später im Kontext des Dualraums noch aus einer anderen Perspektive verständlich machen (vgl. auch den letzten Absatz von Kapitel 1.3).

Definition 3.1.8. Für $M = (\mu_{ij})$ und $N = (\nu_{ij})$ aus $K^{m\times n}$ (also $1 \leq i \leq m$ und $1 \leq j \leq n$) erklären wir die Addition und die Multiplikation mit einem Skalar $\lambda \in K$ durch $M + N = (\mu_{ij} + \nu_{ij})$ und $\lambda \cdot M = (\lambda \cdot \mu_{ij})$. Im Fall von $m = n$ sei weiter $\mathbb{1} = (\lambda_{ij}) \in K^{n\times n}$ die Einheitsmatrix mit $\lambda_{ii} = 1$ und $\lambda_{ij} = 0$ für $i \neq j$.

Als Übung möge man sich von den folgenden Tatsachen überzeugen.

Proposition 3.1.9. *(a) Mit der obigen Addition und Skalarmultiplikation ist $K^{m\times n}$ ein K-Vektorraum.*

(b) Mit der Addition und Multiplikation von Matrizen (siehe Definition 1.3.18 und die folgenden Ergebnisse) ist $K^{n\times n}$ ein Ring mit der Einheitsmatrix als Eins.

Wie oben versprochen, können wir nun die Matrizen angeben, die für die Umformung von Gleichungssystemen benötigt werden.

Definition 3.1.10. Für $1 \leq k, l \leq n$ definieren wir $E_{kl} = (e_{ij}) \in K^{n\times n}$ durch $e_{kl} = 1$ und $e_{ij} = 0$ für $(i, j) \neq (k, l)$. Für $\lambda \in K\backslash\{0\}$ und $k \neq l$ setzen wir

$$M_{k\lambda} = \mathbb{1} + (\lambda - 1) \cdot E_{kk},$$

[75]Das angegebene Vorgehen wird manchmal als Gaußsches Eliminationsverfahren oder auch als Gauß-Jordan-Verfahren nach Carl Friedrich Gauß und Wilhelm Jordan bezeichnet. Verfahren dieser Art tauchen aber beispielsweise auch bereits um das Jahr 0 in chinesischen Quellen auf. Wir werden in Lemma 3.1.12 zeigen, dass Zeilenvertauschungen prinzipiell verzichtbar sind, weil sie durch die beiden anderen Operationen simuliert werden können.

$$S_{kl} = \mathbb{1} - E_{lk},$$
$$P_{kl} = \mathbb{1} - E_{kk} - E_{ll} + E_{kl} + E_{lk}.$$

Matrizen dieser drei Typen nennen wir Elementarmatrizen.

Für das Gleichungssystem vom Beginn dieses Teilkapitels (bezeichnet durch die Matrix M) gilt

$$M_{3,-1/2} \cdot M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 1 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 1 \\ 1 & -1/2 & -1/2 \end{pmatrix}.$$

Wie im Abschnitt nach 3.1.7 wurde also hier die dritte Zeile mit $-1/2$ multipliziert. Weiter ergibt sich

$$\begin{aligned} S_{1,3} \cdot S_{1,2} \cdot M_{3,-1/2} \cdot M &= S_{1,3} \cdot \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 1 \\ 1 & -1/2 & -1/2 \end{pmatrix} \\ &= S_{1,3} \cdot \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 1 & -1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & -5/2 & 1/2 \end{pmatrix}. \end{aligned}$$

Hier wurde also – wie vorher bei der Lösung des Gleichungssystems – die erste von der zweiten und von der dritten Zeile abgezogen. Schließlich erhalten wir

$$P_{2,3} \cdot S_{1,3} \cdot S_{1,2} \cdot M_{3,-1/2} \cdot M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & -5/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -5/2 & 1/2 \\ 0 & 0 & 2 \end{pmatrix},$$

wobei also die zweite mit der dritten Zeile vertauscht wurde. Oben hatten wir die beiden letzten Zeilen noch mit $-2/5$ bzw. $1/2$ multipliziert, was sich durch Matrixmultiplikation mit $M_{2,-2/5}$ und $M_{3,1/2}$ realisieren lässt. Man möge sich nun auch vom allgemeinen Fall überzeugen:

Lemma 3.1.11. *Für $N = (\nu_{ij}) \in K^{m \times n}$ gilt*

$$M_{k\lambda} \cdot N = (\mu_{ij}) \quad \textit{with} \quad \mu_{ij} = \begin{cases} \lambda \cdot \nu_{kj} & \textit{if } i = k, \\ \nu_{ij} & \textit{otherwise}, \end{cases}$$

$$S_{kl} \cdot N = (\sigma_{ij}) \quad \textit{with} \quad \sigma_{ij} = \begin{cases} \nu_{lj} - \nu_{kj} & \textit{if } i = l, \\ \nu_{ij} & \textit{otherwise}, \end{cases}$$

$$P_{kl} \cdot N = (\pi_{ij}) \quad \textit{with} \quad \pi_{ij} = \begin{cases} \nu_{lj} & \textit{if } i = k, \\ \nu_{kj} & \textit{if } i = l, \\ \nu_{ij} & \textit{otherwise}. \end{cases}$$

Wie in Fußnote 75 versprochen, lassen sich Zeilenvertauschungen durch die anderen Operationen simulieren:

Lemma 3.1.12. *Es gilt*

$$P_{kl} = M_{l,-1} \cdot S_{kl} \cdot M_{l,-1} \cdot S_{lk} \cdot M_{l,-1} \cdot S_{kl}.$$

Beweis. Wir weisen nach, dass die Matrizen auf beiden Seiten bei Multiplikation mit den Einheitsvektoren e_i dasselbe Ergebnis liefern (und also gleiche Spalten

haben). Dies ist nur für $i \in \{k, l\}$ interessant. Hier gilt $P_{kl} \cdot e_k = e_l$ und $P_{kl} \cdot e_l = e_k$. Das vorige Lemma liefert außerdem

$$\begin{aligned}
S_{kl} \cdot e_k = e_k - e_l \quad &\text{und} \quad S_{kl} \cdot e_l = e_l, \\
M_{l,-1} \cdot (e_k - e_l) = e_k + e_l \quad &\text{und} \quad M_{l,-1} \cdot e_l = -e_l, \\
S_{lk} \cdot (e_k + e_l) = e_l \quad &\text{und} \quad S_{lk} \cdot (-e_l) = e_k - e_l, \\
M_{l,-1} \cdot e_l = -e_l \quad &\text{und} \quad M_{l,-1} \cdot (e_k - e_l) = e_k + e_l, \\
S_{kl} \cdot (-e_l) = -e_l \quad &\text{und} \quad S_{kl} \cdot (e_k + e_l) = e_k, \\
M_{l,-1} \cdot (-e_l) = e_l \quad &\text{und} \quad M_{l,-1} \cdot e_k = e_k.
\end{aligned}$$

Durch das Produkt der Matrizen werden also e_k und e_l vertauscht. □

In der Beispielrechnung vor Lemma 3.1.11 haben wir die Umformungen nur für die linke Seite der Gleichung $M \cdot v = w$ ausgeschrieben und also nur $E \cdot M$ für bestimmte Produkte E von Elementarmatrizen angegeben. Das ist insofern sachgemäß, als die Wahl der Elementarmatrizen nur von der Form von M und nicht von w abhängt. Wendet man die Umformungen auf beide Seiten an, so erhält man die Gleichung

$$E \cdot M \cdot v = E \cdot w.$$

Um zu zeigen, dass diese dieselbe Lösungsmenge hat, betrachten wir den folgenden Begriff. Man beachte, dass Theorem 1.3.16 die Wohldefiniertheit sichert.

Definition 3.1.13. Man nennt $M \in K^{n\times n}$ invertierbar, wenn $\mathrm{Abb}(M) : K^n \to K^n$ ein Isomorphismus ist. In diesem Fall definiert man die Matrix $M^{-1} \in K^{n\times n}$ durch die Setzung $\mathrm{Abb}(M^{-1}) = \mathrm{Abb}(M)^{-1}$. Wir schreiben $\mathrm{GL}_n(K)$ für die Menge der invertierbaren Matrizen in $K^{n\times n}$.

In Anbetracht von Beispiel 1.2.4(b) möge man sich von folgendem überzeugen.

Proposition 3.1.14. *Die Menge* $\mathrm{GL}_n(K)$ *mit der Matrixmultiplikation ist eine Gruppe, wobei* $\mathbb{1}$ *neutral und* M^{-1} *invers zu* M *ist.*

Eine geschlossene Formel für die Inverse einer Matrix wird sich im nächsten Teilkapitel mittels Determinanten ergeben. Alternativ kann die Inverse mit einem Algorithmus zur Lösung linearer Gleichungssysteme bestimmt werden, wie wir noch im vorliegenden Teilkapitel sehen werden.

Lemma 3.1.15. *Für* $M, N \in \mathrm{GL}_n(K)$ *gilt* $(M \cdot N)^{-1} = N^{-1} \cdot M^{-1}$.

Beweis. Mit Lemma 1.1.8 erhält man

$$\begin{aligned}
&\mathrm{Abb}\left((M \cdot N)^{-1}\right) = \mathrm{Abb}(M \cdot N)^{-1} = \left(\mathrm{Abb}(M) \circ \mathrm{Abb}(N)\right)^{-1} \\
&\quad = \mathrm{Abb}(N)^{-1} \circ \mathrm{Abb}(M)^{-1} = \mathrm{Abb}\left(N^{-1}\right) \circ \mathrm{Abb}\left(M^{-1}\right) = \mathrm{Abb}\left(N^{-1} \cdot M^{-1}\right).
\end{aligned}$$

Wegen Theorem 1.3.16 folgt die Behauptung. □

Für das Lösen linearer Gleichungssysteme ist folgendes entscheidend.

Proposition 3.1.16. *Jede Elementarmatrix ist invertierbar.*

Beweis. Es genügt, zu jeder Elementarmatrix $E \in K^{n\times n}$ eine Matrix $E' \in K^{n\times n}$ mit $E \cdot E' = \mathbb{1}$ zu finden. Dann ist nämlich $\mathrm{Abb}(E) \circ \mathrm{Abb}(E') = \mathrm{Abb}(\mathbb{1})$ die Identität. Somit ist $\mathrm{Abb}(E) : K^n \to K^n$ nach Lemma 1.1.6 surjektiv und nach Proposition 2.4.19 dann schon ein Isomorphismus.

Aus Lemma 3.1.11 ergibt sich direkt $M_{k\lambda} \cdot M_{k,1/\lambda} = \mathbb{1} = P_{kl} \cdot P_{kl}$ (beachte $\lambda \neq 0$ laut Definition 3.1.10), was man unter Verwendung von Proposition 3.1.9 und

$$E_{ij} \cdot E_{kl} = \begin{cases} E_{il} & \text{wenn } j = k, \\ 0 & \text{sonst} \end{cases}$$

auch explizit nachrechnen kann. Für die Elementarmatrix S_{kl} hat man wegen $k \neq l$ (laut Definition 3.1.10) zunächst $E_{lk}^2 = 0$ und also $S_{kl} \cdot (\mathbb{1} + E_{lk}) = \mathbb{1} - E_{lk}^2 = \mathbb{1}$. □

Wie versprochen, können wir folgern, dass unsere Zeilenoperationen die Lösungsmenge eines linearen Gleichungssystems nicht ändern.

Korollar 3.1.17. *Man betrachte $M \in K^{m\times n}$ und $w \in K^n$. Für jede Elementarmatrix $E \in K^{m\times m}$ gilt dann*

$$\mathrm{Lsg}(M, w) = \mathrm{Lsg}(E \cdot M, E \cdot w).$$

Beweis. Man hat

$$M \cdot v = w \;\Rightarrow\; E \cdot M \cdot v = E \cdot w \;\Rightarrow\; M \cdot v = E^{-1} \cdot E \cdot M \cdot v = E^{-1} \cdot E \cdot w = w,$$

wobei die vorige Proposition die Existenz der Inversen sichert. □

Es bleibt zu zeigen, dass jede Matrix durch Multiplikation mit Elementarmatrizen in Stufenform gebracht werden kann. Im folgenden ist $s(i)$ intuitiv die Position des ersten Eintrags ungleich Null in der i-ten Zeile, wobei wir $s(i) = n + 1$ setzen, wenn es keinen solchen Eintrag gibt.

Definition 3.1.18. Eine Matrix $M = (\lambda_{ij}) \in K^{m\times n}$ ist in Stufenform, wenn für

$$s(i) = s_M(i) = \min\big(\{j \in \{1, \dots, n\} : \lambda_{ij} \neq 0\} \cup \{n + 1\}\big)$$

jeweils $s(i) < s(i + 1)$ oder $s(i) = n + 1 = s(i + 1)$ gilt. Sie ist in normierter Stufenform, wenn für $s(i) < n + 1$ zusätzlich $\lambda_{i,s(i)} = 1$ gilt.

Das Lösungsverfahren für lineare Gleichungssysteme kann nun wie folgt beschrieben werden.

Theorem 3.1.19. *Für jedes $M \in K^{m\times n}$ gibt es Elementarmatrizen $E_k \in K^{m\times m}$, für die $E_l \cdot \ldots \cdot E_1 \cdot M$ in normierter Stufenform ist.*

Beweis. Für eine einheitliche Darstellung schreiben wir $i <_n j$, wenn $i < j \leq n+1$ oder $i = n+1 = j$ gilt. Per Induktion über $i \leq m$ wählen wir Elementarmatrizen E_k so, dass für $N(i) = E_{l(i)} \cdot \ldots \cdot E_1 \cdot M$ gilt:

(i) es ist $s_{N(i)}(1) <_n \ldots <_n s_{N(i)}(i)$ und $s_{N(i)}(i) <_n s_{N(i)}(j)$ für $i < j \leq m$,
(ii) für $1 \leq i' \leq i$ mit $s_{N(i)}(i') \leq n$ hat man $\lambda_{1,s_{N(i)}(i')} = 1$.

Als Induktionsanfang betrachten wir $i = 0$, wo nichts zu zeigen ist. Im Induktionsschritt wählen wir $j \in \{i+1, \dots, m\}$ so, dass $s_{N(i)}(j)$ kleinstmöglich ist. Gilt dabei $s_{N(i)}(j) = n + 1$, so sind wir fertig. Im Folgenden nehmen wir $s_{N(i)}(j) \leq n$ an. Für $N := P_{i+1,j} \cdot N(i)$ gilt dann neben $s_N(1) < \ldots < s_N(i+1)$ auch $s_N(i+1) \leq s_N(j)$ für $i + 1 < j \leq m$ (wie auch im Folgenden unter Verwendung von Lemma 3.1.11). Sei J die Menge aller $j \in \{i+1, \dots, m\}$ mit $s_N(j) = s_N(i+1)$. Für $j \in J$ sei $\lambda(j)$ der Eintrag von N an der Position $(j, s_N(j))$. Wir normieren nun die entsprechenden Zeilen J durch Multiplikation mit den Matrizen $M_{j,1/\lambda(j)}$ für $j \in J$. Schließlich ziehen wir durch Multiplikation mit $S_{i+1,j}$ für $j \in J\backslash\{i + 1\}$ jeweils die $(i + 1)$-te von der j-ten Zeile ab. Das resultierende Produkt $N(i + 1)$ erfüllt (i) und (ii). □

Im Fall des obigen Theorems ist der Beweis mindestens so wichtig wie das Ergebnis. Er ist nämlich konstruktiv und liefert ein explizites Verfahren für die Lösung von Gleichungssystemen. In der Praxis wird man das Verfahren wohl abkürzen (indem man etwa nicht an jeder Stelle normiert) und anders notieren (nämlich nicht mittels Elementarmatrizen sondern mit den gewohnten Zeilenumformungen). Welche Information anhand der normierten Stufenform ablesbar ist, lässt sich durch ein Beispiel am besten erklären.

Beispiel 3.1.20. Man betrachte die Matrix

$$M = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^3.$$

Gemäß Bemerkung 2.4.15 gilt $\operatorname{rank}(M) = 2$. In der Tat gilt allgemein $\operatorname{rank}(N) = k$, wenn N in Stufenform ist und genau k Zeilen hat, in denen nicht alle Einträge Null sind. Zum einen sind nämlich die k Spalten, welche jeweils den ersten Eintrag Eins einer Zeile enthalten, linear unabhängig (wie man prüfen möge). Zum anderen erhält man aus einer linear unabhängigen Menge von Spalten durch Weglassung der letzten $n - k$ Zeilen (nur Einträge Null) eine linear unabhängige Menge im $\mathbb{R}^k$.

Aus dem Rang lässt sich mit Korollar 3.1.6 und der Dimensionsformel die Dimension der Lösungsmengen ermitteln. Im vorliegenden Fall ergibt sich

$$\operatorname{Lsg}(M, w) \neq \emptyset \quad \Rightarrow \quad \dim(\operatorname{Lsg}(M, w)) = \dim(\mathbb{R}^3) - \operatorname{rank}(M) = 1.$$

Weiter gilt $\operatorname{Lsg}(M, w) = \emptyset$ genau dann, wenn der dritte Eintrag von w ungleich Null ist.[76] Für $w = (w_1, w_2, 0)^\mathsf{T}$ ist $v_0 := (w_1, 0, w_2)^\mathsf{T}$ eine spezielle Lösung der Gleichung $M \cdot v_0 = w$. Weiter besteht $\operatorname{Lsg}(M, 0)$ aus genau den Vektoren $(x, y, z)^\mathsf{T}$, für die $z = 0$ und $x = -2y$ gilt. Mit Lemma 3.1.5 erhält man also

$$\operatorname{Lsg}\left(M, \begin{pmatrix} w_1 \\ w_2 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} w_1 \\ 0 \\ w_2 \end{pmatrix} + \operatorname{Span}\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}\right).$$

Im Fall von invertierbaren Matrizen können wir das Ergebnis noch verbessern:

Theorem 3.1.21. *Eine Matrix in $K^{n\times n}$ ist invertierbar genau dann, wenn sie sich als Produkt von Elementarmatrizen schreiben lässt.*

Beweis. Produkte von Elementarmatrizen sind wegen Proposition 3.1.14 und 3.1.16 invertierbar. Umgekehrt finden wir für $M \in \mathrm{GL}_n(K)$ wegen Theorem 3.1.19 ein Produkt E von Elementarmatrizen, für das $E \cdot M$ in normierter Stufenform ist. Dabei ist $E \cdot M$ wieder invertierbar und kann also keine Zeile enthalten, in der alle Einträge Null sind (mit dem Argument aus Beispiel 3.1.20). Es muss dann also $E \cdot M = (\lambda_{ij})$ mit $\lambda_{ii} = 1$ und $\lambda_{ij} = 0$ für $i > j$ gelten ('obere Dreiecksform'). Wir müssen nun durch weitere Multiplikationen mit Elementarmatrizen erreichen, dass auch die Einträge λ_{ij} mit $i < j$ auf Null gesetzt werden (ohne dass die anderen Einträge sich ändern). Nehme dazu zunächst an, dass $\lambda_{n-1,n} \neq 0$ gilt. Wir können

[76] Hierbei ist zu beachten, dass man M in einer Anwendungen wohl als $M = E \cdot N$ für ein Produkt E von Elementarmatrizen erhalten hat. Dann ist also $\operatorname{Lsg}(N, w') = \operatorname{Lsg}(M, E \cdot w') = \emptyset$ genau dann, wenn der dritte Eintrag von $E \cdot w'$ ungleich Null ist, woraus man noch eine äquivalente Bedingung an w' herleiten muss (indem man etwa E^{-1} bestimmt und dann $E^{-1} \cdot (w_1', w_2', 0)^\mathsf{T}$ in Abhängigkeit von $w_i' \in \mathbb{R}$ berechnet).

diesen Eintrag Null setzen, indem wir von der vorletzten Zeile das $1/\lambda_{n-1,n}$-fache der letzten Zeile abziehen. Wie im Absatz nach Korollar 3.1.7 erklärt, lässt sich dies in der Tat durch ein Produkt von Elementarmatrizen bewerkstelligen. Analog setzt man alle relevanten Einträge von unten nach oben auf Null. Insgesamt erhält man also Elementarmatrizen E_k mit

$$E_l \cdot \ldots \cdot E_1 \cdot M = \mathbb{1}.$$

Entsprechend gilt

$$M = (E_l \cdot \ldots \cdot E_1)^{-1} = E_1^{-1} \cdot \ldots \cdot E_l^{-1},$$

wobei wir Lemma 3.1.15 verwendet haben. Laut dem Beweis von Proposition 3.1.16 sind $M_{k\lambda}^{-1} = M_{k,1/\lambda}$ und $P_{kl}^{-1} = P_{kl}$ wieder Elementarmatrizen. Aufgrund desselben Beweises erhalten wir

$$\begin{aligned} M_{k,-1} \cdot S_{kl} \cdot M_{k,-1} &= (\mathbb{1} - 2 \cdot E_{kk}) \cdot (\mathbb{1} - E_{lk}) \cdot (\mathbb{1} - 2 \cdot E_{kk}) \\ &= (\mathbb{1} - 2 \cdot E_{kk}) \cdot (\mathbb{1} - 2 \cdot E_{kk} - E_{lk} + 2 \cdot E_{lk}) \\ &= \mathbb{1} - 2 \cdot E_{kk} - E_{lk} + 2 \cdot E_{lk} - 2 \cdot E_{kk} + 4 \cdot E_{kk} = \mathbb{1} + E_{lk} = S_{kl}^{-1}, \end{aligned}$$

sodass also auch S_{kl}^{-1} ein Produkt von Elementarmatrizen ist. □

Auch bei diesem Theorem ist der Beweis mindestens so wichtig wie das Resultat. Für $M \in \mathrm{GL}_n(K)$ liefert er einen Algorithmus zur Bestimmung eines Produktes E von Elementarmatrizen mit $E \cdot M = \mathbb{1}$. Dieser berechnet also $E = M^{-1}$. In der Praxis möchte man mit Zeilenumformungen rechnen und die Elementarmatrizen aus E nicht explizit notieren. Als Trick bietet es sich an, die Zeilenumformungen nicht nur auf M sondern parallel auf $\mathbb{1}$ anzuwenden. Hat man $E \cdot M = \mathbb{1}$ erreicht, so gilt gleichzeitig natürlich $E \cdot \mathbb{1} = E$. Die Anwendung der Zeilenumformungen auf $\mathbb{1}$ bieten also eine Möglichkeit, die Matrix E zu bestimmen. Wir machen dies wieder anhand eines Beispiels verständlich.

Beispiel 3.1.22. Wir beginnen mit dem Schema

$$\left.\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array}\right.$$

Indem wir das Zweifache der ersten von der zweiten Zeile abziehen, erhalten wir

$$\left.\begin{array}{cc|cc} 1 & -1 & 1 & 0 \\ 0 & 3 & -2 & 1 \end{array}\right. .$$

Wenn wir die zweite Zeile durch drei teilen und zur ersten addieren, ergibt sich

$$\left.\begin{array}{cc|cc} 1 & 0 & 1/3 & 1/3 \\ 0 & 1 & -2/3 & 1/3 \end{array}\right. .$$

Man erhält also

$$\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1/3 & 1/3 \\ -2/3 & 1/3 \end{pmatrix} = \frac{1}{3} \cdot \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix}.$$

Alternativ kann man das Verfahren zur Bestimmung der Inversen so interpretieren, dass wir simultan die Gleichungen $M \cdot v_i = e_i$ für alle Einheitsvektoren e_i

lösen.[77] Für $N = (v_1 \mid \ldots \mid v_n)$ gilt dann also

$$\mathrm{Abb}(M) \circ \mathrm{Abb}(N)(e_i) = M \cdot N \cdot e_i = M \cdot v_i = e_i.$$

Hieraus folgt $\mathrm{Abb}(N) = \mathrm{Abb}(M)^{-1} = \mathrm{Abb}(M^{-1})$ und also $N = M^{-1}$.

3.2. **Die Determinante.** In diesem Teilkapitel betrachten wir die sogenannte Determinante $\det : K^{n\times n} \to K$, mit der sich insbesondere testen lässt, ob eine Matrix invertierbar ist. Die Determinante hat viele weitere Anwendungen in der linearen Algebra und in anderen Gebieten. Beispielsweise liefert die Determinante in der Analysis Aufschluss über lokale Extrema von Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ in mehreren Veränderlichen.

Wir führen die Determinante zunächst über ihre charakteristischen Eigenschaften ein, wie sie durch das folgende Resultat gegeben sind. Später werden wir sehen, wie sich Determinanten konkret berechnen lassen.

Theorem 3.2.1. *Es gibt genau eine Funktion $\Delta : (K^n)^n \to K$ mit den folgenden Eigenschaften:*

(i) für die Einheitsvektoren e_i gilt

$$\Delta(e_1, \ldots, e_n) = 1,$$

(ii) es gilt $\Delta(v_1, \ldots, v_n) = 0$, wann immer es $i < j$ gibt mit $v_i = v_j$,

(iii) die Abbildung Δ ist multilinear, das heißt, für $1 \leq i \leq n$ und feste Vektoren $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$ ist

$$K^n \ni v \mapsto \Delta(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n) \in K$$

stets eine lineare Abbildung.

Bevor wir das Theorem beweisen, merken wir an, dass $\Delta(v_1, \ldots, v_n)$ zumindest intuitiv das Volumen des von $v_1, \ldots, v_n$ aufgespannten Parallelepipeds ist (vgl. auch die untenstehenden Bemerkungen zu negativen Werten). In der Tat sollte klar sein, dass ein solches Volumen (wenn es denn überhaupt definierbar ist), die Eigenschaften (i) bis (iii) haben sollte. Insbesondere möchte man mit (ii), dass eine Figur, welche nicht die volle Dimension des Raumes ausschöpft, Volumen Null hat (sodass etwa ein zweidimensionales Parallelogramm innerhalb des dreidimensionalen Raumes kein Volumen hat). Hiermit ist auch bereits angedeutet, wie der oben erwähnte Test auf Invertierbarkeit funktioniert, da eine Matrix M ja genau dann invertierbar ist, wenn ihre Spaltenvektoren eine Figur voller Dimension aufspannen.

Fraglich mag einzig sein, ob man wie in (iii) auch negatives Volumen zulassen oder Beträge nehmen möchte. Hier ist zunächst anzumerken, dass man in allgemeinen Körpern ohnehin kein Vorzeichen ausmachen kann (denn in $\mathbb{F}_2 = \mathbb{B}$ gilt etwa $1 = -1$). Aber auch über Körpern wie $\mathbb{R}$ ist es sinnvoll, negative Werte von $\Delta(v_1, \ldots, v_n)$ zuzulassen. Das Vorzeichen kann dann nämlich als Orientierung der durch $v_1, \ldots, v_n$ aufgespannten Figur gedeutet werden, was unter anderem für die erwähnte Anwendung in der Analysis relevant ist. Das folgende Ergebnis zeigt, dass die Orientierung sich bei der Vertauschung von zwei Vektoren ändert.

[77] Hierbei verwendet man, dass die Zeilenumformungen, die man zur Lösung des Gleichungssystems $M \cdot v = w$ verwendet, nur von M und nicht von w abhängen (wie bereits vor Definition 3.1.13 bemerkt).

Lemma 3.2.2. *Hat Δ die Eigenschaften (ii) und (iii) aus dem vorigen Theorem, so gilt für $1 \leq i < j \leq n$ stets*

$$\Delta(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_n) = -\Delta(v_1, \dots, v_n).$$

Beweis. Wir schreiben

$$f(v, w) = \Delta(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_{j-1}, w, v_{j+1}, \dots, v_n).$$

Man erhält dann

$$\begin{aligned} 0 &= f(v_i + v_j, v_i + v_j) = f(v_i, v_i + v_j) + f(v_j, v_i + v_j) \\ &= f(v_i, v_i) + f(v_i, v_j) + f(v_j, v_i) + f(v_j, v_j) \\ &= f(v_i, v_j) + f(v_j, v_i). \end{aligned}$$

Es folgt $f(v_j, v_i) = -f(v_i, v_j)$, wie gewünscht. □

In Körpern wie $\mathbb{R}$, in denen $x + x = 0$ stets $x = 0$ impliziert, folgt auch umgekehrt (ii) aus der Eigenschaft im Lemma: Hat man nämlich $i < j$ mit $v_i = v_j$, so liefert die Vertauschung der beiden gleichen Vektoren

$$\Delta(v_1, \dots, v_n) = -\Delta(v_1, \dots, v_n)$$

und also $\Delta(v_1, \dots, v_n) + \Delta(v_1, \dots, v_n) = 0$. Hingegen ist die Eigenschaft aus dem Lemma etwa im Körper $\mathbb{B} = \mathbb{F}_2$ (vgl. Definition 1.2.1) schwächer als (ii). Beispielsweise sind für $\Delta : \mathbb{B}^2 \to \mathbb{B}$ mit

$$\Delta\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right) = (a_1 + a_2) \cdot (b_1 + b_2)$$

die Eigenschaften (i) und (iii) sowie die Eigenschaft aus dem Lemma erfüllt (da man in $\mathbb{B}$ stets $x = -x$ hat), während (ii) verletzt ist (da nämlich $(1+0)\cdot(1+0) = 1 \neq 0$ auch in $\mathbb{B}$ gilt).[78]

Beweis von Theorem 3.2.1. Um die Eindeutigkeit nachzuweisen, betrachten wir zwei Funktionen Δ und Δ', welche die gegebenen Bedingungen erfüllen. Wir müssen

$$\Delta(\bar{v}) = \Delta'(\bar{v})$$

für beliebiges $\bar{v} = (v_1, \dots, v_n)$ zeigen. Hierfür betrachten wir sukzessive allgemeinere Fälle. Im ersten Fall gelte $v_i = 0$ für ein i. Mit der Multilinearität erhält man dann $\Delta(\bar{v}) = 0 = \Delta'(\bar{v})$, wie im Beweis von Lemma 1.3.2. Im zweiten Fall gelte jeweils $v_i = e_{\pi(i)}$ für eine Funktion $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$. Ist π nicht injektiv, so liefert Bedingung (ii) wieder $\Delta(\bar{v}) = 0 = \Delta'(\bar{v})$. Wir betrachten nun den Fall, in dem π injektiv und also bijektiv ist. Hier argumentieren wir per Induktion über $|I_\pi|$ für

$$I_\pi := \{(i, j) : 1 \leq i < j \leq n \text{ und } \pi(i) > \pi(j)\}.$$

Gilt $|I_\pi| = 0$, so hat man jeweils $\pi(i) = i$ und also $\Delta(\bar{v}) = 1 = \Delta'(\bar{v})$ gemäß Bedingung (i). Für den Induktionsschritt mit $|I_\pi| > 0$ wählen wir ein beliebiges Element $(i_0, j_0) \in I_\pi$. Betrachte die Bijektion $\rho : \{1, \dots, n\} \to \{1, \dots, n\}$ mit

$$\rho(i) = \begin{cases} \pi(j_0) & \text{wenn } i = i_0, \\ \pi(i_0) & \text{wenn } i = j_0, \\ \pi(i) & \text{sonst.} \end{cases}$$

[78] Dieses Gegenbeispiel habe ich aus Vorlesungsnotizen von Jens Jordan übernommen.

Um die Induktionshypothese anwenden zu können, zeigen wir $|I_\rho| < |I_\pi|$. In Anbetracht von $(i_0, j_0) \notin I_\rho$ definieren wir $f : I_\rho \to I_\pi \backslash \{(i_0, j_0)\}$ durch

$$\begin{aligned} f((i,j)) &= (i,j) \quad \text{wenn} \quad i,j \notin \{i_0, j_0\}, \\ f((k,i_0)) = (k,j_0) \quad &\text{und} \quad f((k,j_0)) = (k,i_0) \quad \text{für} \quad k < i_0, \\ f((i_0,k)) = (i_0,k) \quad &\text{und} \quad f((k,j_0)) = (k,j_0) \quad \text{für} \quad i_0 < k < j_0, \\ f((i_0,k)) = (j_0,k) \quad &\text{und} \quad f((j_0,k)) = (i_0,k) \quad \text{für} \quad j_0 < k. \end{aligned}$$

Um zu prüfen, dass die Werte von f in I_π liegen, beachte man etwa

$$(i,i_0) \in I_\rho \quad \Rightarrow \quad \pi(i) = \rho(i) > \rho(i_0) = \pi(j_0) \quad \Rightarrow \quad (i,j_0) \in I_\pi.$$

Wegen $(i_0, j_0) \in I_\pi$ gilt für $i_0 < k < j_0$ außerdem

$$(i_0,k) \in I_\rho \quad \Rightarrow \quad \pi(i_0) > \pi(j_0) = \rho(i_0) > \rho(k) = \pi(k) \quad \Rightarrow \quad (i_0,k) \in I_\pi.$$

Die restlichen Fälle überlassen wir als Übung. Weiter möge man sich überzeugen, dass f injektiv ist. Wir können nun die Induktionshypothese anwenden. Mit Lemma 3.2.2 erhalten wir (immer noch für $\bar{v} = (v_1, \ldots, v_n)$ mit $v_i = e_{\pi(i)}$) die Gleichungen

$$\begin{aligned} \Delta(\bar{v}) &= \Delta\big(e_{\pi(1)}, \ldots, e_{\pi(n)}\big) = -\Delta\big(e_{\rho(1)}, \ldots, e_{\rho(n)}\big) \\ &= -\Delta'\big(e_{\rho(1)}, \ldots, e_{\rho(n)}\big) = \Delta'\big(e_{\pi(1)}, \ldots, e_{\pi(n)}\big) = \Delta'(\bar{v}). \end{aligned}$$

Als Grundlage für ein späteres Argument zeigen wir an dieser Stelle noch, dass $|I_\rho|$ genau dann gerade ist, wenn $|I_\pi|$ ungerade ist. Man betrachte

$$K = \{k \in \mathbb{N} : i_0 < k < j_0 \text{ und } \pi(i_0) > \pi(k) > \pi(j_0)\}$$

und überzeuge sich von

$$\operatorname{img}(f) = I_\pi \backslash \big(\{(i_0,j_0)\} \cup \{(i_0,k) : k \in K\} \cup \{(k,j_0) : k \in K\}\big).$$

Hieraus ergibt sich

$$|I_\rho| = |I_\pi| - 1 - 2 \cdot |K|.$$

Im dritten und etwas allgemeineren Spezialfall sei jeweils $v_i = \lambda_i \cdot e_{\pi(i)}$ für eine Funktion $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. Wir zeigen $\Delta(\bar{v}) = \Delta'(\bar{v})$ durch Induktion über $|\{i : \lambda_i \neq 1\}|$. Ist diese Größe Null, so sind wir im zuvor behandelten Fall. Anderenfalls wählen wir ein i mit $\lambda_i \neq 1$. Mit der Multilinearität erhalten wir induktiv

$$\begin{aligned} \Delta(\bar{v}) &= \Delta\big(v_1, \ldots, v_{i-1}, \lambda_i \cdot e_{\pi(i)}, v_{i+1}, \ldots, v_n\big) \\ &= \lambda_i \cdot \Delta\big(v_1, \ldots, v_{i-1}, e_{\pi(i)}, v_{i+1}, \ldots, v_n\big) \\ &= \lambda_i \cdot \Delta'\big(v_1, \ldots, v_{i-1}, e_{\pi(i)}, v_{i+1}, \ldots, v_n\big) = \Delta'(\bar{v}). \end{aligned}$$

Schließlich beweisen wir $\Delta(\bar{v}) = \Delta'(\bar{v})$ allgemein. Dazu schreiben wir temporär $\#v$ für die Anzahl der von Null erschiedenen Einträge eines Vektors $v \in K^n$. Wir argumentieren nun per Induktion über $\#v_1 + \ldots + \#v_n$. Sind wir in keinem der zuvor behandelten Fälle, so gibt es ein i, für das wir $v_i = u + w$ mit $\#u, \#w < \#v_i$ schreiben können. Erneut mit der Multilinearität ergibt sich induktiv

$$\begin{aligned} \Delta(\bar{v}) &= \Delta\big(v_1, \ldots, v_{i-1}, u + w, v_{i+1}, \ldots, v_n\big) \\ &= \Delta\big(v_1, \ldots, v_{i-1}, u, v_{i+1}, \ldots, v_n\big) + \Delta\big(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n\big) \\ &= \Delta'\big(v_1, \ldots, v_{i-1}, u, v_{i+1}, \ldots, v_n\big) + \Delta'\big(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n\big) = \Delta'(\bar{v}). \end{aligned}$$

Dies schließt den Beweis der Eindeutigkeit ab.

Für die Existenz definieren wir geeignete Funktionen $\Delta_n : (K^n)^n \to K$ per Rekursion über n. Hat man $n = 1$, so lassen sich die Vektoren in K^n eindeutig als $\lambda \cdot e_1$ schreiben. Wir setzen

$$\Delta_1(\lambda \cdot e_1) = \lambda.$$

Man möge prüfen, dass dies die Bedingungen aus Theorem 3.2.1 erfüllt. Als Vorbereitung auf den Rekursionsschritt definieren wir

$$\mathrm{hd}(v) = \lambda_1 \text{ und } \mathrm{tl}(v) := (\lambda_2, \dots, \lambda_{n+1})^\mathsf{T} \in K^n \text{ für } v = (\lambda_1, \dots, \lambda_{n+1})^\mathsf{T} \in K^{n+1}.$$

Nun definieren wir $\Delta_{n+1}(v_1, \dots, v_{n+1})$ rekursiv als

$$\sum_{j=1}^{n+1} (-1)^{1+j} \cdot \mathrm{hd}(v_j) \cdot \Delta_n\big(\mathrm{tl}(v_1), \dots, \mathrm{tl}(v_{j-1}), \mathrm{tl}(v_{j+1}), \dots, \mathrm{tl}(v_{n+1})\big).$$

Im Fall der Einheitsvektoren gilt $\mathrm{hd}(e_1) = 1$ und $\mathrm{hd}(e_j) = 0$ für $j > 1$. Weiter hat man $\mathrm{tl}(e_{j+1}) = e_j$. Bedingung (i) aus dem Theorem ist also induktiv erfüllt vermöge

$$\Delta_{n+1}\big(e_1, \dots, e_{n+1}\big) = (-1)^{1+1} \cdot \Delta_n\big(e_1, \dots, e_n\big) = 1.$$

Für Bedingung (ii) betrachten wir $\bar v = (v_1, \dots, v_{n+1})$ mit $v_i = v_j$ für ein Paar von Indizes $i < j$. Da dann auch $\mathrm{tl}(v_i) = \mathrm{tl}(v_j)$ gilt, ergibt sich induktiv

$$\begin{aligned}\Delta_{n+1}(\bar v) = \quad & (-1)^{1+i} \cdot \mathrm{hd}(v_i) \cdot \Delta_n\big(\mathrm{tl}(v_1), \dots, \mathrm{tl}(v_{i-1}), \mathrm{tl}(v_{i+1}), \dots, \mathrm{tl}(v_{n+1})\big) \\ & + (-1)^{1+j} \cdot \mathrm{hd}(v_j) \cdot \Delta_n\big(\mathrm{tl}(v_1), \dots, \mathrm{tl}(v_{j-1}), \mathrm{tl}(v_{j+1}), \dots, \mathrm{tl}(v_{n+1})\big).\end{aligned}$$

Die Listen der Argumente von Δ_n lassen sich durch $j - i - 1$ Vertauschungen ineinander überführen (tausche v_j aus der ersten Liste sukzessive mit $v_{j-1}, \dots, v_{i+1}$). Mit Lemma 3.2.2 folgt somit

$$\begin{aligned}&\Delta_n\big(\mathrm{tl}(v_1), \dots, \mathrm{tl}(v_{i-1}), \mathrm{tl}(v_{i+1}), \dots, \mathrm{tl}(v_{n+1})\big) \\ &\qquad = (-1)^{j-i-1} \cdot \Delta_n\big(\mathrm{tl}(v_1), \dots, \mathrm{tl}(v_{j-1}), \mathrm{tl}(v_{j+1}), \dots, \mathrm{tl}(v_{n+1})\big).\end{aligned}$$

In Anbetracht von $\mathrm{hd}(v_i) = \mathrm{hd}(v_j)$ hat man in $\Delta_{n+1}(\bar v)$ also einen Faktor

$$(-1)^{1+i} \cdot (-1)^{j-i-1} + (-1)^{1+j} = (-1)^j + (-1)^{1+j} = 0.$$

Dies ergibt $\Delta_{n+1}(\bar v) = 0$, wie von Bedingung (ii) gefordert. Bedingung (iii) prüft man durch eine Rechnung, die wir zur Übung überlassen. □

Die Determinante einer Matrix erklären wir über die Spaltenvektoren $M \cdot e_i$.

Definition 3.2.3. Wir definieren $\det : K^{n\times n} \to K$ durch

$$\det(M) = \Delta(M \cdot e_1, \dots, M \cdot e_n)$$

mit der Funktion Δ aus Theorem 3.2.1. Für $M = (\lambda_{ij}) \in K^{n\times n}$ schreibt man auch

$$\det(M) = \begin{vmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{n1} & \dots & \lambda_{nn} \end{vmatrix}.$$

Die Abbildung $\mathrm{Abb}(M) : K^n \to K^n$ 'verzerrt' den durch $e_1, \dots, e_n$ aufgespannten Einheitswürfel in das durch $M \cdot e_1, \dots, M \cdot e_n$ aufgespannte Parallelepiped. Gemäß der Bemerkung nach Theorem 3.2.1 misst die Determinante, wie diese Verzerrung das (orientierte) Volumen ändert.

Der Beweis der Existenz in Theorem 3.2.1 liefert einen rekursiven Algorithmus zur Bestimmung von Determinanten. Dieser wird durch das folgende Resultat explizit gemacht. Tatsächlich wurde im Beweis von Theorem 3.2.1 nur der Fall $i = 1$ betrachtet. Der allgemeine Fall ergibt sich analog, wobei allerdings zu beachten ist, dass das Vorzeichen zwischen geraden und ungeraden Zeilen wechseln muss (damit man stets $\Delta(e_1, \ldots, e_n) = 1$ erhält). Wir werden später noch ein ergänzendes Resultat beweisen, bei dem nach Spalten statt nach Zeilen entwickelt wird.

Theorem 3.2.4 (Laplacescher Entwicklungssatz, Teil 1[79])**.** *Die Determinante einer Matrix $M = (\mu_{ij}) \in K^{n\times n}$ mit $n > 1$ lässt sich für $1 \leq i \leq n$ durch Entwicklung nach der i-ten Zeile bestimmen als*

$$\det(M) = \sum_{j=1}^{n} (-1)^{i+j} \cdot \mu_{ij} \cdot \det\left(M_{ij}\right),$$

wobei $M_{ij} \in K^{(n-1)\times(n-1)}$ aus M entsteht, indem man die i-te Zeile und die j-te Spalte streicht.

Für niedrige Dimension geben wir konkrete Formeln an:

Beispiel 3.2.5. Es gilt

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = (-1)^{1+1} \cdot a_1 \cdot |b_2| + (-1)^{1+2} \cdot b_1 \cdot |a_2| = a_1 \cdot b_2 - a_2 \cdot b_1.$$

Weiter hat man

$$\begin{aligned} \begin{vmatrix} a_1 & b_1 & c_1 \\ c_2 & a_2 & b_2 \\ b_3 & c_3 & a_3 \end{vmatrix} &= a_1 \cdot \begin{vmatrix} a_2 & b_2 \\ c_3 & a_3 \end{vmatrix} - b_1 \cdot \begin{vmatrix} c_2 & b_2 \\ b_3 & a_3 \end{vmatrix} + c_1 \cdot \begin{vmatrix} c_2 & a_2 \\ b_3 & c_3 \end{vmatrix} \\ &= a_1 \cdot (a_2 a_3 - c_3 b_2) - b_1 \cdot (c_2 a_3 - b_3 b_2) + c_1 \cdot (c_2 c_3 - b_3 a_2) \\ &= a_1 a_2 a_3 + b_1 b_2 b_3 + c_1 c_2 c_3 - b_3 a_2 c_1 - c_2 b_1 a_3 - a_1 c_3 b_2. \end{aligned}$$

Durch die Benennung und farbliche Markierung der Einträge sollte hier gut ersichtlich sein, wie die Haupt- und Nebendiagonalen durchlaufen werden. Die Formel für Matrizen der Dimension 3×3 ist bekannt als Regel von Sarrus[80] (oder manchmal aus mnemotechnischen Gründen auch als Lattenzaun-Regel).

Für Matrizen in oberer oder unterer Dreiecksform ergibt sich die Determinante als Produkt der Diagonalelemente, wie man gut anhand eines Beispiels sehen kann:

Beispiel 3.2.6. Durch Entwicklung nach Zeilen – von der letzten hin zur ersten – ergibt sich

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & 0 & c_3 \end{vmatrix} = a_1 \cdot \begin{vmatrix} b_2 & c_2 \\ 0 & c_3 \end{vmatrix} = a_1 b_2 \cdot |c_3| = a_1 b_2 c_3.$$

Für untere Dreiecksmatrizen entwickelt man entsprechend von der ersten hin zur letzten Zeile.

Wir haben zuvor angemerkt, dass $\det(M)$ misst, wie stark M – oder genauer die assoziierte Abbildung $\mathrm{Abb}(M)$ – den Einheitswürfel verzerrt. Tatsächlich ist es willkürlich, hier den Würfel zu betrachten. Wir könnten als Ausgangspunkt genauso

[79] Nach Pierre-Simon Laplace.
[80] Nach Pierre Frédéric Sarrus.

gut ein anderes Parallelepiped betrachten und dessen Verzerrung messen. Diese Beobachtung hilft bei der geometrischen Interpretation des folgenden Theorems: Das Produkt $\det(M)\cdot\det(N)$ misst die Verzerrung, welche die Komposition von $\mathrm{Abb}(M)$ und $\mathrm{Abb}(N)$ in zwei Schritten bewirkt. Wegen $\mathrm{Abb}(M)\circ\mathrm{Abb}(N) = \mathrm{Abb}(M\cdot N)$ entspricht dies aber gerade der Verzerrung durch $M\cdot N$.

Theorem 3.2.7. *Für alle $M, N \in K^{n\times n}$ gilt* $\det(M\cdot N) = \det(M)\cdot\det(N)$.

Beweis. Wir betrachten zunächst den Fall, in dem N eine Elementarmatrix ist. Die Matrix $M_{k\lambda}$ enthält nur Einträge auf der Diagonalen, wobei ein Eintrag gleich λ ist und die übrigen Einträge Eins sind. Mit dem vorigen Beispiel hat man also $\det(M_{k,\lambda}) = \lambda$. Die Matrix $M\cdot M_{k\lambda}$ erhält man aus M, indem man die k-te Spalte von M mit λ multipliziert (so wie sich die Multiplikation von links laut Lemma 3.1.11 auf die k-te Zeile auswirkt). Wegen der Multilinearität der Determinante folgt also

$$\det(M\cdot M_{k\lambda}) = \lambda\cdot\det(M) = \det(M)\cdot\det(M_{k\lambda}).$$

Für $k\neq l$ hat man weiter $\det(S_{kl}) = 1$, da alle Einträge auf der Diagonalen Eins sind und man abseits der Diagonalen nur einen Eintrag ungleich Null hat, sodass die Matrix in Dreiecksform ist. Die Matrix $M\cdot S_{kl}$ entsteht aus M, indem man die l-te Spalte von der k-ten Spalte abzieht. Sei M' die Matrix, die aus M entsteht, wenn man die k-te Spalte durch die l-te Spalte ersetzt. Da M' dann zwei gleiche Spalten hat, gilt $\det(M') = 0$. Zusammen mit der Multilinearität erhält man

$$\det(M\cdot S_{kl}) = \det(M) - \det(M') = \det(M) = \det(M)\cdot\det(S_{kl}).$$

Die Elementarmatrizen P_{kl} müssen wir wegen Lemma 3.1.12 nicht berücksichtigen (wobei man mit dem Lemma $\det(P_{kl}) = -1$ erhält, was der Tatsache entspricht, dass in $M\cdot P_{kl}$ zwei Spalten vertauscht sind).

Wir betrachten nun den allgemeineren Fall, in dem N eine beliebige invertierbare Matrix ist. Theorem 3.1.21 liefert $N = E_1\cdot\ldots\cdot E_k$ für Elementarmatrizen E_i. Es ergibt sich

$$\begin{aligned}\det(M\cdot N) &= \det(M\cdot E_1\cdot\ldots\cdot E_{k-1})\cdot\det(E_k)\\ &= \ldots = \det(M)\cdot\det(E_1)\cdot\ldots\cdot\det(E_k)\\ &= \det(M)\cdot\det(E_1\cdot E_2)\cdot\det(E_3)\cdot\ldots\cdot\det(E_k)\\ &= \ldots = \det(M)\cdot\det(N).\end{aligned}$$

Abschließend untersuchen wir den Fall, in dem N nicht invertierbar ist. Hier ist $\mathrm{Abb}(N)$ und also auch $\mathrm{Abb}(M\cdot N) = \mathrm{Abb}(M)\circ\mathrm{Abb}(N)$ nicht injektiv, sodass $M\cdot N$ ebenfalls nicht invertierbar ist. Es genügt also, wenn wir zeigen, dass für $N\notin \mathrm{GL}_n(K)$ stets $\det(N) = 0$ gilt. Wir schreiben v_i für die Spalten $N\cdot e_i$ von N. Im nicht invertierbaren Fall können wir einen der Vektoren $v_1,\ldots,v_n$ als Linearkombination der anderen schreiben.Wir dürfen also etwa $v_n = \sum_{i=1}^{n-1}\lambda_i\cdot v_i$ annehmen. Die Multilinearität und Eigenschaft (ii) aus Theorem 3.2.1 liefern

$$\det(N) = \sum_{i=1}^{n-1}\lambda_i\cdot\Delta(v_1,\ldots,v_{n-1},v_i) = 0,$$

wie gewünscht. □

Es ergibt sich nun der folgende Test auf Invertierbarkeit.

Korollar 3.2.8. *Für $M \in K^{n\times n}$ hat man*

$$M \in \mathrm{GL}_n(K) \quad \Leftrightarrow \quad \det(M) \neq 0.$$

Weiter ist $\det(M^{-1}) = 1/\det(M)$ *für alle* $M \in \mathrm{GL}_n(K)$.

Beweis. Im vorigen Beweis haben wir schon die (Kontraposition der) Richtung von rechts nach links gezeigt. Ist umgekehrt M invertierbar, so führt das Theorem auf

$$1 = \det(\mathbb{1}) = \det(M) \cdot \det(M^{-1}),$$

sodass in der Tat $\det(M) \neq 0$ und $\det(M^{-1}) = 1/\det(M)$ gelten muss. □

Tatsächlich ist die Determinante mit der Multiplikation nicht nur verträglich sondern sogar in einem hohen Maß durch ihre multiplikative Struktur charakterisiert, wie das nächste Theorem zeigt. Um dieses zu formulieren, führen wir den folgenden Begriff ein.[81]

Definition 3.2.9. Eine Funktion $h : G_1 \to G_2$ zwischen Gruppen $(G_i, \circ_i)$ heißt Homomorphismus, wenn $h(g \circ_1 g') = h(g) \circ_2 h(g')$ für alle $g, g' \in G_1$ gilt.

Wir halten die folgenden elementaren Eigenschaften fest.

Lemma 3.2.10. *Für jeden Gruppenhomomorphismus $h : G_1 \to G_2$ gilt*

(a) $h(e_1) = e_2$ für die neutralen Elemente $e_i \in G_i$,
(b) $h(g^{-1}) = h(g)^{-1}$ für alle $g \in G_1$.

Beweis. (a) Dies gilt wegen der Eindeutigkeit des neutralen Elements oder explizit wegen

$$e_2 \circ_2 h(e_1) = h(e_1) = h(e_1 \circ_1 e_1) = h(e_1) \circ_2 h(e_1),$$

woraus die Aussage durch Kürzen folgt.

(b) Dies folgt aus der Eindeutigkeit des inversen Elements oder explizit mit

$$\begin{aligned} h(g^{-1}) &= e_2 \circ_2 h(g^{-1}) = h(g)^{-1} \circ_2 h(g) \circ_2 h(g^{-1}) \\ &= h(g)^{-1} \circ_2 h(g \circ_1 g^{-1}) = h(g)^{-1} \circ_2 h(e_1) = h(g)^{-1} \circ_2 e_2 = h(g)^{-1}, \end{aligned}$$

wobei wir Teil (a) verwendet haben. □

Das folgende Ergebnis zeigt insbesondere, dass die Determinante die stärkste Invariante auf invertierbaren Matrizen ist, welche sich mit der Multiplikation verträgt. Hat man im Folgenden nämlich $\det(M) = \det(N)$ mit $M, N \in \mathrm{GL}_n(K)$, so folgt $D'(M) = h \circ D(M) = h \circ D(N) = D'(N)$. Also kann kein Homomorphismus D' zwischen Matrizen unterscheiden, die nicht schon von det unterschieden werden.

Theorem 3.2.11. [82] *Sei $D : \mathrm{GL}_n(K) \to K^*$ die Einschränkung der Determinante auf invertierbare Matrizen,*[83] *wobei wir $K^* = K\backslash\{0\}$ als multiplikative Gruppe*

[81] Man vergleiche mit Definition 1.3.1 für den Fall von Vektorräumen.

[82] Der Autor des vorliegenden Skripts kennt dieses Resultat aus dem Buch von Bertram Huppert und Wolfgang Willems, Lineare Algebra, Vieweg+Teubner 2010[2]. Dort wird es Kurt Hensel zugeschrieben. Auch der Beweis ist (mit Modifikationen) aus dem genannten Buch übernommen.

[83] Also $D(M) = \det(M)$ für $M \in \mathrm{GL}_n(K)$.

auffassen. Dann hat D die folgende universelle Eigenschaft: Für jeden Gruppenhomomorphismus $D' : \mathrm{GL}_n(K) \to K^*$ gibt es einen eindeutig bestimmten Gruppenhomomorphismus $h : K^* \to K^*$ so, dass

$$\begin{array}{ccc} \mathrm{GL}_n(K) & \xrightarrow{D} & K^* \\ & \searrow^{D'} & \downarrow h \\ & & K^* \end{array}$$

ein kommutatives Diagramm ist, also so, dass $D' = h \circ D$ gilt. Hat dabei D' dieselbe universelle Eigenschaft wie D, so ist das genannte h ein Gruppenisomorphismus.[84]

Beweis. Wir zeigen zunächst, dass D die universelle Eigenschaft erfüllt. Sei dazu ein Gruppenhomomorphismus $D' : \mathrm{GL}_n(K) \to K^*$ gegeben. Wenn die Gleichung $D'(M_{1\lambda}) = h \circ D(M_{1\lambda})$ gelten soll, kommt wegen $\det(M_{1\lambda}) = \lambda$ höchstens

$$h(\lambda) = D'(M_{1\lambda})$$

als Definition von h in Frage. Das so definierte h ist jedenfalls ein Gruppenhomomorphismus, denn man kann $M_{1,\lambda\cdot\mu} = M_{1\lambda} \cdot M_{1\mu}$ prüfen und erhält dann

$$h(\lambda \cdot \mu) = D'(M_{1,\lambda\cdot\mu}) = D'(M_{1\lambda} \cdot M_{1\mu}) = D'(M_{1\lambda}) \cdot D'(M_{1\mu}) = h(\lambda) \cdot h(\mu).$$

Es bleibt zu zeigen, dass für $M \in \mathrm{GL}_n(K)$ stets $D'(M) = h \circ D(M)$ gilt. Gemäß Theorem 3.1.21 können wir $M = E_1 \cdot \ldots \cdot E_k$ für Elementarmatrizen E_i schreiben. Da D' ein Gruppenhomomorphismus ist, ergibt sich

$$D'(M) = D'(E_1) \cdot \ldots \cdot D'(E_k).$$

Selbiges gilt mit $h \circ D$ anstelle von D'. Laut Theorem 3.2.7 ist nämlich D und damit auch $h \circ D$ ein Gruppenhomomorphismus, wie man prüfen möge. Die Behauptung reduziert also auf $D'(E) = h \circ D(E)$ für Elementarmatrizen E. Wir betrachten zunächst den Fall $E = M_{k\lambda}$. Für $k = 1$ gilt die Behauptung wegen der Definition von h. Für $k \neq 1$ hat man $M_{k\lambda} = P_{1k} \cdot M_{1\lambda} \cdot P_{1k}$. Dies lässt sich durch Nachrechnen verifizieren, folgt aber auch aus der Überlegung, dass $M_{k\lambda}$ die k-te Zeile mit λ multipliziert, während P_{1k} die erste mit der k-ten Zeile vertauscht. Nun gilt $P_{kl}^2 = \mathbb{1}$ (siehe den Beweis von Proposition 3.1.16). Mit dem vorigen Lemma folgt

$$D'(P_{kl})^2 = D'(P_{kl}^2) = D'(\mathbb{1}) = 1$$

und also

$$D'(M_{k\lambda}) = D'(P_{1k}) \cdot D'(M_{1\lambda}) \cdot D'(P_{1k}) = D'(M_{1\lambda}) = h(\lambda) = h \circ D(M_{k\lambda}).$$

Als nächstes betrachten wir $E = S_{kl}$. Sei allgemeiner $S_{kl\mu}$ die Matrix, welche das μ-fache der k-ten von der l-ten Zeile abzieht, also $S_{kl\mu} = M_{k,1/\mu} \cdot S_{kl} \cdot M_{k\mu}$. Weil nach dem vorigen Lemma $h(\mu^{-1}) = h(\mu)^{-1}$ gilt, erhält man

$$D'(S_{kl\mu}) = D'(M_{k,1/\mu}) \cdot D'(S_{kl}) \cdot D'(M_{k,\mu}) = h\left(\frac{1}{\mu}\right) \cdot D'(S_{kl}) \cdot h(\mu) = D'(S_{kl}).$$

Das gewünschte Ergebnis ist unmittelbar, wenn K^* nur ein Element hat. Im verbliebenen Fall betrachten wir $\mu = (\lambda + 1)/\lambda$ für ein $\lambda \in K^*$ mit $\lambda \neq -1$. Es gilt dann

$$S_{kl} \cdot S_{kl\lambda} = M_{k,1/\mu} \cdot S_{kl\lambda} \cdot M_{k,\mu}.$$

[84] Der Begriff des Gruppenisomorphismus ist durch Fußnote 36 geklärt.

Die linke Seite zieht nämlich insgesamt das $(\lambda + 1)$-fache der k-ten von der l-ten Zeile ab. Die rechte Seite tut effektiv dasselbe, indem sie zunächst die k-te Zeile mit μ multipliziert, dann das λ-fache der modifizierten k-ten Zeile (und also das $\lambda \cdot \mu = (\lambda + 1)$-fache der ursprünglichen k-ten Zeile) von der l-ten Zeile abzieht, bevor die ursprüngliche k-te Zeile durch Multiplikation mit $1/\mu$ wieder hergestellt wird. Wir erhalten nun

$$D'(S_{kl}) \cdot D'(S_{kl\lambda}) = h\left(\frac{1}{\mu}\right) \cdot D'(S_{kl\lambda}) \cdot h(\mu) = D'(S_{kl\lambda}).$$

Durch Kürzen und mit dem vorigen Lemma ergibt sich

$$D'(S_{kl}) = 1 = h(1) = h \circ D(S_{kl}).$$

Der Fall $E = P_{kl}$ muss wegen Lemma 3.1.12 nicht betrachtet werden.

Hat schließlich D' dieselbe universelle Eigenschaft, so erhält man eindeutige Gruppenhomomorphismen h und h' so, dass

$$\begin{array}{ccc} & & K^* \\ & \overset{D}{\nearrow} & \downarrow{\scriptstyle h} \\ \mathrm{GL}_n(K) & \xrightarrow{D'} & K^* \\ & \underset{D}{\searrow} & \downarrow{\scriptstyle h'} \\ & & K^* \end{array}$$

ein kommutatives Diagramm ist. Andererseits ist auch die Identität $\mathrm{id} : K^* \to K^*$ ein Gruppenhomomorphismus mit der Eigenschaft, dass

$$\begin{array}{ccc} \mathrm{GL}_n(K) & \xrightarrow{D} & K^* \\ & \underset{D}{\searrow} & \downarrow{\scriptstyle \mathrm{id}} \\ & & K^* \end{array}$$

kommutiert. Wegen der Eindeutigkeit ergibt sich also $h' \circ h = \mathrm{id}$ und genauso auch $h \circ h' = \mathrm{id}$, sodass h ein Isomorphismus ist.[85]

□

Im Rest dieses Kapitels diskutieren wir weitere Methoden zur Berechnung der Determinanten sowie den Bezug zu linearen Gleichungssystemen. Zunächst zeigen wir, dass die Rollen der Zeilen und Spalten austauschbar sind.

Proposition 3.2.12 (Laplacescher Entwicklungssatz, Teil 2)**.** *Die Determinante einer Matrix $M = (\mu_{ij}) \in K^{n\times n}$ mit $n > 1$ lässt sich für $1 \leq j \leq n$ durch Entwicklung nach der j-ten Spalte bestimmen als*

$$\det(M) = \sum_{i=1}^{n} (-1)^{i+j} \cdot \mu_{ij} \cdot \det\big(M_{ij}\big).$$

Hierbei ist $M_{ij} \in K^{(n-1)\times(n-1)}$ wieder die Matrix, welche aus M entsteht, indem man die i-te Zeile und die j-te Spalte streicht.

[85]Man beachte, dass der letzte Teil des Arguments nicht von der spezifischen Situation abhängt, sondern sich analog auf andere universelle Eigenschaften anwenden lässt.

Beweis. Die Transponierte einer Matrix $A = (\alpha_{ij}) \in K^{m\times n}$ ist definiert als die Matrix $A^\mathsf{T} = (\gamma_{ij}) \in K^{n\times m}$ mit $\gamma_{ij} = \alpha_{ji}$, wobei also Zeilen und Spalten vertauscht werden.[86] Durch Rechnung überzeuge man sich von $(A \cdot B)^\mathsf{T} = B^\mathsf{T} \cdot A^\mathsf{T}$. Für die Elementarmatrizen hat man $M_{k\lambda}^\mathsf{T} = M_{k\lambda}$ und $S_{kl}^\mathsf{T} = S_{lk}$ sowie $P_{kl}^\mathsf{T} = P_{kl}$. Ist nun $M \in K^{n\times n}$ invertierbar und also Produkt $E_1 \cdot \ldots \cdot E_k$ von Elementarmatrizen, so ist

$$M^\mathsf{T} = E_k^\mathsf{T} \cdot \ldots \cdot E_1^\mathsf{T}$$

wieder invertierbar. Wegen $(M^\mathsf{T})^\mathsf{T} = M$ folgt auch die andere Richtung, das heißt, M ist genau dann invertierbar, wenn dasselbe für M^T gilt. Wir folgern nun

$$\det\big(M^\mathsf{T}\big) = \det\big(M\big).$$

Wegen des eben Gezeigten ist nur $M \in \mathrm{GL}_n(M)$ zu betrachten. Hier schreiben wir M wie oben als Produkt von Elementarmatrizen E_i. Für letztere hat man jeweils $\det(E_i^\mathsf{T}) = \det(E_i)$. Mit der Multiplikativität der Determinanten ergibt sich

$$\det\big(M^\mathsf{T}\big) = \det\big(E_k^\mathsf{T}\big) \cdot \ldots \cdot \det\big(E_1^\mathsf{T}\big) = \det(E_1) \cdot \ldots \cdot \det(E_k) = \det(M).$$

Die Proposition folgt nun, weil die Entwicklung von $M = (\mu_{ij})$ nach Spalten der Entwicklung von $N = M^\mathsf{T} = (\nu_{ij})$ nach Zeilen entspricht. Explizit hat man nämlich $\nu_{ij} = \mu_{ji}$ und $N_{ji} = {M_{ij}}^\mathsf{T}$. Mit Theorem 3.2.4 erhält man

$$\det(M) = \det(N) = \sum_{i=1}^{n} (-1)^{j+i} \cdot \nu_{ji} \cdot \det(N_{ji}) = \sum_{i=1}^{n} (-1)^{i+j} \cdot \mu_{ij} \cdot \det({M_{ij}}^\mathsf{T}),$$

was wegen $\det({M_{ij}}^\mathsf{T}) = \det(M_{ij})$ der Behauptung entspricht. □

Das vorige Ergebnis ist nützlich, wenn es eine Spalte (aber keine Zeile) gibt, in der viele Einträge Null sind.

Beispiel 3.2.13. Entwicklung nach der zweiten Spalte ergibt

$$\begin{vmatrix} 1 & 0 & 2 \\ 2 & 3 & -1 \\ -1 & 0 & 4 \end{vmatrix} = 3 \cdot \begin{vmatrix} 1 & 2 \\ -1 & 4 \end{vmatrix} = 18,$$

wie man in diesem Fall freilich auch schnell mit der Regel von Sarrus sieht.

Direkt aus der Multilinearität der Determinanten ergibt sich:

Proposition 3.2.14 (Cramersche Regel[87])**.** *Für $M \in \mathrm{GL}_n(K)$ und $b \in K^n$ ist die eindeutige Lösung von $M \cdot x = b$ gegeben durch*

$$x = (x_1, \ldots, x_n)^\mathsf{T} \quad \textit{mit} \quad x_i = \frac{\det(M_i(b))}{\det(M)},$$

wobei $M_i(b)$ aus M entsteht, indem die i-te Spalte durch b ersetzt wird.

Beweis. Wir schreiben $M = (\lambda_{ij})$ und rechnen

$$M \cdot x = \begin{pmatrix} \lambda_{11} \cdot x_1 + \ldots + \lambda_{1n} \cdot x_n \\ \vdots \\ \lambda_{n1} \cdot x_1 + \ldots + \lambda_{nn} \cdot x_n \end{pmatrix} = \sum_{j=1}^{n} x_j \cdot \begin{pmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{nj} \end{pmatrix} = \sum_{j=1}^{n} x_j \cdot M \cdot e_j.$$

[86]Wir haben den Begriff der Transponierten bisher vermieden, weil er sich besonders gut im Zusammenhang mit dem Dualraum motivieren lässt. Diesen Raum werden wir später untersuchen.
[87]Nach Gabriel Cramer.

In der Notation aus Definition 3.2.3 folgt aus $M \cdot x = b$ also

$$\begin{aligned}\det(A_i(b)) &= \Delta\left(M \cdot e_1, \ldots, M \cdot e_{i-1}, \textstyle\sum_{j=1}^n x_j \cdot M \cdot e_j, M \cdot e_{i+1}, \ldots, M \cdot e_n\right) \\ &= \sum_{j=1}^n x_j \cdot \Delta(M \cdot e_1, \ldots, M \cdot e_{i-1}, M \cdot e_j, M \cdot e_{i+1}, \ldots, M \cdot e_n),\end{aligned}$$

wobei wir die Multilinearität verwendet haben. Die Summanden zu $j \neq i$ sind Null, weil das Argument $M \cdot e_j$ hier doppelt vorkommt. Man erhält also jeweils

$$\det(A_i(b)) = x_i \cdot \det(M).$$

Wegen $M \in \mathrm{GL}_n(K)$ können wir durch $\det(M) \neq 0$ teilen. □

Auch das folgende Resultat wird manchmal als Cramersche Regel bezeichnet.

Korollar 3.2.15. *Für $M \in \mathrm{GL}_n(K)$ gilt*

$$M^{-1} = (\lambda_{ij}) \quad \textit{mit} \quad \lambda_{ij} = (-1)^{i+j} \cdot \frac{\det(M_{ji})}{\det(M)},$$

wobei M_{ji} wieder durch Streichen der j-ten Zeile und i-ten Spalte entsteht.

Beweis. Durch Entwicklung nach der i-ten Spalte erhält man

$$\det(M_i(e_j)) = (-1)^{i+j} \cdot \det(M_{ji}).$$

Für $N = (\lambda_{ij})$ mit den gegebenen λ_{ij} löst deshalb $y = (\lambda_{1j}, \ldots, \lambda_{nj})^\mathsf{T} = N \cdot e_j$ laut der vorigen Proposition die Gleichung $M \cdot y = e_j$. Man hat also jeweils $M \cdot N \cdot e_j = e_j$, sodass N die Inverse von M ist. □

Neben dem rekursiven Verfahren von Laplace kann die Determinante auch durch folgende explizite Formel beschrieben werden.

Theorem 3.2.16 (Leibniz-Formel[88])**.** *Für $M = (\lambda_{ij}) \in K^{n \times n}$ gilt*

$$\begin{aligned}\det(M) = \sum_{\pi \in S(n)} (-1)^{N(\pi)} \cdot \lambda_{\pi(1),1} \cdot \ldots \cdot \lambda_{\pi(n),n} \\ \textit{mit} \quad N(\pi) = |\{(i,j) : 1 \leq i < j \leq n \textit{ und } \pi(i) > \pi(j)\}|,\end{aligned}$$

wobei $S(n)$ die Menge der Bijektionen $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ bezeichnet.

Wir merken an, dass dieselbe Formel mit $\lambda_{i,\pi(i)}$ an der Stelle von $\lambda_{\pi(i),i}$ gilt. Dies folgt aus $\det(M) = \det(M^\mathsf{T})$, was im Beweis von Proposition 3.2.12 gezeigt wurde.

Beweis. Wir schreiben $\det'(M)$ für die rechte Seite der zu zeigenden Gleichung. Wegen der Eindeutigkeit in Theorem 3.2.1 genügt es, die folgenden Eigenschaften für jede Matrix M nachzuweisen:

(i) es gilt $\det'(\mathbb{1}) = 1$,
(ii) man hat $\det'(M) = 0$ wenn es $i < j$ mit $M \cdot e_i = M \cdot e_j$ gibt,
(iii) für $1 \leq j \leq n$ ist $b \mapsto \det'(M_j(b))$ linear, wobei $M_j(b)$ aus M entsteht, indem die j-te Spalte durch b ersetzt wird.

[88] Nach Gottfried Wilhelm Leibniz.

Für $\mathbb{1} = (\lambda_{ij})$ ist $\lambda_{\pi(1),1} \cdot \ldots \cdot \lambda_{\pi(n),1}$ genau dann Eins und nicht Null, wenn π die Identität ist. Hier gilt $N(\pi) = 0$ und also

$$\det{}'(\mathbb{1}) = (-1)^0 \cdot 1 = 1,$$

wie für Eigenschaft (i) benötigt. Sind i, j wie in Punkt (ii) fixiert, so definieren wir für $\pi \in S(n)$ jeweils $\pi' \in S(n)$ durch

$$\pi'(k) = \begin{cases} \pi(j) & \text{wenn } k = i, \\ \pi(i) & \text{wenn } k = j, \\ \pi(k) & \text{sonst.} \end{cases}$$

Wegen $\pi'' = \pi$ ist die Operation $\pi \mapsto \pi'$ eine Bijektion auf $S(n)$. Für den Fall $\pi(i) > \pi(j)$ haben wir im Beweis von Theorem 3.2.1 gezeigt, dass $N(\pi)$ genau dann gerade ist, wenn $N(\pi')$ ungerade ist. Selbiges gilt auch im Fall $\pi(i) < \pi(j)$, wo ja $\pi'(i) > \pi'(j)$ gilt. Wir schreiben nun $M = (\lambda_{ij})$ und erhalten wegen $M \cdot e_i = M \cdot e_j$ also $\lambda_{ki} = \lambda_{kj}$ für alle k. Hieraus folgt jeweils

$$\lambda_{\pi(i),i} \cdot \lambda_{\pi(j),j} = \lambda_{\pi(i),j} \cdot \lambda_{\pi(j),i} = \lambda_{\pi'(j),j} \cdot \lambda_{\pi'(i),i}$$

und also

$$\lambda_{\pi(1),1} \cdot \ldots \cdot \lambda_{\pi(n),n} = \lambda_{\pi'(1),1} \cdot \ldots \cdot \lambda_{\pi'(n),n}.$$

Für $E(n) = \{\pi \in S(n) : N(\pi) \text{ ist gerade}\}$ erhält man also

$$\det{}'(M) = \textstyle\sum_{\pi \in E(n)} \left((-1)^{N(\pi)} + (-1)^{N(\pi')} \right) \cdot \lambda_{\pi(1),1} \cdot \ldots \cdot \lambda_{\pi(n),n} = 0.$$

Die Verifikation von (iii) durch explizite Rechnung überlassen wir als Übung. □

Im Fall von $n = 3$ liefert die Leibniz-Formel die $|S(3)| = 6$ Summanden der Regel von Sarrus, als deren (weitreichende) Verallgemeinerung sie angesehen werden kann. Wegen des enormen Wachstums von $|S(n)| = n! = 1 \cdot 2 \cdot \ldots \cdot n$ ist die Leibniz-Formel und auch die Cramersche Regel für konkrete Berechnungen meist weniger geeignet als etwa das Gaußsche Eliminationsverfahren zur Bestimmung der Stufenform. Hingegen ist die Leibniz-Formel von großem theoretischen Interesse. Sie zeigt etwa, dass die Determinante $\det : K^{n \times n} \to K$ für festes $n \in \mathbb{N}$ als Polynom in n^2 Variablen aufgefasst werden kann und etwa für $K = \mathbb{R}$ in einem geeigneten Sinn stetig ist – ebenso wie mit Korollar 3.2.15 die Operation $A \mapsto A^{-1}$ auf $\mathrm{GL}_n(\mathbb{R})$.[89]

[89]Diese Bemerkung setzt ein gewisses Vorwissen aus der Analysis voraus. Der Autor stellt in seinen Prüfungen zur linearen Algebra daher keine Aufgaben zur Stetigkeit der genannten Operationen.

4. Algebraische Konstruktionen auf Vektorräumen

Über einem festen Körper K gibt es bis auf Isomorphie genau einen Vektorraum gegebener Dimension. Dieser kann allerdings auf vielfältige Weise beschrieben werden. Zwischen verschiedenen Beschreibungen wechseln zu können, ist für Beweise und Rechnungen ein entscheidender Vorteil. In diesem Kapitel behandeln wir zentrale Konstruktionen, die sich zur Beschreibung von Vektorräumen nutzen lassen.

4.1. **Dualität.** Wir rufen in Erinnerung, dass mit $\mathrm{Hom}(V,W)$ die Menge der linearen Abbildungen (oder synonym der Homomorphismen) zwischen zwei Vektorräumen V und W über einem gemeinsamen Körper bezeichnet wird. Eine zentrale Idee dieses Unterkapitels ist, dass man $\mathrm{Hom}(V,W)$ selbst wieder mit einer Vektorraumstruktur ausstatten kann.

Definition 4.1.1. Es seien V und W zwei Vektorräume über einem Körper K. Auf $\mathrm{Hom}(V,W)$ definieren wir eine Addition und eine Skalarmultiplikation durch

$$\begin{aligned}(\varphi+\psi)(v) &:= \varphi(v)+\psi(v) && \text{für } \varphi,\psi\in\mathrm{Hom}(V,W) \text{ sowie } v\in V,\\ (\lambda\cdot\varphi)(v) &:= \lambda\cdot\varphi(v) && \text{für } \lambda\in K \text{ und } \varphi\in\mathrm{Hom}(V,W) \text{ sowie } v\in V.\end{aligned}$$

Für $W=K$ nennen wir $V^*:=\mathrm{Hom}(V,K)$ (mit den eben definierten Operationen) den Dualraum von V.

Die Bezeichnung von V^* als Raum ist mit dem folgenden Resultat gerechtfertigt.

Proposition 4.1.2. *Mit den obigen Operationen ist* $\mathrm{Hom}(V,W)$ *ein Vektorraum. Insbesondere ist der Dualraum jedes K-Vektorraums wieder ein K-Vektorraum.*

Beweis. Wir rechnen beispielhaft

$$\begin{aligned}\big(\lambda\cdot(\varphi+\psi)\big)(v) &= \lambda\cdot(\varphi+\psi)(v) = \lambda\cdot(\varphi(v)+\psi(v))\\ &= \lambda\cdot\varphi(v)+\lambda\cdot\psi(v) = (\lambda\cdot\varphi)(v)+(\lambda\cdot\psi)(v) = (\lambda\cdot\varphi+\lambda\cdot\psi)(v)\end{aligned}$$

für beliebiges $v\in V$ (wobei man sich die genaue Begründung jedes Schrittes vor Augen führen möge). Dies zeigt $\lambda\cdot(\varphi+\psi)=\lambda\cdot\varphi+\lambda\cdot\psi$, was eine der Bedingungen an einen Vektorraum ist. Die übrigen Bedingungen möge man als Übung prüfen. □

In Definition 3.1.8 haben wir bereits eine explizite Beschreibung der Addition und Skalarmultiplikation auf Matrizen gegeben, mit der die Menge $K^{m\times n}$ laut Proposition 3.1.9 zu einem Vektorraum wird. Unsere neue Konstruktion kann in folgendem Sinn als Verallgemeinerung aufgefasst werden.

Proposition 4.1.3. *Die bijektive Funktion* $\mathrm{Abb}: K^{m\times n}\to\mathrm{Hom}(K^n,K^m)$ *aus Theorem 1.3.16 ist ein Isomorphismus von Vektorräumen.*

Beweis. Um zu zeigen, dass $\mathrm{Abb}(M+N)$ und $\mathrm{Abb}(M)+\mathrm{Abb}(N)$ übereinstimmen, prüfen wir, dass die beiden linearen Abbildungen auf jedem Vektor e_i aus der Standardbasis denselben Wert annehmen. Der Vektor

$$\mathrm{Abb}(M+N)(e_i) = (M+N)\cdot e_i$$

ist durch die i-te Spalte von $M+N$ gegeben. Diese stimmt aber laut Definition 3.1.8 mit der Summe

$$M\cdot e_i+N\cdot e_i = \mathrm{Abb}(M)(e_i)+\mathrm{Abb}(N)(e_i) = \big(\mathrm{Abb}(M)+\mathrm{Abb}(N)\big)(e_i)$$

der i-ten Spalten von M und N überein. Analog ist $\mathrm{Abb}(\lambda\cdot M)=\lambda\cdot\mathrm{Abb}(M)$. □

Für $m = n$ trägt $K^{n\times n}$ laut Proposition 3.1.9 sogar eine Ringstruktur (bezüglich der Matrixmultiplikation). Wegen der Gleichung $\mathrm{Abb}(M \cdot N) = \mathrm{Abb}(M) \circ \mathrm{Abb}(N)$ aus Definition 1.3.18 wird der Raum $\mathrm{Hom}(V, V)$ für $V = K^n$ zu einem isomorphen Ring mit der Komposition von Abbildungen als multiplikativer Verknüpfung. Wir verzichten an dieser Stelle auf eine Erweiterung auf allgemeines V. Es ist aber eine Bemerkung wert, dass die Ringstruktur (die wir bei Proposition 3.1.9 als Übung aufgegeben hatten) sich auf Ebene von linearen Abbildungen mit weniger Rechenaufwand nachweisen lässt. Für Homomorphismen φ und ψ, ψ' mit passendem Definitions- und Wertebereich hat man etwa

$$\varphi \circ (\psi + \psi')(v) = \varphi\big(\psi(v) + \psi'(v)\big) = \varphi \circ \psi(v) + \varphi \circ \psi'(v),$$

woraus sich für Matrizen unmittelbar $M \cdot (N + N') = M \cdot N + M \cdot N'$ ergibt.

Im Folgenden legen wir unseren Fokus auf den Dualraum. Die zu verschiedenen Vektorräumen gehörigen Dualräume lassen sich wie folgt in Beziehung setzen.

Definition 4.1.4. Ist $\varphi : V \to W$ eine lineare Abbildung, so sei $\varphi^* : W^* \to V^*$ durch $\varphi^*(\psi) = \psi \circ \varphi$ gegeben.

In der Sprache der Kategorientheorie (die wir im Folgenden hin und wieder erwähnen aber ansonsten nicht voraussetzen) sagt das folgende Ergebnis, dass die Konstruktion des Dualraums ein kontravarianter Funktor ist.

Lemma 4.1.5. *(a) Für jede lineare Abbildung φ ist auch φ^* wieder linear.*

(b) Ist $\mathrm{id} : V \to V$ *die Identitätsabbildung auf V, so ist* $\mathrm{id}^* : V^* \to V^*$ *die Identitätsabbildung auf V^*.*

(c) Für lineare Abbildungen $\theta : U \to V$ und $\varphi : V \to W$ gilt $(\varphi \circ \theta)^* = \theta^* \circ \varphi^*$.

Beweis. (a) Für $\varphi : V \to W$ und $\psi, \psi' \in W^*$ sowie beliebiges $v \in V$ rechnet man

$$\varphi^*(\psi + \psi')(v) = (\psi + \psi') \circ \varphi(v) = \psi \circ \varphi(v) + \psi' \circ \varphi(v) = \big(\varphi^*(\psi) + \varphi^*(\psi')\big)(v),$$

was $\varphi^*(\psi + \psi') = \varphi^*(\varphi) + \varphi^*(\psi')$ zeigt. Analog erhält man $\varphi^*(\lambda \cdot \psi) = \lambda \cdot \varphi^*(\psi)$.

(b) Es gilt $\mathrm{id}^*(\psi) = \psi \circ \mathrm{id} = \psi$.

(c) Man rechnet

$$(\varphi \circ \theta)^*(\psi) = \psi \circ (\varphi \circ \theta) = (\psi \circ \varphi) \circ \theta = \theta^*(\psi \circ \varphi) = \theta^*(\varphi^*(\psi)) = \theta^* \circ \varphi^*(\psi)$$

für beliebiges $\psi \in W^*$. □

Die folgende Tatsache kann man nicht nur für den Dualraum sondern analog für jeden Funktor ableiten.

Korollar 4.1.6. *Aus $V \cong W$ folgt $V^* \cong W^*$.*

Beweis. Sei $\varphi : V \to W$ ein Isomorphismus. Es ist dann

$$\varphi^* \circ (\varphi^{-1})^* = (\varphi^{-1} \circ \varphi)^* = \mathrm{id}^*$$

die Identitätsabbildung auf V^* und ebenso $(\varphi^{-1})^* \circ \varphi^*$ die Identitätsabbildung auf W^*. Somit ist $\varphi^* : W^* \to V^*$ ein Isomorphismus mit $(\varphi^*)^{-1} = (\varphi^{-1})^*$. □

Die Rechnung aus dem Absatz vor Definition 4.1.4 zeigt $(\psi + \theta)^* = \psi^* + \theta^*$. Analog erhält man $(\lambda \cdot \psi)^* = \lambda \cdot \psi^*$ und also:

Lemma 4.1.7. *Die Abbildung* $\mathrm{Hom}(V, W) \ni \varphi \mapsto \varphi^* \in \mathrm{Hom}(W^*, V^*)$ *ist linear.*

Wir untersuchen nun, inwieweit man V nach V^* einbetten kann. Zunächst betrachten wir eine Einbettung in den sogenannten Bidualraum $V^{**} := (V^*)^*$, der aus den linearen Abbildungen von V^* in den Grundkörper besteht.

Definition 4.1.8. Für jeden Vektorraum V definieren wir $\mathrm{ev}_V : V \to V^{**}$ (für ‚Evaluierung') durch $\mathrm{ev}_V(v)(\varphi) = \varphi(v)$.

Wieder in der Sprache der Kategorientheorie besagt das folgende Ergebnis, dass die Evaluierung eine natürliche Transformation darstellt.

Proposition 4.1.9. *(a) Die Abbildung ev_V ist für jeden Vektorraum V linear.*
(b) Für jede lineare Abbildung $\varphi : V \to W$ ist

$$\begin{array}{ccc} V & \xrightarrow{\mathrm{ev}_V} & V^{**} \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi^{**}=(\varphi^*)^*} \\ W & \xrightarrow{\mathrm{ev}_W} & W^{**} \end{array}$$

ein kommutatives Diagram, das heißt, es gilt $\varphi^{**} \circ \mathrm{ev}_V = \mathrm{ev}_W \circ \varphi$.

Beweis. (a) Wir rechnen

$$\begin{aligned} \mathrm{ev}_V(v+w)(\varphi) &= \varphi(v+w) = \varphi(v) + \varphi(w) \\ &= \mathrm{ev}_V(v)(\varphi) + \mathrm{ev}_V(w)(\varphi) = \big(\mathrm{ev}_V(v) + \mathrm{ev}_V(w)\big)(\varphi). \end{aligned}$$

Analog zeigt man $\mathrm{ev}_V(\lambda \cdot v) = \lambda \cdot \mathrm{ev}_V(v)$.
(b) Man rechnet zunächst

$$(\varphi^{**} \circ \mathrm{ev}_V)(v) = (\varphi^*)^*(\mathrm{ev}_V(v)) = \mathrm{ev}_V(v) \circ \varphi^*.$$

Für $\psi \in W^*$ ergibt sich nun

$$\begin{aligned} (\varphi^{**} \circ \mathrm{ev}_V)(v)(\psi) &= \mathrm{ev}_V(v)(\varphi^*(\psi)) = \mathrm{ev}_V(v)(\psi \circ \varphi) \\ &= \psi \circ \varphi(v) = \psi(\varphi(v)) = \mathrm{ev}_W(\varphi(v))(\psi) = (\mathrm{ev}_W \circ \varphi)(v)(\psi). \end{aligned}$$

Man hat also $(\varphi^{**} \circ \mathrm{ev}_V)(v) = (\mathrm{ev}_W \circ \varphi)(v)$ für beliebiges v, wie gewünscht. □

Wir würden nun gerne auch eine Einbettung von V nach V^* (statt nach V^{**}) konstruieren. Hier lässt sich zunächst feststellen, dass dies nicht auf (im Sinne der Kategorientheorie) natürliche Weise möglich ist:

Lemma 4.1.10. *Wir nehmen an, jedem Vektorraum V ist eine lineare Abbildung $e_V : V \to V^*$ so zugeordnet, dass das Diagramm*

$$\begin{array}{ccc} V & \xrightarrow{e_V} & V^* \\ {\scriptstyle\varphi}\downarrow & & \uparrow{\scriptstyle\varphi^*} \\ W & \xrightarrow{e_W} & W^* \end{array}$$

stets kommutiert, man also $\varphi^* \circ e_W \circ \varphi = e_V$ für jedes lineare $\varphi : V \to W$ hat. Dann muss e_V konstant Null sein.

Beweis. Betrachte $\varphi : V \to W$ mit $\varphi(v) = 0$ für alle $v \in V$. Mit der Linearität von e_W und φ^* folgt stets $e_V(v) = \varphi^* \circ e_W \circ \varphi(v) = \varphi^* \circ e_W(0) = 0$. □

Zumindest heuristisch ist eine Abbildung natürlich, wenn sie nicht von einer willkürlichen Wahl abhängt – etwa von der Wahl einer bestimmten Basis. Umgekehrt können wir hoffen, mittels einer Basis doch noch eine Einbettung von V nach V^* zu konstruieren.

Definition 4.1.11. Sei B eine Basis eines Vektorraums V. Für $b \in B$ definieren wir $\overline{b} \in V^*$ durch $\overline{b}(w) = \lambda$ für $w = \lambda \cdot b + w'$ mit $w' \in \operatorname{Span}(B \backslash \{b\})$. Ist $v \in V$ ein beliebiger Vektor, so setzen wir nun

$$\iota_B(v) = \sum_{i=1}^{n} \lambda_i \cdot \overline{b}_i \in V^* \quad \text{für } v = \sum_{i=1}^{n} \lambda_i \cdot b_i \text{ mit paarweise verschiedenen } b_i \in B.$$

Ist B die Standardbasis im Vektorraum K^n, so schreiben wir ι_n anstelle von ι_B.

Im endlichdimensionalen Fall heißt die Menge der Vektoren $\overline{b}$ wegen des folgenden Resultats auch die zu B duale Basis.

Proposition 4.1.12. *Ist V ein Vektorraum mit Basis B, so hat $B^* = \{\overline{b} : b \in B\}$ die folgenden Eigenschaften:*

(a) Die Menge $B^* \subseteq V^*$ ist linear unabhängig. Weiter ist $B \ni b \mapsto \overline{b} \in B^*$ eine injektive Funktion.

(b) Hat V endliche Dimension, so ist B^* eine Basis von V^*.

Beweis. (a) Für $b, c \in B$ ist

$$\overline{b}(c) = \begin{cases} 1 & \text{wenn } b = c, \\ 0 & \text{sonst.} \end{cases}$$

Dies liefert zunächst die Injektivität. Hat man weiter $\sum_{i=1}^{n} \lambda_i \cdot \overline{b}_i = 0$ mit paarweise verschiedenen b_i, so erhält man jeweils

$$0 = \left(\sum_{i=1}^{n} \lambda_i \cdot \overline{b}_i \right)(b_j) = \sum_{i=1}^{n} \lambda_i \cdot \overline{b}_i(b_j) = \lambda_j.$$

(b) Um zu zeigen, dass $B^* = \{\overline{b}_1, \dots, \overline{b}_n\}$ ein Erzeugendensystem ist, betrachten wir $\varphi \in V^*$ und setzen $\lambda_i^{\varphi} := \varphi(b_i)$. Wie in (a) gilt dann jeweils

$$\left(\sum_{i=1}^{n} \lambda_i^{\varphi} \cdot \overline{b}_i \right)(b_j) = \lambda_j^{\varphi} = \varphi(b_j),$$

sodass die Abbildungen φ und $\sum_{i=1}^{n} \lambda_i^{\varphi} \cdot \overline{b}_i$ übereinstimmen. □

Definiert man $f : B \to V^*$ durch $f(b) = \overline{b}$, so stimmt ι_B per Konstruktion mit der Abbildung $\operatorname{Abb}(f)$ aus Definition 2.1.11 überein. Mit Proposition 2.1.13 folgt:

Korollar 4.1.13. *Für jeden Vektorraum V mit Basis B gilt:*

(a) Die Abbildung $\iota_B : V \to V^*$ ist linear und injektiv.

(b) Hat V endliche Dimension, so ist ι_B ein Isomorphismus.

Die Beschränkung auf endliche Dimension ist im Allgemeinen nötig:

Bemerkung 4.1.14. Wir betrachten den $\mathbb{Q}$-Vektorraum $\bigoplus_{\mathbb{N}} \mathbb{Q}$ mit abzählbarer Dimension $\dim(\bigoplus_{\mathbb{N}} \mathbb{Q}) = |\mathbb{N}|$. Wir wählen eine Basis $B = \{b_i : i \in \mathbb{N}\}$. Jede Teilmenge $A \subseteq B$ induziert eine Funktion $f_A : B \to \{0, 1\} \subseteq \mathbb{Q}$ mit $f_A(b) = 1$ genau für $b \in B$. Es ergibt sich eine lineare Abbildung $\varphi_A = \operatorname{Abb}(f_A) : \bigoplus_{\mathbb{N}} \mathbb{Q} \to \mathbb{Q}$. Für $A \neq C$ hat man $\varphi_A \neq \varphi_C$, denn gilt etwa $b \in A$ aber $b \notin C$, so ergibt sich $\varphi_A(b) = f_A(b) = 1$ aber $\varphi_C(b) = f_C(b) = 0$. Mit Proposition 2.2.10 folgt

$$|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| = |\mathcal{P}(B)| \leq \left| (\textstyle\bigoplus_{\mathbb{N}} \mathbb{Q})^* \right|.$$

Da der Grundkörper $\mathbb{Q}$ abzählbar ist, folgt hieraus, dass $(\bigoplus_{\mathbb{N}} \mathbb{Q})^*$ überabzählbare Dimension hat (vgl. Bemerkung 2.3.12) und also nicht zu $\bigoplus_{\mathbb{N}} \mathbb{Q}$ isomorph ist.

Im Fall der Standardbasis erhalten wir eine bekannte Konstruktion:

Beispiel 4.1.15. Für die Einheitsvektoren im K^n erhält man

$$\overline{e}_i(w) = w_i \quad \text{für} \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \sum_{i=1}^{n} w_i \cdot e_i.$$

Die Funktion $\iota_n : K^n \to (K^n)^*$ erfüllt deshalb

$$\iota_n(v)(w) = \sum_{i=1}^{n} v_i \cdot \overline{e}_i(w) = \sum_{i=1}^{n} v_i \cdot w_i \text{ für } v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \text{ und } w = \begin{pmatrix} w_1 \\ \vdots \\ w_3 \end{pmatrix}.$$

Für $K = \mathbb{R}$ ist also $\iota_n(\cdot)(\cdot)$ das zumeist mit $\langle \cdot, \cdot \rangle$ bezeichnete Skalarprodukt. Für später vermerken wir die hieraus ersichtliche Symmetrie $\iota_n(v)(w) = \iota_n(w)(v)$ sowie die Darstellung $\iota_n(v)(w) = v^\mathsf{T} \cdot w$.

Durch zweifache Einbettung wird die Abhängigkeit von einer Basis aufgehoben:

Lemma 4.1.16. *Wir betrachten einen Vektorraum V mit einer Basis B und dualer Basis $C := B^* = \{\overline{b} : b \in B\} \subseteq V^*$. Das Diagramm*

$$\begin{array}{ccc} V & \xrightarrow{\quad \mathrm{ev}_V \quad} & V^{**} \\ & {\scriptstyle \iota_B} \searrow \quad \nearrow {\scriptstyle \iota_C} & \\ & V^* & \end{array}$$

kommutiert, das heißt, die Evaluierung faktorisiert als $\mathrm{ev}_V = \iota_C \circ \iota_B$.

Beweis. Für beliebige $b, c \in B$ rechnet man

$$(\iota_C \circ \iota_B)(b)(\overline{c}) = \iota_C(\overline{b})(\overline{c}) = \overline{\overline{b}}(\overline{c}) = \left\{ \begin{array}{ll} 1 & \text{wenn } b = c \\ 0 & \text{sonst} \end{array} \right\} = \overline{c}(b) = \mathrm{ev}_V(b)(\overline{c}).$$

Dies zeigt zunächst, dass $\mathrm{ev}_V(b)$ und $(\iota_C \circ \iota_B)(b)$ für jedes $b \in B$ als lineare Abbildungen von V^* in den Grundkörper übereinstimmen. Die Behauptung folgt. □

Für Vektorräume V und W endlicher Dimension $\dim(V) = m$ und $\dim(W) = n$ steht $\mathrm{Hom}(V^*, W^*)$ laut Korollar 4.1.6 und 4.1.13 in Bijektion zu $\mathrm{Hom}(K^m, K^n)$. Über diese Bijektion sind lineare Abbildungen zwischen Dualräumen durch Matrizen darstellbar – insbesondere die durch Definition 4.1.4 induzierten Abbildungen:

Definition 4.1.17. Ist $\varphi : K^n \to K^m$ eine lineare Abbildung, so ist die transponierte Abbildung $\varphi^\mathsf{T} : K^m \to K^n$ dadurch bestimmt, dass sie das Diagramm

$$\begin{array}{ccc} K^m & \xrightarrow{\iota_m} & (K^m)^* \\ {\scriptstyle \varphi^\mathsf{T}} \downarrow & & \downarrow {\scriptstyle \varphi^*} \\ K^n & \xrightarrow{\iota_n} & (K^n)^* \end{array}$$

kommutativ macht. Man hat also $\varphi^\mathsf{T} = (\iota^n)^{-1} \circ \varphi^* \circ \iota^m$. Für $M \in K^{m \times n}$ ist die transponierte Matrix $M^\mathsf{T} \in K^{n \times m}$ durch $\mathrm{Abb}(M^\mathsf{T}) = \mathrm{Abb}(M)^\mathsf{T}$ bestimmt.

Eine Ad-hoc-Definition der transponierten Matrix hatten wir bereits im Beweis von Proposition 3.2.12 gegeben. Wir zeigen nun, dass diese mit unserer neuen Definition übereinstimmt.

Lemma 4.1.18. *Für $M = (\mu_{ij}) \in K^{m\times n}$ gilt $M^\mathsf{T} = (\nu_{kl}) \in K^{n\times m}$ with $\nu_{kl} = \mu_{lk}$.*

Beweis. Es genügt, zu zeigen, dass die Matrix M^T aus dem Lemma die Gleichung

$$\iota_n \circ \mathrm{Abb}(M^\mathsf{T}) = \mathrm{Abb}(M)^* \circ \iota_m$$

erfüllt. Für $i = 1, \dots, m$ rechnet man einerseits

$$\left(\iota_n \circ \mathrm{Abb}(M^\mathsf{T})\right)(e_i) = \iota_n\left(M^T \cdot e_i\right) = \iota_n\left(\begin{pmatrix}\nu_{1i}\\ \vdots \\ \nu_{ni}\end{pmatrix}\right) = \iota_n\left(\sum_{j=1}^n \nu_{ji}\cdot e_j\right)$$
$$= \iota_n\left(\sum_{j=1}^n \mu_{ij}\cdot e_j\right) = \sum_{j=1}^n \mu_{ij}\cdot \overline{e}_j$$

sowie andererseits

$$\left(\mathrm{Abb}(M)^* \circ \iota_m\right)(e_i) = \mathrm{Abb}(M)^*(\iota_m(e_i)) = \mathrm{Abb}(M)^*(\overline{e}_i) = \overline{e}_i \circ \mathrm{Abb}(M).$$

Angewandt auf e_k für $k = 1, \dots, n$ ergibt sich

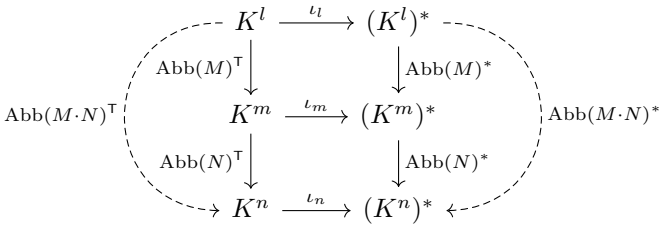
$$\left(\iota_n \circ \mathrm{Abb}(M^\mathsf{T})\right)(e_i)(e_k) = \sum_{j=1}^n \mu_{ij}\cdot\overline{e}_j(e_k) = \mu_{ik} = \overline{e}_i\left(\begin{pmatrix}\mu_{1k}\\ \vdots \\ \mu_{nk}\end{pmatrix}\right)$$
$$= \overline{e}_i(M\cdot e_k) = \overline{e}_i \circ \mathrm{Abb}(M)(e_k) = \left(\mathrm{Abb}(M)^* \circ \iota_m\right)(e_i)(e_k),$$

wie benötigt. □

Im Beweis von Proposition 3.2.12 hatten wir es zur Übung überlassen, Teil (b) des folgenden Lemmas durch Matrizenrechnung zu prüfen. Wir können nun auch einen abstrakten Beweis geben. Dieser mag zunächst schwieriger erscheinen, beruht aber auf der einfachen Idee, das Ergebnis auf die Gleichung $(\varphi \circ \theta)^* = \theta^* \circ \varphi^*$ zurückzuführen.

Proposition 4.1.19. *(a) Für jede Matrix M hat man $(M^\mathsf{T})^\mathsf{T} = M$.*
(b) Für alle $M \in K^{l\times m}$ und $N \in K^{m\times n}$ ist $(M\cdot N)^\mathsf{T} = N^\mathsf{T}\cdot M^\mathsf{T}$.
(c) Für $M, N \in K^{m\times n}$ und $\lambda \in K$ ist $(M+N)^\mathsf{T} = M^\mathsf{T} + N^\mathsf{T}$ und $(\lambda\cdot M)^\mathsf{T} = \lambda\cdot M^\mathsf{T}$.

Beweis. Teil (a) ist auf der Ebene von Matrizen unmittelbar (kann aber darauf zurückgeführt werden, dass φ und φ^{**} laut Proposition 4.1.9 bis auf natürliche Isomorphie gleich sind). Teil (b) kann als Reformulierung von Lemma 4.1.5(c) aufgefasst werden. Letzteres stellt sicher, dass das folgende Diagramm auch nach Hinzunahme der gestrichelten Pfeile kommutiert:



Arbeitet man mit solch einem Diagramm zum ersten Mal, so ist es sicher sinnvoll, es in die Gleichungskette

$$\begin{aligned}\mathrm{Abb}(M\cdot N)^*\circ\iota_l &= \big(\mathrm{Abb}(M)\circ\mathrm{Abb}(N)\big)^*\circ\iota_l = \mathrm{Abb}(N)^*\circ\mathrm{Abb}(M)^*\circ\iota_l\\ &= \mathrm{Abb}(N)^*\circ\iota_m\circ\mathrm{Abb}(M)^{\mathsf{T}} = \iota_n\circ\mathrm{Abb}(N)^{\mathsf{T}}\circ\mathrm{Abb}(M)^{\mathsf{T}}\end{aligned}$$

zu übersetzen. Gemäß Definition 4.1.17 folgt dann

$$\iota_n\circ\mathrm{Abb}(M\cdot N)^{\mathsf{T}} = \mathrm{Abb}(M\cdot N)^*\circ\iota_l = \iota_n\circ\mathrm{Abb}(N)^{\mathsf{T}}\circ\mathrm{Abb}(M)^{\mathsf{T}}$$

und also

$$\begin{aligned}\mathrm{Abb}\left((M\cdot N)^{\mathsf{T}}\right) &= \mathrm{Abb}(M\cdot N)^{\mathsf{T}} = \mathrm{Abb}(N)^{\mathsf{T}}\circ\mathrm{Abb}(M)^{\mathsf{T}}\\ &= \mathrm{Abb}\left(N^{\mathsf{T}}\right)\circ\mathrm{Abb}\left(M^{\mathsf{T}}\right) = \mathrm{Abb}\left(N^{\mathsf{T}}\cdot M^{\mathsf{T}}\right).\end{aligned}$$

Dies ergibt $(M\cdot N)^{\mathsf{T}} = N^{\mathsf{T}}\cdot M^{\mathsf{T}}$, wie gewünscht. Teil (c) lässt sich analog auf Lemma 4.1.7 zurückführen (oder für Matrizen nachrechnen). □

Wir können folgenden Schluss ziehen.

Korollar 4.1.20. *Eine Matrix $M\in K^{n\times n}$ ist genau dann invertierbar, wenn die transponierte Matrix M^{T} invertierbar ist. In diesem Fall hat man*

$$\left(M^{\mathsf{T}}\right)^{-1} = \left(M^{-1}\right)^{\mathsf{T}}.$$

Beweis. Die Äquivalenz haben wir bereits im Beweis von Proposition 3.2.12 abgeleitet. Wegen

$$M^{\mathsf{T}}\cdot(M^{-1})^{\mathsf{T}} = (M^{-1}\cdot M)^{\mathsf{T}} = \mathbb{1}^{\mathsf{T}} = \mathbb{1}$$

lautet die Inverse wie angegeben. Hierbei ist $\mathbb{1}^{\mathsf{T}} = \mathbb{1}$ auf der Ebene von Matrizen trivial, kann aber auch auf $\mathrm{id}^* = \mathrm{id}$ aus Lemma 4.1.5 zurückgeführt werden. □

Dualisierung erhellt auch die Eigenschaften der Elementarmatrizen:

Bemerkung 4.1.21. Laut Lemma 3.1.11 sind Zeilenoperationen durch Multiplikation mit Elementarmatrizen von links realisierbar. Im Beweis von Theorem 3.2.7 haben wir dann verwendet, dass Multiplikation von rechts entsprechende Spaltenoperationen liefert. Man kann dies für Multiplikation von links und rechts jeweils separat durch Matrizenrechnung verifizieren. Die beiden Fälle lassen sich aber auch per Dualität aufeinander zurückführen (sodass dann also nur noch ein Fall nachgerechnet werden muss). In Anbetracht von Definition 3.1.10 hat man etwa

$$S_{kl}^{\mathsf{T}} = (\mathbb{1}-E_{lk})^{\mathsf{T}} = \mathbb{1}^{\mathsf{T}} - E_{lk}^{\mathsf{T}} = \mathbb{1}-E_{kl} = S_{lk}.$$

Es ergibt sich

$$N\cdot S_{kl} = \big((N\cdot S_{kl})^{\mathsf{T}}\big)^{\mathsf{T}} = \big(S_{kl}^{\mathsf{T}}\cdot N^{\mathsf{T}}\big)^{\mathsf{T}} = \big(S_{lk}\cdot N^{\mathsf{T}}\big)^{\mathsf{T}}.$$

Hat man Lemma 3.1.11, so weiß man, dass die Matrix $S_{lk}\cdot N^{\mathsf{T}}$ entsteht, indem in N^{T} die l-te von der k-ten Zeile abgezogen wird. Bei der Transponierten sind aber die Rollen von Zeilen und Spalten vertauscht. Also entsteht $N\cdot S_{kl} = (S_{lk}\cdot N^{\mathsf{T}})^{\mathsf{T}}$ aus N, indem die l-te von der k-ten Spalte abgezogen wird.

Im Folgenden verwenden wir Dualisierung, um den bekannten Slogan ‚Zeilenrang gleich Spaltenrang' zu rechtfertigen. Wir führen zunächst den folgenden Begriff ein.

Definition 4.1.22. Der Annulator eines Untervektorraums $U\subseteq K^n$ ist

$$U^0 = \{v\in K^n : U\subseteq\ker(\iota_n(v))\}.$$

In den meisten Quellen wird der Annulator als $\{\varphi \in (K^n)^* : U \subseteq \ker(\varphi)\}$ definiert (und dann auch für beliebige Vektorräume anstelle des K^n). Wir haben den Isomorphismus $\iota_n : K^n \to (K^n)^*$ eingebaut, weil es praktische Vorteile hat, wenn U und U^0 im selben Raum enthalten sind. Dies ist analog zu den unterschiedlichen Vorzügen der Dualisierung $\varphi \mapsto \varphi^*$ und der Transposition $\varphi \mapsto \varphi^\mathsf{T}$. Während erstere natürlich ist (nämlich unabhängig von der Wahl einer Basis), hat letztere den Vorteil, dass etwa $(\varphi^\mathsf{T})^\mathsf{T} = \varphi$ wörtlich und nicht nur bis auf Isomorphie gilt. Freilich gewöhnt man sich in der Mathematik daran, bis auf implizite Isomorphie zu arbeiten. Dennoch stellt dies zu Beginn eine praktische Schwierigkeit dar.

Lemma 4.1.23. *Für jedes $U \subseteq K^n$ ist U^0 ein Untervektorraum des K^n.*

Beweis. Hat man $v_1, v_2 \in U^0$ und also jeweils $\iota_n(v_i)(u) = 0$ für alle $u \in U$, so ist

$$\iota_n(v_1 + v_2)(u) = \big(\iota_n(v_1) + \iota_n(v_2)\big)(u) = \iota_n(v_1)(u) + \iota_n(v_2)(u) = 0$$

für $u \in U$, sodass $v_1 + v_2 \in U^0$ gilt. Analog folgt $\lambda \cdot v \in U^0$ für $v \in U^0$. □

Wir zeigen, dass U und U^0 komplementäre Untervektorräume sind.

Proposition 4.1.24. *Für jeden Untervektorraum $U \subseteq K^n$ gilt*

$$\dim(U) + \dim(U^0) = n.$$

Beweis. Wähle eine Basis $u_1, \dots, u_m$ von U und ergänze sie durch $v_1, \dots, v_{n-m}$ zu einer Basis des K^n. Bezüglich der gesamten Basis bilden wir die linear unabhängigen Vektoren $\overline{v}_1, \dots, \overline{v}_{n-m}$ im Raum $(K^n)^*$. Es genügt zu zeigen, dass diese den Raum $U' := \{\varphi \in (K^n)^* : U \subseteq \ker(\varphi)\}$ aufspannen. Laut Definition 4.1.11 gilt jeweils $\overline{v}_i(u) = 0$ für alle $u \in U$, wodurch $\overline{v}_i \in U'$ gezeigt ist. Umgekehrt lässt sich ein beliebiges $\varphi \in U'$ darstellen als

$$\varphi = \sum_{i=1}^{n} \varphi(v_i) \cdot \overline{v}_i.$$

Die linearen Abbildungen auf den beiden Seiten dieser Gleichung stimmen nämlich auf den Basisvektoren u_j und v_k überein (mit Wert 0 beziehungsweise $\varphi(v_k)$). □

Zur Dualität gehört, dass die Konstruktion des Annulators selbstinvers ist.

Proposition 4.1.25. *Für jeden Untervektorraum $U \subseteq K^n$ gilt $U^{00} = U$.*

Beweis. Sei $u \in U$ beliebig. Um $u \in U^{00}$ zu zeigen, müssen wir $U^0 \subseteq \ker(\iota_n(u))$ nachweisen. Wir betrachten ein beliebiges $w \in U^0$, sodass also $U \subseteq \ker(\iota_n(w))$ gilt. Wegen der Symmetrie aus Beispiel 4.1.15 ergibt sich $\iota_n(u)(w) = \iota_n(w)(u) = 0$ und also $w \in \ker(\iota_n(u))$. Dies zeigt $U \subseteq U^{00}$. Wegen

$$\dim(U) + \dim(U^0) = n = \dim(U^0) + \dim(U^{00})$$

ist aber $\dim(U) = \dim(U^{00})$ und damit schon $U = U^{00}$. □

Ein wesentlicher Vorzug des Annulators ist, dass wir nun auch Bild und Kern als zueinander duale Konzepte verstehen können:

Proposition 4.1.26. *Für jede lineare Abbildung $\varphi : K^m \to K^n$ gilt*

$$\ker\big(\varphi^\mathsf{T}\big) = \operatorname{img}(\varphi)^0 \quad \textit{und} \quad \operatorname{img}\big(\varphi^\mathsf{T}\big) = \ker(\varphi)^0.$$

Beweis. Die erste Gleichung ergibt sich aus

$$\begin{aligned} v \in \ker(\varphi^\mathsf{T}) \Leftrightarrow \iota_n(v) \in \ker(\varphi^*) &\Leftrightarrow \varphi^*(\iota_n(v)) = \iota_n(v) \circ \varphi = 0 \\ &\Leftrightarrow \operatorname{img}(\varphi) \subseteq \ker(\iota_n(v)) \Leftrightarrow v \in \operatorname{img}(\varphi)^0. \end{aligned}$$

Die zweite Gleichung folgt per Dualität. Mit der ersten Gleichung hat man nämlich erst $\operatorname{img}(\varphi^\mathsf{T})^0 = \ker(\varphi^{\mathsf{TT}}) = \ker(\varphi)$ und dann $\operatorname{img}(\varphi^\mathsf{T}) = \operatorname{img}(\varphi^\mathsf{T})^{00} = \ker(\varphi)^0$. □

Der Slogan ‚Zeilenrang gleich Spaltenrang' lässt sich durch das folgende Ergebnis ausdrücken. Explizit wissen wir aus Bemerkung 2.4.15, dass der Rang von M^T mit der Kardinalität einer maximal linear unabhängigen Menge von Spalten von M^T übereinstimmt. Also ergibt sich der Rang von M auch als die Kardinalität einer maximal linear unabhängigen Menge von Zeilen der Matrix M. Insbesondere ist $M \in K^{n\times n}$ genau dann invertierbar, wenn die Zeilen linear unabhängig sind.

Korollar 4.1.27. *Für $M \in K^{m\times n}$ ist* $\operatorname{rank}(M) = \operatorname{rank}(M^\mathsf{T})$.

Beweis. Hat man $\varphi = \operatorname{Abb}(M)$ und also $\varphi^\mathsf{T} = \operatorname{Abb}(M)^\mathsf{T} = \operatorname{Abb}(M^\mathsf{T})$, so ist

$$\begin{aligned} \operatorname{rank}(M^\mathsf{T}) = \dim(\operatorname{img}(\varphi^\mathsf{T})) &= \dim(\ker(\varphi)^0) \\ &= n - \dim(\ker(\varphi)) = \dim(\operatorname{img}(\varphi)) = \operatorname{rank}(M), \end{aligned}$$

wobei die vorangegangenen Ergebnisse und der Rangsatz einfließen. □

Zum Abschluss dieses Unterkapitels besprechen wir eine Anwendung von Dualität im Kontext von Gleichungssystemen. In Abschnitt 3.1 haben wir eine lineare Abbildung $\varphi : K^m \to K^n$ und einen Vektor $w \in K^n$ als gegeben vorausgesetzt und dann nach einer Beschreibung der Lösungsmenge

$$\operatorname{Lsg}(\varphi, w) = \{v \in K^m : \varphi(v) = w\}$$

gefragt. Umgekehrt kann man einen affinen Teilraum $A \subseteq K^m$ als Lösungsmenge vorgeben und dann nach φ und w mit $\operatorname{Lsg}(\varphi, w) = A$ suchen. Sei $A = v_0 + U$ für einen Untervektorraum $U \subseteq K^m$. Mit den Lemmata 3.1.3 und 3.1.5 gilt

$$\operatorname{Lsg}(\varphi, w) = v_0 + U \quad \Leftrightarrow \quad \varphi(v_0) = w \text{ und } \ker(\varphi) = U.$$

Auf der rechten Seite ist w durch den gegebenen Punkt $v_0 \in A$ und die zu findende Lösung φ eindeutig bestimmt. Wir haben es also äquivalent mit dem Problem zu tun, für einen gegebenen Untervektorraum $U \subseteq K^m$ ein lineares $\varphi : K^m \to K^n$ mit $\ker(\varphi) = U$ zu finden. Diese Bedingung ist äquivalent zu $\operatorname{img}(\varphi^\mathsf{T}) = U^0$ wegen

$$\begin{aligned} \ker(\varphi) = U &\Rightarrow \operatorname{img}(\varphi^\mathsf{T}) = \operatorname{img}(\varphi^\mathsf{T})^{00} = \ker(\varphi^{\mathsf{TT}})^0 = \ker(\varphi)^0 = U^0 \\ &\Rightarrow \ker(\varphi) = \ker(\varphi)^{00} = U^{00} = U. \end{aligned}$$

Hat man eine Basis $u_1, \ldots, u_n$ von U^0 gegeben – wobei $n = \dim(U^0) = m - \dim(U)$ gilt –, so betrachte man die Matrix $N \in K^{m\times n}$ mit den Vektoren u_i als Spalten. Für $M = N^\mathsf{T}$ und $\varphi = \operatorname{Abb}(M)$ gilt $\varphi^\mathsf{T} = \operatorname{Abb}(N)$ und also $\operatorname{img}(\varphi^\mathsf{T}) = U^0$. Wie gezeigt, ist also die gewünschte Bedingung $\ker(\varphi) = U$ erfüllt. Es bleibt also, zu gegebenem U eine Basis von U^0 zu bestimmen. Sei dazu $w_1, \ldots, w_{m-n}$ eine Basis von U. Man hat

$$\begin{aligned} v \in U^0 &\Leftrightarrow U \subseteq \ker(\iota_m(v)) \\ &\Leftrightarrow \iota_m(v)(w_i) = \iota_m(w_i)(v) = w_i^\mathsf{T} \cdot v = 0 \text{ für } i = 1, \ldots, m-n. \end{aligned}$$

Für die Matrix $A \in K^{(m-n)\times m}$ mit Zeilen w_i^T ist also $v \in U^0$ zu $A \cdot v = 0 \in K^{m-n}$ äquivalent. Mit anderen Worten gilt $U^0 = \mathrm{Lsg}(A, 0)$. Wir haben die Aufgabe also darauf zurückgeführt, ein lineares Gleichungssystem zu lösen, was wir schon können. Die allgemeine Idee, ein neues Problem auf ein bereits bekanntes duales Problem zurückzuführen, ist in den verschiedensten Gebieten der Mathematik fruchtbar.

Beispiel 4.1.28. Im $\mathbb{R}^3$ betrachten wir die Ebene

$$E = \mathrm{Span}(w_1, w_2) \quad \text{mit} \quad w_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \text{ und } w_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}.$$

Wie oben sei A die Matrix mit Zeilen w_i^T, also

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix}.$$

Man überzeuge sich von

$$E^0 = \mathrm{Lsg}(A, 0) = \mathrm{Span}(u_1) \quad \text{mit} \quad u_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

Setzt man nun $\varphi = \mathrm{Abb}(M)$ mit $M = \begin{pmatrix} 1 & 2 & -1 \end{pmatrix} \in K^{1\times 3}$, so erhält man gemäß der obigen Herleitung $E = \ker(\varphi)$, also

$$E = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x + 2y - z = 0 \right\}.$$

4.2. **Äquivalenzrelationen und Faktorraum.** In diesem Unterkapitel besprechen wir insbesondere, wie man einen Untervektorraum $U \subseteq V$ durch Übergang zu einem Quotienten V/U herausteilen und dadurch lineare Abbildungen injektiv machen kann.

Definition 4.2.1. Eine binäre Relation $R \subseteq M^2$ auf einer Menge M heißt Äquivalenzrelation, wenn für alle $x, y, z \in M$ gilt:

(Reflexivität) $\qquad (x, x) \in R,$

(Symmetrie) $\qquad (x, y) \in R \quad \Rightarrow \quad (y, x) \in R,$

(Transitivität) $\qquad (x, y) \in R \text{ und } (y, z) \in R \quad \Rightarrow \quad (x, z) \in R.$

Im Fall von Äquivalenzrelationen schreibt man oft infix $x \sim y$ statt $(x, y) \in R$. Es folgt das wohl wichtigste Beispiel für die lineare Algebra.

Lemma 4.2.2. *Seit V ein Vektorraum mit Untervektorraum U. Durch*

$$v \sim w \quad :\Leftrightarrow \quad v - w \in U$$

ist eine Äquivalenzrelation auf V gegeben.

Beweis. Wegen $v - v = 0 \in U$ gilt stets $v \sim v$. Aus $v \sim w$ folgt $w - v = -(v - w) \in U$ und also $w \sim v$. Hat man $v \sim v'$ und $v' \sim v''$, so ergibt sich

$$v - v'' = (v - v') + (v' - v'') \in U$$

und damit $v \sim v''$. □

Die Idee der folgenden Definition ist, äquivalente Elemente zu identifizieren, also in ein Objekt zusammenzufassen.

Definition 4.2.3. Ist $\sim$ eine Äquivalenzrelation auf einer Menge M, so definieren wir die Äquivalenzklasse von $x \in M$ durch

$$[x]_\sim = \{y \in M : x \sim y\}.$$

Die Faktormenge (oder synonym der Quotient) bezüglich $\sim$ ist gegeben durch

$$M/\sim = \{[x] : x \in M\}.$$

Die Funktion $\pi_\sim : M \to M/\sim$ mit $\pi_\sim(x) = [x]$ heißt Quotientenabbildung (oder auch kanonische Abbildung).

Ist $\sim$ aus dem Kontext klar, so schreibt man oft $[\cdot]$ und π statt $[\cdot]_\sim$ und $\pi_\sim$.

Lemma 4.2.4. *Für jede Äquivalenzrelation hat man*

$$x \sim y \quad \Leftrightarrow \quad \pi(x) = [x] = [y] = \pi(y).$$

Beweis. Man nehme zunächst $[x] = [y]$ an. Wegen der Reflexivität gilt $y \in y$ und somit $y \in [y]$, also auch $y \in [x]$ und daher $x \sim y$. Nun nehmen wir $x \sim y$ an. Wegen der Symmetrie genügt es, $[x] \supseteq [y]$ zu zeigen. Sei dazu $z \in [y]$ beliebig. Wir haben dann $y \sim z$ und mit der Transitivität auch $x \sim z$, also $z \in [x]$. □

In unserem Beispiel aus der linearen Algebra haben wir die Äquivalenzklassen bereits unter einem anderen Namen kennengelernt:

Beispiel 4.2.5. Ist $\sim$ die Relation aus Lemma 4.2.2, so ist die Äquivalenzklasse $[v]$ gleich dem affinen Teilraum $v + U$. Hat man nämlich $w \in [v]$ und also $v \sim w$, so ergibt sich $u := v - w \in U$ und damit $w = v - u \in v + U$. Ist umgekehrt $w \in v + U$ und also $w = v + u'$ für ein $u' \in U$, so erhält man $v - w = -u' \in U$, was $v \sim w$ und damit $w \in [v]$ zeigt. Lemma 3.1.3 besagt unter anderem, dass $v + U = w + U$ äquivalent ist zu $w \in v + U$. Dies ergibt sich nun als Spezialfall von Lemma 4.2.4.

Unter einer Partition einer Menge M versteht man eine Menge $\mathcal{M} \subseteq \mathcal{P}(M)\backslash\{\emptyset\}$ mit $M = \bigcup \mathcal{M}$ und $N \cap N' = \emptyset$ für alle $N \neq N'$ aus $\mathcal{M}$. Intuitiver schreibt man oft $\mathcal{M} = \{M_i : i \in I\}$, sodass man M als disjunkte Vereinigung $M = \bigcup_{i \in I} M_i$ erhält.

Proposition 4.2.6. *Ist $\sim$ eine Äquivalenzrelation auf einer Menge M, so ist $M/\sim$ eine Partition von M.*

Beweis. Für jedes $x \in M$ ist wegen der Reflexivität $x \sim x$ und also $x \in [x] \neq \emptyset$. Dies zeigt $M = \bigcup M/\sim$. Wir nehmen nun $[x] \cap [y] \neq \emptyset$ an und leiten $[x] = [y]$ ab. Ist $z \in [x] \cap [y]$ und also $x \sim z$ sowie $y \sim z$, so erhalten wir mit der Symmetrie $z \sim y$ und mit der Transitivität $x \sim y$. Nun folgt $[x] = [y]$ mit dem vorigen Lemma. □

Tatsächlich lassen sich Äquivalenzrelationen und Partitionen als zwei Perspektiven auf dasselbe Phänomen auffassen. Wir skizzieren dies in der folgenden Bemerkung, wobei wir die Details zur Übung überlassen.

Bemerkung 4.2.7. Ist $\mathcal{M}$ eine Partition von M, so erhält man durch

$$x \sim_{\mathcal{M}} y \quad :\Leftrightarrow \quad \text{es gibt ein } N \in \mathcal{M} \text{ mit } x, y \in N$$

eine Äquivalenzrelation. Die Operationen $\sim \mapsto M/\sim$ und $\mathcal{M} \mapsto \sim_{\mathcal{M}}$ sind zueinander invers, sodass wir eine Bijektion zwischen den Äquivalenzrelationen und den Partitionen gefunden haben.

Aus der Definition erhält man unmittelbar:

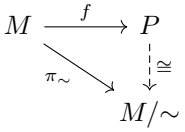
Korollar 4.2.8. *Die Quotientenabbildung $\pi : M \to M/\!\sim$ ist surjektiv.*

In der Tat lassen sich surjektive Abbildungen als dritte Perspektive auf ein gemeinsames Phänomen auffassen (neben Äquivalenzrelationen und Partitionen), wobei wir dies erneut nur skizzieren.

Bemerkung 4.2.9. Für eine Surjektion $f : M \to P$ ist durch

$$x \sim^f y \quad :\Leftrightarrow \quad f(x) = f(y)$$

eine Äquivalenzrelation auf M gegeben. Die Operationen $\sim \mapsto \pi_\sim$ und $f \mapsto \sim^f$ sind im Wesentlichen zueinander invers. Eine gegebene Äquivalenzrelation $\sim$ stimmt wegen Lemma 4.2.4 mit der Relation $\sim^f$ für $f := \pi_\sim$ überein. Ist umgekehrt eine Surjektion $f : M \to P$ gegeben, so kann man nicht erwarten, dass $\pi_\sim$ mit $\sim = \sim^f$ dieselbe Funktion wie f ist (weil die Wertebereiche im Allgemeinen verschieden sind). Ist aber die Funktion $g : P \to M$ rechtsinvers zu f, so ist $\pi_\sim \circ g$ eine Bijektion, welche den gestrichelten Pfeil im kommutativen Diagramm



bezeugt. In diesem Sinn lassen sich f und $\pi_\sim$ identifizieren. Formal mag man dafür zwei surjektive Funktionen $f_i : M \to P_i$ als äquivalent definieren, wenn es eine Bijektion $h : P_1 \to P_2$ mit $h \circ f_1 = f_2$ gibt.

Oben haben wir den Wertebereich mit P für ‚property' (‚Eigenschaft') bezeichnet. Man kann nämlich $f : M \to P$ so verstehen, dass jedem Element von M genau eine Eigenschaft aus P zugeordnet wird. Die Äquivalenzklassen in $M/\!\sim_f$ bestehen dann jeweils aus den Objekten, die unter eine der Eigenschaften fallen.

Beispiel 4.2.10. Für $n \in \mathbb{N}\backslash\{0\}$ ordne $r : \mathbb{Z} \to \{0, \dots, n-1\}$ den Rest bei Division durch n zu, sodass also jeweils $p = q \cdot n + r(p)$ mit eindeutig bestimmtem $q \in \mathbb{Z}$ gilt. Für die zugehörige Äquivalenzrelation $\sim^r$ (vgl. die vorige Bemerkung) schreiben wir auch $\sim^n$. Man hat dann

$$\begin{aligned} p \sim^n p' &\Leftrightarrow p \text{ und } p' \text{ haben denselben Rest bei Division durch } n \\ &\Leftrightarrow p - p' \text{ ist durch } n \text{ teilbar.} \end{aligned}$$

Da $p - r(p)$ stets durch n teilbar ist, lassen sich die Äquivalenzklassen schreiben als

$$[p] = [r(p)] = \{q \cdot n + r(p) : q \in \mathbb{Z}\} = r(p) + n\mathbb{Z}.$$

Für $0 \leq k < l < n$ hat man $k \not\sim^n l$. Damit bilden die sogenannten Restklassen $k + n\mathbb{Z}$ mit $0 \leq k < n$ eine Partition der ganzen Zahlen. Für die Faktormenge schreibt man oft $\mathbb{Z}_n = \mathbb{Z}/(n)$. Die Quotientenabbildung

$$\pi : \mathbb{Z} \to \mathbb{Z}_n \quad \text{mit} \quad \pi(p) = r(p) + n\mathbb{Z}$$

stimmt im Wesentlichen (bis auf die Bijektion $\{0, \dots, n-1\} \ni i \mapsto i + n\mathbb{Z}$) mit der Funktion r überein. Entsprechend kann man $\mathbb{Z}_n$ als abstrakte Repräsentation der Menge $\{0, \dots, n-1\}$ der Reste auffassen. Ein Vorteil des abstrakten Zugangs ist, dass $\mathbb{Z}_n$ auf kanonische Weise die Ringstruktur von $\mathbb{Z}$ erbt, wie wir später zeigen.

Wir konzentrieren uns nun auf den wichtigen Fall eines Quotienten durch einen Untervektorraum. Insbesondere wollen wir zeigen, dass die Faktormenge selbst wieder als Vektorraum aufgefasst werden kann, den man als Faktorraum bezeichnet. Die folgende Definition bedarf einer Rechtfertigung, welche wir unten nachliefern.

Definition 4.2.11. Im Fall der Äquivalenzrelation aus Lemma 4.2.2 schreiben wir meist V/U anstelle von $V/\sim$ (und mit Beispiel 4.2.5 manchmal $v+U$ statt $[v]$). Auf der Menge V/U definieren wir eine Addition und eine Skalarmultiplikation durch

$$[v] + [w] = [v + w] \quad \text{und} \quad \lambda \cdot [v] = [\lambda \cdot v].$$

Als Rechtfertigung der Definition ist hier die Wohldefiniertheit nachzuweisen. Für die Addition ist nämlich $a + b$ mit $a, b \in V/U$ zu erklären, ohne dass die Summanden schon in der Form $a = [v]$ und $b = [w]$ gegeben wären. Der Punkt ist, dass diese zwar in der angegebenen Form geschrieben werden können, dass v und w dabei aber nicht eindeutig bestimmt sind. Um diese Schwierigkeit zu umgehen, können wir aus jeder Äquivalenzklasse ein Element auswählen und für unsere Definition verwenden (wobei auch eine andere Rechtfertigung ohne Auswahlaxiom möglich ist). Es bleibt dann aber nachzuweisen, dass die Gleichungen aus der Definition nicht nur für die gewählten ‚Repräsentaten' sondern für alle $v, w \in V$ gelten. Dies leistet der folgende Beweis.

Lemma 4.2.12. *Addition und Skalarmultiplikation auf V/U sind wohldefiniert.*

Beweis. In Bezug auf die Addition zeigen wir, dass aus $[v] = [v']$ und $[w] = [w']$ schon $[v + w] = [v' + w']$ folgt. Wir nehmen also an, dass $v - v'$ und $w - w'$ in U liegen. Es ergibt sich dann auch

$$(v + w) - (v' + w') = (v - v') + (w - w') \in U.$$

Die Skalarmultiplikation behandelt man analog. □

Die Vektorraumstruktur vererbt sich unmittelbar, wie man prüfen möge.

Proposition 4.2.13. *Ist U ein Untervektorraum eines Vektorraums V, so ist V/U wieder ein Vektorraum.*

Die Vektorräume V und V/U lassen sich wie folgt in Beziehung setzen.

Proposition 4.2.14. *Die Quotientenabbildung $\pi : V \to V/U$ ist linear und surjektiv mit* $\ker(\pi) = U$.

Beweis. Die Gleichungen aus Definition 4.2.11 lassen sich auch als

$$\pi(v) + \pi(w) = \pi(v + w) \quad \text{und} \quad \lambda \cdot \pi(v) = \pi(\lambda \cdot v)$$

schreiben. Weiter hat jedes Element von V/U die Form $[v] = \pi(v)$ mit $v \in V$. Schließlich gilt $\pi(v) = [v] = 0$ genau für $v \sim 0$ und also für $v = v - 0 \in U$. □

Insbesondere zeigt das vorige Ergebnis, dass π genau für $U = \{0\}$ injektiv ist.

Proposition 4.2.15. *Man hat* $\dim(U) + \dim(V/U) = \dim(V)$.

Beweis. Wir wählen eine Basis B von U und ergänzen sie zu einer Basis C von V. Angenommen, es gilt $\sum_{i=1}^{n} \lambda_i \cdot \pi(v_i) = 0$ für paarweise verschiedene $v_i \in C \backslash B$. Man hat dann $\pi(\sum_{i=1}^{n} \lambda_i \cdot v_i) = 0$ in V/U und also $\sum_{i=1}^{n} \lambda_i \cdot v_i \in U$. Wegen der Eindeutigkeit der Basisdarstellung folgt $\lambda_1 = \ldots = \lambda_n = 0$. Dies zeigt, dass die Einschränkung von π auf $C \backslash B$ injektiv ist und linear unabhängiges Bild hat.

Um zu zeigen, dass letzteres auch ein Erzeugendensystem ist, betrachten wir ein beliebiges Element $[v] = \pi(v) \in V/U$ und schreiben

$$v = \sum_{i=1}^{m} \lambda_i \cdot v_i + \sum_{j=1}^{n} \mu_j \cdot w_j \quad \text{mit } v_i \in B \text{ und } w_j \in C \backslash B.$$

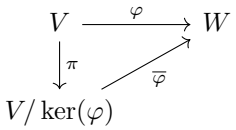
Wegen $B \subseteq U$ gilt jeweils $\pi(v_i) = 0$ und also $[v] = \sum_{j=1}^{n} \mu_j \cdot \pi(w_j)$. □

Für ein intuitives Verständnis der Dimension mag das Folgende hilfreich sein.

Beispiel 4.2.16. Wir nehmen Beispiel 4.2.5 auf und betrachten den Fall eines Untervektorraums $U \subseteq V = \mathbb{R}^2$ mit $\dim(U) = 1$, sodass also U eine Gerade durch den Ursprung ist. Hier besteht V/U aus den zu U parallelen Geraden $v + U$. Für festes $v_0 \notin U$ ist die lineare Abbildung $p : \mathbb{R} \to V/U$ mit $p(\lambda) = \lambda \cdot v_0 + U$ nicht konstant Null und hat also ein Bild positiver Dimension. Da die vorige Proposition $\dim(V/U) = 1$ liefert, ist p surjektiv und damit bijektiv. Intuitiv ist p eine Parametrisierung unserer Parallelenschar. Dass V/U Dimension Eins hat, entspricht der Tatsache, dass p von einer Variablen λ abhängt.

Um eine gegebene Abbildung $\varphi : V \to W$ surjektiv zu machen, genügt es, den Wertebereich auf das Bild $\operatorname{img}(\varphi)$ einzuschränken. Als wichtige Anwendung des Faktorraums besprechen wir nun, wie man eine Abbildung injektiv machen kann.

Theorem 4.2.17 (Homomorphiesatz)**.** *Für jede lineare Abbildung $\varphi : V \to W$ gibt es eine eindeutig bestimmte Funktion $\overline{\varphi} : V/\ker(\varphi) \to W$ mit der Eigenschaft, dass*



kommutiert, sodass man also $\varphi = \overline{\varphi} \circ \pi$ hat. Die Funktion $\overline{\varphi}$ ist linear und injektiv. Insbesondere gilt

$$V/\ker(\varphi) \cong \operatorname{img}(\varphi).$$

Beweis. Wir betrachten zunächst allgemeiner den Quotienten V/U durch einen beliebigen Untervektorraum U, der in $\ker(\varphi)$ enthalten ist. Da π surjektiv ist, kann es höchstens eine Funktion $\overline{\varphi} : V/U \to W$ mit $\varphi = \overline{\varphi} \circ \pi$ geben. Für die Existenz würden wir gerne

$$\overline{\varphi} : V/U \to W \quad \text{mit} \quad \overline{\varphi}([v]) = \varphi(v)$$

betrachten. Der entscheidende Punkt ist hier erneut die Wohldefiniertheit. Um letztere nachzuweisen, nehmen wir $[v] = [w]$ an. Es gilt dann $v - w \in U \subseteq \ker(\varphi)$ und also $\varphi(v - w) = 0$. Man hat daher $\varphi(v) = \varphi(w)$, wie gewünscht. Die Gleichung $\varphi = \overline{\varphi} \circ \pi$ gilt per Konstruktion. Für die Linearität rechnet man etwa

$$\overline{\varphi}([v] + [w]) = \overline{\varphi}([v + w]) = \varphi(v + w) = \varphi(v) + \varphi(w) = \overline{\varphi}([v]) + \overline{\varphi}([w]).$$

Per Konstruktion haben φ und $\overline{\varphi}$ dasselbe Bild. Ist U sogar gleich dem Kern von φ, so gilt $\overline{\varphi}([v]) = \varphi(v) = 0$ genau für $v \in U$. Es ergibt sich

$$\ker(\overline{\varphi}) = \{[v] : v \in U\} = \{0\},$$

sodass $\overline{\varphi}$ injektiv ist und also $V/\ker(\varphi) = \operatorname{img}(\overline{\varphi}) = \operatorname{img}(\varphi)$ gilt. □

Als erste Anwendung erhalten wir einen zweiten und besonders transparenten Beweis eines bereits bekannten Resultats:

Bemerkung 4.2.18. Es sei $\varphi : V \to W$ eine lineare Abbildung. Mit dem obigen Theorem ergibt sich

$$\operatorname{rank}(\varphi) = \dim(\operatorname{img}(\varphi)) = \dim(V/\ker(\varphi)).$$

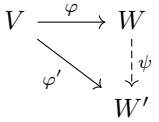
Zusammen mit Proposition 4.2.15 erhält man die Gleichung

$$\dim(\ker(\varphi)) + \operatorname{rank}(\varphi) = \dim(V),$$

welche wir bereits zuvor als Rangsatz bewiesen hatten.

Der folgende Beweis stellt eine weitere Anwendung dar. Konkret wird hier der Faktorraum verwendet, um das Ergebnis nicht mehr für allgemeines sondern nur noch für injektives φ zeigen zu müssen. Dieser Spezialfall ist einfacher, obwohl sich der allgemeine Fall auch direkt beweisen lässt.

Lemma 4.2.19. *Wir betrachten lineare Abbildungen $\varphi : V \to W$ und $\varphi' : V \to W'$. Es gilt genau dann $\ker(\varphi) \subseteq \ker(\varphi')$, wenn es eine lineare Abbildung $\psi : W \to W'$ gibt, für die das Diagramm*

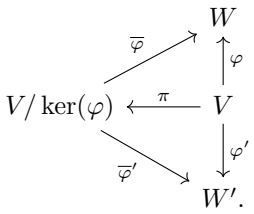


kommutiert, das heißt, für die $\psi \circ \varphi = \varphi'$ gilt.

Beweis. Hat man ein ψ wie im Diagramm, so folgt aus $v \in \ker(\varphi)$ in der Tat

$$\varphi'(v) = \psi \circ \varphi(v) = \psi(0) = 0$$

und also $v \in \ker(\varphi')$. Wir nehmen nun umgekehrt $\ker(\varphi) \subseteq \ker(\varphi')$ an. Aufgrund von Theorem 4.2.17 und seinem Beweis haben wir ein kommutatives Diagramm



Hierbei ist $\overline{\varphi}$ injektiv und hat also eine Linksinverse $\rho : W \to V/\ker(\varphi)$. Man kann annehmen, dass diese linear ist (indem man $\overline{\varphi} = \operatorname{Abb}(f)$ schreibt und $\rho = \operatorname{Abb}(g)$ mit $g \circ f = \mathrm{id}$ setzt). Wir definieren nun $\psi : W \to W'$ durch $\psi = \overline{\varphi}' \circ \rho$. Man erhält

$$\psi \circ \overline{\varphi} = \overline{\varphi}' \circ \rho \circ \overline{\varphi} = \overline{\varphi}'.$$

Mit dem obigen Diagramm ergibt sich

$$\psi \circ \varphi = \psi \circ \overline{\varphi} \circ \pi = \overline{\varphi}' \circ \pi = \varphi',$$

wie gewünscht. □

Um zu sehen, wofür das vorige Lemma gut ist, kann man einen direkten Beweis der Gleichung $\operatorname{img}(\varphi^{\mathsf{T}}) = \ker(\varphi)^0$ aus Proposition 4.1.26 geben (welche zuvor per Dualität gezeigt wurde).

Im Rest dieses Unterkapitels beschäftigen wir uns mit Quotienten von Ringen. Diese werden wir im nächsten Unterkapitel für eine transparente Konstruktion der komplexen Zahlen verwenden.

Definition 4.2.20. Es sei R ein kommutativer Ring. Eine nicht-leere Menge $I \subseteq R$ heißt Ideal, wenn für $x, y \in I$ und beliebiges $z \in R$ jeweils $x + y \in I$ und $-x \in I$ sowie $x \cdot z \in I$ gilt.[90] Ist I ein Ideal, so schreiben wir R/I für den Quotienten $R/\sim$ durch die Äquivalenzrelation $\sim$ mit

$$x \sim y \quad :\Leftrightarrow \quad x - y \in I.$$

Auf R/I definieren wir $[x] + [y] = [x + y]$ und $[x \cdot y] = [x] \cdot [y]$.

Man möge prüfen, dass es sich bei der Relation $\sim$ aus der Definition tatsächlich um eine Äquivalenzrelation handelt. Hierbei kommt die additive Abgeschlossenheit von I zum Tragen, welche auch die Wohldefiniertheit der Addition auf R/I garantiert. Für die Wohldefiniertheit der Multiplikation beobachten wir

$$[x] = [x'] \;\Rightarrow\; x - x' \in I \;\Rightarrow\; x \cdot y - x' \cdot y = (x - x') \cdot y \in I \;\Rightarrow\; [x \cdot y] = [x' \cdot y].$$

Hierbei beachte man, dass y nicht unbedingt in R liegt. Dies erklärt, warum in der Definition eines Ideals $x \cdot z \in I$ für $x \in I$ und allgemeines $z \in R$ gefordert wurde (statt nur für $z \in I$). Analog folgt $[x \cdot y] = [x \cdot y']$ aus $[y] = [y']$. Hat man letzteres und $[x] = [x']$, so ergibt sich also $[x \cdot y] = [x \cdot y'] = [x' \cdot y']$.

Lemma 4.2.21. *Für jeden kommutativen Ring R und jedes Ideal $I \subseteq R$ ist der Quotient R/I auch wieder ein kommutativer Ring. Ist $1 \in R$ multiplikativ neutral, so gilt dies auch für $[1] \in R/I$.*

Die Quotientenabbildung $\pi : R \to R/I$ ist wieder surjektiv und respektiert per Konstruktion die Addition und Multiplikation (ist also ein Ringhomomorphismus). Wir interessieren uns besonders für den Fall, dass R/I ein Körper ist.

Proposition 4.2.22. *Sei R ein kommutativer Ring mit Eins. Für jedes Ideal $I \subsetneq R$ sind äquivalent:*

(i) Der Faktorring R/I ist ein Körper.
(ii) Das Ideal I ist maximal, das heißt, es gibt kein Ideal $J \subsetneq R$ mit $I \subsetneq J$.

Beweis. Es gelte zunächst (i). Für ein gegebenes Ideal $J \supsetneq I$ wählen wir ein Element $x \in J \backslash I$. Wegen $[x] \neq 0$ in R/I finden wir ein $y \in R$ mit $[x \cdot y] = [x] \cdot [y] = 1$ und also $x \cdot y - 1 \in I$. Es ist dann $1 = x \cdot y - (x \cdot y - 1) \in J$ und damit $z = 1 \cdot z \in J$ für alle $z \in R$, was $J = R$ zeigt.

Wir nehmen nun (ii) an. Betrachte ein beliebiges Element $[x] \neq 0$ von R/I, sodass also $x \notin I$ gilt. Man kann sich überzeugen, dass

$$J := \{s + x \cdot t : s \in I \text{ und } t \in R\}$$

ein Ideal ist. Wegen $x \in J$ hat man $I \subsetneq J$ und mit (ii) deshalb $J = R$, also $1 \in J$. Schreibe $1 = s + x \cdot t$ mit $s \in I$. Es ergibt sich $x \cdot t - 1 \in I$ und also $[x] \cdot [t] = 1$. □

Mit dem folgenden Beispiel wird die Konstruktion des Körpers $\mathbb{B} = \mathbb{Z}_2$ aus Kapitel 1 in einen größeren Kontext gestellt.

[90] Die ersten beiden dieser drei Eigenschaften machen I zu einer Untergruppe der additiven Gruppe von R. Man kann Ideale auch in nicht-kommutativen Ringen definieren, wobei man dann entweder auch $z \cdot x \in I$ für $x \in I$ und $z \in R$ fordert (‚beidseitiges Ideal') oder auf die Symmetrie verzichtet und von Links- beziehungsweise Rechtsidealen spricht.

Beispiel 4.2.23. Für einen kommutativen Ring R und $x \in R$ ist

$$(x) := \{x \cdot z : z \in R\}$$

ein Ideal, wie man prüfen möge. Ideale der Form (x) heißen Hauptideale. In Fortführung von Beispiel 4.2.10 hat man

$$p \sim^n p' \quad \Leftrightarrow \quad p - p' \text{ ist durch } n \text{ teilbar} \quad \Leftrightarrow \quad p - p' \in (n) \subseteq \mathbb{Z}.$$

Dies rechtfertigt die bereits zuvor verwendete Schreibweise $\mathbb{Z}/(n)$ für den Quotientenring, der auch mit $\mathbb{Z}_n$ bezeichnet wird. Wie versprochen erhalten wir nun eine Ringstruktur auf $\mathbb{Z}_n$. Diese erlaubt uns das Rechnen mit Resten bei der Division durch n. Beispielsweise hat 7 bei der Division durch 3 den Rest 1, sodass man $[7] = [1]$ in $\mathbb{Z}_3$ hat (wofür man auch $7 \equiv 1 \mod 3$ schreibt). Es ergibt sich

$$\left[7^{10}\right] = [7]^{10} = [1]^{10} = \left[1^{10}\right] = [1],$$

sodass auch 7^{10} bei der Division durch 3 wieder den Rest 1 hat. Wir haben nun auch eine systematische Erklärung für die Gleichung $1 + 1 = 0$ in $\mathbb{B} = \mathbb{Z}_2$. Mit dem vorigen Ergebnis und dem sogenannten Lemma von Bézout[91] folgt, dass $\mathbb{Z}_n$ für $n > 0$ genau dann ein Körper ist, wenn n prim ist.

4.3. **Intermezzo: Polynome und komplexe Zahlen.** In einem Körper K hat eine lineare Gleichung $x \cdot s + t = 0$ mit $s, t \in K$ und $s \neq 0$ stets eine Lösung, die durch $x = -t/s$ gegeben ist. Für quadratische Gleichungen und Polynome höheren Grades ist dies im Allgemeinen nicht der Fall: So hat etwa $x^2 - 2 = 0$ in $\mathbb{Q}$ keine Lösung, weil $\sqrt{2}$ irrational ist. Weiter hat $x^2 + 1 = 0$ keine Lösung in $\mathbb{R}$, weil Quadratzahlen dort nicht negativ sind. Wir überzeugen uns zunächst informell, dass es nützlich wäre, die gegebene Gleichung lösen zu können:

Beispiel 4.3.1. Wir werden $\mathbb{R}$ später in einen größeren Körper $\mathbb{C}$ einbetten, dessen Elemente man komplexe Zahlen nennt. Dabei gibt es ein $i \in \mathbb{C}$ mit $i^2 = -1$. Um dessen Nutzen zu illustrieren, betrachten wir (im Zusammenhang mit den sogenannten Cardanischen Formeln) die Gleichung

$$p(x) = x^3 - 15x - 4 \stackrel{!}{=} 0.$$

Wir substituieren $x = s + t$ und rechnen

$$p(s + t) = s^3 + t^3 + 3 \cdot (s \cdot t - 5) \cdot (s + t) - 4.$$

Die Gleichung $p(x) = 0$ reduziert sich damit auf die Gleichungen

$$s \cdot t - 5 = 0 \quad \text{und} \quad s^3 + t^3 - 4 = 0.$$

Aus der ersten Gleichung erhalten wir $t = 5/s$. Einsetzen in die zweite Gleichung führt auf die Bedingung

$$w^2 - 4w + 5^3 \stackrel{!}{=} 0 \quad \text{mit} \quad w = s^3.$$

Rechnen wir – an dieser Stelle noch informell – mit $i = \sqrt{-1}$, so liefert die bekannte Lösungsformel für quadratische Gleichungen

$$w = \frac{4 \pm \sqrt{4^2 - 4 \cdot 5^3}}{2} = 2 \pm \sqrt{4 - 5^3} = 2 \pm \sqrt{(-1) \cdot 121} = 2 \pm 11i.$$

Für $s = 2 + i$ rechnen wir

$$s^3 = (4 + 4i + i^2) \cdot (2 + i) = (3 + 4i)(2 + i) = 6 + 3i + 8i + 4i^2 = 2 + 11i = w,$$

[91] Wir werden später eine Version dieses Lemmas für Polynome beweisen.

was der obigen Wahl von w entspricht. Weiter erhält man nun

$$t = \frac{5}{s} = \frac{5}{2+i} = \frac{5 \cdot (2-i)}{(2+i) \cdot (2-i)} = \frac{5 \cdot (2-i)}{4-i^2} = 2-i.$$

Schließlich ergibt sich

$$x = s + t = (2+i) + (2-i) = 4.$$

Man prüft, dass in der Tat $p(4) = 0$ gilt.

Das vorige Beispiel deutet zumindest informell an, dass die Verwendung von komplexen Zahlen große Vorteile bringen kann – selbst, wenn man letztlich nur an reellen Lösungen interessiert sein sollte. Wir werden im nächsten Kapitel sehen, dass die komplexen Zahlen auch in der linearen Algebra von entscheidender Bedeutung sind. Im vorliegenden Unterkapitel geben wir den komplexen Zahlen eine präzise mathematische Bedeutung (mit der dann auch die Rechnungen im vorigen Beispiel einen offiziellen Status gewinnen).

Wir beschäftigen uns zunächst mit Polynomen. Dabei beginnen wir mit einer formalen Definition, die wir dann im Nachgang durch eine intuitivere (und üblichere) Schreibweise ergänzen.

Definition 4.3.2. Für einen Körper K heißt $K[x] := \bigoplus_{\mathbb{N}} K$ die Menge der Polynome über K.[92] Die Addition auf $K[x]$ ist wie in Definition 2.3.10 gegeben. Um die Multiplikation von $p, q \in K[x]$ zu erklären, definieren wir $p \cdot q : \mathbb{N} \to K$ durch

$$(p \cdot q)(n) = \sum_{i=0}^{n} p(i) \cdot q(n-i).$$

Ein Polynom ist für uns also formal eine Funktion $p : \mathbb{N} \to K$, die nur für endlich viele Argumente $n(1) < \ldots < n(l)$ Werte $p(n(i)) \neq 0$ annimmt. Intuitiver schreibt man solch ein p in der Form

$$p(n(l)) \cdot x^{n(l)} + \ldots + p(n(1)) \cdot x^{n(1)},$$

wobei die $p(n(i)) \in K$ Koeffizienten genannt werden. Indem man auch die Null als Koeffizient zulässt, erhält man für jedes Polynom ungleich Null eine Darstellung

$$\sum_{i=0}^{d} \lambda_i \cdot x^i,$$

welche wir eindeutig machen können, indem wir $\lambda_d \neq 0$ fordern. Die Schreibweise ist dabei nicht willkürlich, weil das angegebene Polynom tatsächlich durch Addition und Multiplikation im Sinn der obigen Definition aus den konstanten Polynomen $\lambda_i = \lambda_i \cdot x^0$ und dem Polynom $x = 1 \cdot x^1$ hervorgeht. Die Multiplikation von Polynomen lässt sich ebenfalls durch die intuitive Schreibweise transparent machen, wofür wir auf den Beweis des nächsten Lemmas verweisen.

Definition 4.3.3. Der Grad eines Polynoms $p = \sum_{i=0}^{d} \lambda_i \cdot x^i \in K[X]$ mit $\lambda_d \neq 0$ ist $\deg(p) = d$. Weiter setzen wir $\deg(0) = -\infty$.

[92]Man kann Polynome genauso über Ringen definieren. Wir verzichten auf die allgemeinere Darstellung, um wörtlich auf Definition 2.3.10 aufbauen zu können. Auch gelten nicht alle der folgenden Ergebnisse über beliebigen Ringen.

Der folgende Beweis zeigt insbesondere, dass für $p, q \in K[x]$ wieder $p \cdot q \in K[x]$ gilt, dass die Funktion $p \cdot q$ also auch nur endlich viele Werte ungleich Null annimmt. Für beliebiges $d \in \{-\infty\} \cup \mathbb{N}$ vereinbaren wir dabei $-\infty + d = -\infty = d + (-\infty)$.

Lemma 4.3.4. *Es gilt* $\deg(p \cdot q) = \deg(p) + \deg(q)$.

Beweis. Ist p oder q Null, so ist das Ergebnis unmittelbar. Im verbliebenen Fall schreiben wir

$$p = \sum_{i=0}^{c} \lambda_i \cdot x^i \quad \text{und} \quad q = \sum_{j=0}^{d} \mu_j \cdot x^j$$

mit $\lambda_c, \mu_d \neq 0$. Es gilt dann

$$p \cdot q = \sum_{n=0}^{c+d} \sum_{i=0}^{c} \lambda_i \cdot \mu_{n-i} \cdot x^n,$$

wobei wir $\mu_j = 0$ für $j > d$ setzen. Insbesondere ist damit

$$\sum_{i=0}^{c} \lambda_i \cdot \mu_{c+d-i} = \lambda_c \cdot \mu_d \neq 0$$

und also $\deg(p \cdot q) = c + d$. □

Die Multiplikation auf $K[x]$ kann als Erweiterung der Skalarmultiplikation aus Definition 2.3.10 aufgefasst werden (über die Einbettung $\lambda \mapsto \lambda \cdot x^0$). Wir wissen bereits, dass $K[x]$ mit dieser Skalarmultiplikation und der Addition ein K-Vektorraum ist. In Erweiterung dieser Tatsache prüfe man:

Proposition 4.3.5. *Die Menge $K[x]$ der Polynome bildet einen Ring.*

Man beachte, dass $K[x]$ nie ein Körper ist. Das Polynom x hat nämlich wegen

$$\deg(x \cdot q) = 1 + \deg(q) > 0 = \deg(1)$$

kein Inverses. Wie ganze Zahlen kann man auch Polynome mit Rest dividieren:

Proposition 4.3.6. *Für $p, q \in K[x]$ mit $q \neq 0$ gibt es eindeutige $r, s \in K[x]$ mit*

$$p = s \cdot q + r \quad \textit{und} \quad \deg(r) < \deg(q).$$

Beweis. Für die Eindeutigkeit nehmen wir an, dass für $i \in \{0, 1\}$ je $p = s_i \cdot q + r_i$ mit $\deg(r_i) < \deg(q)$ gilt. Man hat dann $r_0 - r_1 = q \cdot (s_1 - s_0)$. Dabei ist

$$\deg(q) > \deg(r_0 - r_1) = \deg(q) + \deg(s_1 - s_0),$$

was nur für $s_0 = s_1$ und also $r_0 = r_1$ möglich ist.

Für die Existenz wählen wir s, r mit $p = s \cdot q + r$ so, dass $\deg(r)$ kleinstmöglich ist (wobei zumindest $s = 0$ und $r = p$ eine mögliche Wahl darstellt). Für einen Widerspruch nehmen wir an, dass $m := \deg(q) \leq \deg(r) =: n$ gilt. Wir schreiben dann $q = q' + \lambda \cdot x^m$ und $r = r' + \mu \cdot x^n$ mit $\deg(q') < m$ und $\deg(r') < n$. Für das Polynom $s' := s + \mu/\lambda \cdot x^{n-m}$ hat man

$$\begin{aligned} p - s' \cdot q &= p - s \cdot q - \frac{\mu}{\lambda} \cdot x^{n-m} \cdot q = r - \frac{\mu}{\lambda} \cdot x^{n-m} \cdot (q' + \lambda \cdot x^m) \\ &= r - \frac{\mu}{\lambda} \cdot x^{n-m} \cdot q' - \mu \cdot x^n = r' - \frac{\mu}{\lambda} \cdot x^{n-m} \cdot q' =: r''. \end{aligned}$$

Hierbei ist $\deg(x^{n-m} \cdot q') < n$ und also $\deg(r'') < n$. Andererseits ist $p = s' \cdot q + r''$, was der Minimalität von $\deg(r)$ widerspricht. □

Der vorige Beweis enthält einen Algorithmus, den wir nun explizit machen.

Bemerkung 4.3.7. Um p mit Rest durch $q \neq 0$ zu dividieren, setzen wir zunächst $s_0 = 0$ und $r_0 = p$. Induktiv erhalten wir so Darstellungen $p = s_i \cdot q + r_i$. Solange dabei $n = \deg(r_i) \geq \deg(q) = m$ gilt, setzen wir

$$s_{i+1} = s_i + \frac{\mu}{\lambda} \cdot x^{n-m} \quad \text{und} \quad r_{i+1} = r_i' - \frac{\mu}{\lambda} \cdot x^{n-m} \cdot q'$$

für $q = q' + \lambda \cdot x^m$ mit $\deg(q') < m$ und $r_i = r_i' + \mu \cdot x^n$ mit $\deg(r_i') < n$. Für konkrete Berechnungen bietet es sich an, die Polynome s_i und r_i in Form einer Tabelle aufzuschreiben, wobei man zusätzlich auch noch $\mu = \mu_i$ und $n = n_i$ festhalten kann. Für $p = x^4 + 4x - 2$ und $q = x^2 + x + 2$ hat man etwa $\lambda = 1$ und $m = 2$ sowie $q' = x + 2$, was zu folgender Tabelle führt:

s_i	r_i	μ_i	n_i
0	$x^4 + 4x - 2$	1	4
x^2	$-x^3 - 2x^2 + 4x - 2$	-1	3
$x^2 - x$	$-x^2 + 6x - 2$	-1	2
$x^2 - x - 1$	$7x$		

Man kann prüfen, dass in der Tat $p = (x^2 - x - 1) \cdot q + 7x$ gilt.

Die intuitive Schreibweise legt bereits nahe, dass sich Polynome als Funktionen auf dem Grundkörper auffassen lassen:

Definition 4.3.8. Für $p = \sum_{i=0}^{d} \lambda_i \cdot x^i \in K[x]$ sei $E[p] : K \to K$ gegeben durch

$$E[p](\zeta) = \sum_{i=0}^{d} \lambda_i \cdot \zeta^i.$$

Die Abbildung $p \mapsto E[p]$ ist mit der Addition und Multiplikation verträglich (also ein Ringhomomorphismus), aber für endliches K nicht injektiv. Letzteres folgt zum Beispiel, weil hier $K[x]$ unendlich ist, während die Menge der Funktionen auf K endlich ist. Es ist aber auch instruktiv, sich als Übung ein konkretes Gegenbeispiel zu überlegen. Für unendliches K ist $p \mapsto E[p]$ tatsächlich injektiv, was man bei Interesse als anspruchsvolle Übung zeigen möge (wobei man für den Nachweis der linearen Unabhängigkeit von $\{E[x^n] : n \in \mathbb{N}\}$ die Vandermonde-Matrizen nachschlagen kann). Jedenfalls ist es grundsätzlich wichtig, dass man zwischen einem Polynom p und der zugehörigen Polynomfunktion $E[p]$ unterscheidet. Wenn man sich den Unterschied ausreichend bewusst gemacht hat, schreibt man in der Praxis oft p anstelle von $E[p]$. Speziell in unserem Formalismus ist dann allerdings zu beachten, dass man es nun mit einer Funktion $p : K \to K$ zu tun hat, die nicht mit der repräsentierenden Funktion $p : \mathbb{N} \to K$ aus Definition 4.3.2 übereinstimmt.

Von großer Bedeutung ist der Zusammenhang von Nullstellen und Teilbarkeit. Als Nullstelle von $p \in K[x]$ bezeichnen wir ein $\lambda \in K$ mit $p(\lambda) = E[p](\lambda) = 0$. Für $p, q \in K[x]$ sagt man, q teile p, wenn es ein $s \in K[x]$ mit $p = s \cdot q$ gibt.

Proposition 4.3.9. *Für $p \in K[x]$ und $\lambda \in K$ hat man*

$$\lambda \textit{ ist Nullselle von } p \quad \Leftrightarrow \quad x - \lambda \textit{ teilt } p.$$

Beweis. In der Rückrichtung schreiben wir $p = (x - \lambda) \cdot s$. Da $q \mapsto E[q]$ ein Ringhomomorphismus ist, ergibt sich

$$E[p](\lambda) = E[x - \lambda](\lambda) \cdot E[s](\lambda) = (\lambda - \lambda) \cdot E[s](\lambda) = 0.$$

Für die andere Richtung schreiben wir $p = s \cdot (x - \lambda) + r$ für einen Rest $r \in K[x]$ mit $\deg(r) < 1$. Dabei ist $r = \mu$ konstant und man erhält

$$0 = p(\lambda) = s(\lambda) \cdot (\lambda - \lambda) + \mu = \mu,$$

also $p = s \cdot q$. □

Das folgende Ergebnis impliziert insbesondere, dass ein nicht-konstantes Polynom nur endlich viele Nullstellen haben kann.

Proposition 4.3.10. *Sind $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Nullstellen eines Polynoms $p \in K[x]$ mit $p \neq 0$, so hat man eine Darstellung*

$$p = q \cdot \prod_{i=1}^{k} (x - \lambda_i)^{n(i)}$$

für ein Polynom $q \in K[x]$ ohne Nullstellen, wobei die Exponenten $n(i)$ und das Polynom q eindeutig bestimmt sind.

Beweis. Für die Existenz argumentieren wir per Induktion über $k \in \mathbb{N}$. Hat das Polynom p keine Nullstelle, so setzt man schlicht $q = p$. Anderenfalls wählen wir eine Nullstelle λ_1 und schreiben $p = (x - \lambda_1)^{n(1)} \cdot p'$ für den größtmöglichen Exponenten $n(1)$. Hierbei hat p' nur noch Nullstellen $\lambda_2, \ldots, \lambda_k$, sodass wir induktiv schließen können.

Die Eindeutigkeit von $n(i)$ ergibt sich wie folgt: Es sei $(x - \lambda_i)^m \cdot s = (x - \lambda_i)^n \cdot t$ mit $m \leq n$, wobei λ_i Nullstelle weder von s noch von t ist. Durch Kürzen im Integritätsring $K[x]$ erhält man $s = (x - \lambda_i)^{n-m} \cdot t$, wobei dann $n = m$ gelten muss. Die Eindeutigkeit von q folgt nun erneut durch Kürzen. □

Wir können nun den folgenden Begriff einführen.

Definition 4.3.11. In der Darstellung aus Proposition 4.3.10 heißt $n(i)$ die Vielfachheit der Nullstelle λ_i.

Für Körper mit der folgenden Eigenschaft ist die Darstellung aus der vorigen Proposition besonders einfach, weil q konstant sein muss. Für $p \neq 0$ ergibt sich wegen Lemma 4.3.4 dann $\deg(p) = \sum_{i=1}^{k} n(i)$.

Definition 4.3.12. Ein Körper K heißt algebraisch abgeschlossen, wenn jedes nicht-konstante Polynom aus $K[x]$ eine Nullstelle in K hat.

Hat ein Polynom in einem gegebenen Körper keine Nullstelle, kann man nach Nullstellen in einer Erweiterung suchen. So hat etwa $x^3 - 2$ keine Nullstelle in $\mathbb{Q}$ aber schon in $\mathbb{R}$. Nun ist $\mathbb{R}$ nicht algebraisch abgeschlossen (da $x^2 + 1$ keine reelle Nullstelle besitzt), hat gegenüber $\mathbb{Q}$ aber den folgenden Vorzug:

Proposition 4.3.13. *Ist $\deg(p) > 0$ ungerade, so hat $p \in \mathbb{R}[x]$ eine Nullstelle in $\mathbb{R}$.*

[93] Der Autor prüft diesen Beweis im Rahmen der linearen Algebra nicht ab, da er (freilich elementare) Analysis enthält.

Beweis[93]. Für $p = \sum_{i=0}^{d} \lambda_i \cdot x^i$ mit $\lambda_d \neq 0$ setzen wir $M := \max\{|\lambda_i| : i \leq n\}$. Es ergibt sich

$$\left|\sum_{i=0}^{d-1} \lambda_i \cdot x^i\right| \leq (d-1) \cdot M \cdot |x|^{d-1} \quad \text{für} \quad |x| \geq 1.$$

Setzt man $N = \max(d \cdot M/\lambda_d, 1)$, so erhält man

$$p(-N) \leq -\lambda_d \cdot N^d + (d-1) \cdot M \cdot |x|^{d-1} = N^{d-1} \cdot \big((d-1) \cdot M - \lambda_d \cdot N\big) < 0$$

und analog $p(N) > 0$. Der Zwischenwertsatz liefert eine reelle Nullstelle. □

Im Folgenden wollen wir $\mathbb{R}$ in den algebraisch abgeschlossenen Körper $\mathbb{C}$ der komplexen Zahlen einbetten (wobei wir die algebraische Abgeschlossenheit erst im nächsten Kapitel nachweisen). Die entscheidende Idee ist, dass man Nullstellen geeigneter Polynome durch Quotientenbildung hinzufügen kann.

Definition 4.3.14. Ein nicht-konstantes Polynom $p \in K[x]$ heißt irreduzibel, wenn es keine nicht-konstanten $s, t \in K[x]$ mit $p = s \cdot t$ gibt.

Als Teil des folgenden Beweises zeigen wir das Lemma von Bézout für Polynomringe. Dieses hatten wir für $\mathbb{Z}$ bereits in Beispiel 4.2.23 erwähnt, wo auch die Definition des Hauptideals (f) zu finden ist.

Theorem 4.3.15. *Ist $p \in K[x]$ irreduzibel, so ist $K[x]/(p)$ ein Körper.*

Beweis. Laut Lemma 4.2.21 haben wir einen kommutativen Ring mit Eins. Um mit Proposition 4.2.22 zu schließen, zeigen wir, dass (p) ein maximales Ideal ist. Zunächst ist (p) von $K[x]$ verschieden, da man sonst $1 \in (p)$ hätte, sodass p ein Teiler von 1 wäre, was für nicht-konstantes p unmöglich ist. Für ein beliebiges Ideal $J \supsetneq (p)$ zeigen wir nun $J = K[x]$. Betrachte dazu $q \in J \backslash (p)$ und wähle $t, t' \in K[x]$ so, dass $s := t \cdot p + t' \cdot q$ ungleich Null ist und dabei kleinstmöglichen Grad hat. Teilen mit Rest liefert

$$p = t'' \cdot s + r \quad \text{mit} \quad \deg(r) < \deg(s).$$

Mit $r = (1 - t'' \cdot t) \cdot p + (-t'' \cdot t') \cdot q$ und wegen der Minimalität von $\deg(s)$ gilt $r = 0$. Somit ist s Teiler von p und analog auch von q. Da p irreduzibel ist, muss s oder t'' konstant und also invertierbar sein. Wegen $q \notin (p)$ kann dies nicht für t'' gelten (da man mit $q = s' \cdot s$ sonst $q = p \cdot s'/t''$ hätte). Man erhält also

$$1 = \frac{t}{s} \cdot p + \frac{t'}{s} \cdot q \in J$$

und damit $J = K[x]$, wie gewünscht. □

Um $K[x]/(p)$ als Erweiterung von K aufzufassen, betrachten wir die Elemente von K zunächst als konstante Polynome in $K[x]$. Explizit ist dies (wie bereits angedeutet) vermittelt durch die Funktion

$$\iota : K \to K[x] \quad \text{mit} \quad \iota(\lambda) = \lambda \cdot x^0,$$

die sich mit der Addition und Multiplikation verträgt (also ein Ringhomomorphismus ist). Oft schreibt man schlicht λ anstelle von $\iota(\lambda)$. Unsere Einbettung ergibt sich durch Komposition mit der Quotientenabbildung $\pi : K[x] \to K[x]/(p)$.

Lemma 4.3.16. *Der Ringhomomorphismus $\pi \circ \iota : K \to K[x]/(p)$ ist injektiv.*

Beweis. Aus $\pi \circ \iota(\lambda) = \pi \circ \iota(\mu)$ folgt, dass das konstante Polynom $\iota(\lambda) - \iota(\mu)$ in (p) liegt. Mit $(p) \neq K[x]$ muss dann aber $\iota(\lambda) = \iota(\mu)$ und also $\lambda = \mu$ gelten. □

Wir hatten angedeutet, dass p in der Erweiterung $K' := K[x]/(p)$ eine Nullstelle haben soll. Genauer müssen wir dazu p als Polynom in $K'[x]$ auffassen, was über die Einbettung aus dem vorigen Lemma möglich ist (wobei wir $\pi(q) = [q]$ schreiben).

Definition 4.3.17. Wir schreiben $K' = K[x]/(p)$ und definieren

$$\hat{q} = \sum_{i=0}^{d} [\iota(\lambda_i)] \cdot x^i \in K'[x] \quad \text{für} \quad q = \sum_{i=0}^{d} \lambda_i \cdot x^i \in K[x].$$

Identifiziert man K mit seinem Bild unter $\pi \circ \iota$, so stimmen $\hat{q}$ und q überein. Das folgende Ergebnis ist zwar äußerst wichtig, wird durch den Beweis aber als direkte Folgerung aus der Konstruktion ausgewiesen.

Theorem 4.3.18. *Ist $p \in K[x]$ ein irreduzibles Polynom über einem Körper K, so hat $\hat{p}$ in $K[x]/(p)$ eine Nullstelle.*

Beweis. Die gewünschte Nullstelle ist als das Bild $\pi(x) = [x]$ von $x \in K[x]$ unter der Quotientenabbildung gegeben. Um dies zu sehen, schreiben wir $p = \sum_{i=0}^{d} \lambda_i \cdot x^i$. In Anbetracht von $\iota(\lambda_i) \cdot x^i = \lambda_i \cdot x^0 \cdot x^i = \lambda_i \cdot x^i$ ergibt sich

$$\hat{p}([x]) = \sum_{i=0}^{d} [\iota(\lambda_i)] \cdot [x]^i = \left[\sum_{i=0}^{d} \iota(\lambda_i) \cdot x^i \right] = [p] = 0,$$

wie gewünscht. □

Angesichts von Beispiel 4.3.13 und Proposition 4.3.13 erscheint es besonders wichtig, die reellen Zahlen durch eine Quadratwurzel von -1 zu erweitern. Unser allgemeiner Ansatz ist wegen des folgenden Ergebnisses anwendbar.

Lemma 4.3.19. *Das Polynom $x^2 + 1$ ist über $\mathbb{R}$ irreduzibel.*

Beweis. Wäre die Aussage falsch, so hätte $x^2 + 1$ aufgrund seines Grades einen Teiler der Form $x - \lambda$. Nach Proposition 4.3.9 wäre dann $\lambda \in \mathbb{R}$ eine Nullstelle von $x^2 + 1$, was aber $\lambda^2 + 1 > 0$ widerspricht. □

Im Folgenden sind die Körpereigenschaften durch Theorem 4.3.15 garantiert.

Definition 4.3.20. Der Körper der komplexen Zahlen ist durch $\mathbb{C} := \mathbb{R}[x]/(x^2+1)$ definiert. Wir schreiben i (‚imaginäre Einheit') für das Element $[x] \in \mathbb{C}$ (also für den Wert der Quotientenabbildung auf dem Polynom $x \in \mathbb{R}[x]$).

Wie im allgemeinen Fall oben ist $\mathbb{R} \ni \lambda \mapsto [\lambda \cdot x^0] \in \mathbb{C}$ ein injektiver Ringhomomorphismus. Wir schreiben meist λ anstelle von $[\lambda \cdot x^0]$ und betrachten also $\mathbb{R}$ als Teilkörper von $\mathbb{C}$. Das folgende Ergebnis zeigt, dass unsere Konstruktion mit der üblichen Ad-hoc-Definition von $\mathbb{C}$ übereinstimmt.

Proposition 4.3.21. *(a) Jede komplexe Zahl lässt sich eindeutig als $\lambda + \mu \cdot i$ für reelle Zahlen λ und μ schreiben.*

(b) Es gilt

$$(\lambda + \mu \cdot i) + (\lambda' + \mu' \cdot i) = (\lambda + \lambda') + (\mu + \mu') \cdot i,$$
$$(\lambda + \mu \cdot i) \cdot (\lambda' + \mu' \cdot i) = (\lambda \cdot \lambda' - \mu \cdot \mu') + (\lambda \cdot \mu' + \lambda' \cdot \mu) \cdot i.$$

Aus (b) erhält man insbesondere $i^2 = (0 + 1 \cdot i)^2 = -1$.

Beweis. (a) Für die Eindeutigkeit nehmen wir $\lambda + \mu \cdot i = \lambda' + \mu' \cdot i$ an. Es gilt dann $[\mu \cdot x + \lambda] = [\mu' \cdot x + \lambda']$ in $\mathbb{R}[x]/(x^2 + 1)$. Somit liegt $p := (\mu - \mu') \cdot x + (\lambda - \lambda')$ im Ideal $(x^2 + 1)$ und wird also von $x^2 + 1$ geteilt. Wegen $\deg(p) \leq 1$ folgt daraus $p = 0$ und also $\lambda = \lambda'$ sowie $\mu = \mu'$.

Für die Existenz argumentieren wir per Induktion über $\deg(p)$, um zu zeigen, dass es $\lambda, \mu \in \mathbb{R}$ mit $[p] = \lambda + \mu \cdot i$ gibt. Hat man $\deg(p) \leq 1$, so ist dies unmittelbar. Im verbliebenen Fall schreiben wir

$$p = \sum_{i=0}^{d} \lambda_i \cdot x^i \quad \text{mit} \quad d = \deg(p) \geq 2.$$

Für $p' := p - \lambda_d \cdot x^{d-2} \cdot (x^2 + 1)$ gilt dann $[p] = [p']$ sowie $\deg(p') < \deg(p)$, sodass wir induktiv schließen können.

(b) Die Gleichung für die Addition gilt schlicht, weil $\mathbb{C}$ ein Körper ist. Die Gleichung für die Multiplikation erhält man wegen $[x^2 + 1] = 0$ und also

$$i^2 = [x]^2 = [x^2] = [x^2] - [x^2 + 1] = [x^2 - x^2 - 1] = -1$$

durch Ausmultiplizieren. □

Als Einschränkung der Multiplikation auf $\mathbb{C}$ erhält man eine Skalarmultiplikation mit Elementen von $\mathbb{R}$, sodass sich $\mathbb{C}$ als $\mathbb{R}$-Vektorraum auffassen lässt. Aus Teil (a) der vorigen Proposition folgt, dass die Elemente $1, i \in \mathbb{C}$ eine Basis bilden.

Korollar 4.3.22. *Als $\mathbb{R}$-Vektorraum hat $\mathbb{C}$ die Dimension zwei.*

Man nennt $\mathrm{Re}(\lambda + \mu \cdot i) = \lambda$ und $\mathrm{Im}(\lambda + \mu \cdot i) = \mu$ auch den Real- und Imaginärteil von $\lambda + \mu \cdot i \in \mathbb{C}$. Mit dem Korollar hat man einen Isomorphismus $\mathbb{C} \cong \mathbb{R}^2$, der sich explizit durch

$$\mathbb{C} \ni \lambda + \mu \cdot i \mapsto (\lambda, \mu) \in \mathbb{R}^2$$

angeben lässt. Dies führt zur Interpretation von $\mathbb{C}$ als sogenannte Gaußsche Zahlenebene. Die Addition entspricht dabei der üblichen Vektoraddition, während für die Multiplikation die Längen der Vektoren multipliziert und ihre Winkel zur positiven reellen Achse addiert werden (wie man sich anhand von $(2 + i) \cdot (1 + 2i) = 5i$ verdeutlichen möge). Es ergibt sich ein tiefer Zusammenhang mit der komplexen Exponentialfunktion, die wir hier aber nicht untersuchen (wobei wir darauf hinweisen, dass sich eine Summe im Exponenten in ein Produkt übersetzt).

Zum Abschluss dieses Unterkapitels behandeln wir noch einige elementare Operationen auf den komplexen Zahlen.

Bemerkung 4.3.23. Um durch eine komplexe Zahl $\lambda + \mu \cdot i \neq 0$ zu teilen, erweitern wir zu einer binomischen Formel und rechnen

$$\frac{1}{\lambda + \mu \cdot i} = \frac{\lambda - \mu \cdot i}{(\lambda + \mu \cdot i) \cdot (\lambda - \mu \cdot i)} = \frac{\lambda}{\lambda^2 + \mu^2} - \frac{\mu}{\lambda^2 + \mu^2} \cdot i.$$

Die vorige Rechnung erhellt auch das Folgende.

Definition 4.3.24. Die konjugierte Zahl zu $z = \lambda + \mu \cdot i \in \mathbb{C}$ ist definiert durch

$$\overline{z} = \lambda - \mu \cdot i.$$

Die Norm von z ist durch $|z| = \sqrt{z \cdot \overline{z}}$ gegeben.

Explizit hat man also $|\lambda+\mu\cdot i| = \sqrt{\lambda^2+\mu^2} \in \mathbb{R}$, woraus ersichtlich wird, dass wir es mit der üblichen Länge eines Vektors zu tun haben. Die folgenden elementaren Eigenschaften überlassen wir zur Übung.

Lemma 4.3.25. *Für $z, w \in \mathbb{C}$ hat man*

$$\overline{z+w} = \overline{z} + \overline{w} \quad \textit{und} \quad \overline{z\cdot w} = \overline{z}\cdot\overline{w} \quad \textit{und} \quad |z\cdot w| = |z|\cdot|w|.$$

Das folgende Resultat zeigt, dass unsere Konstruktion tatsächlich unter den allgemeinen Begriff der Norm fällt.

Proposition 4.3.26. *Für $z, w \in \mathbb{C}$ und $\lambda \in \mathbb{R}$ gilt*

(Positive Definitheit) $\quad |z| \geq 0 \quad \textit{und} \quad |z| = 0 \Leftrightarrow z = 0,$

(Absolute Homogenität) $\quad |\lambda \cdot z| = |\lambda| \cdot |z|,$

(Dreiecksungleichung) $\quad |z + w| \leq |z| + |w|.$

Hierbei verweist $|\lambda|$ auf den Betrag als reelle Zahl, welcher aber mit der Norm als komplexe Zahl übereinstimmt.

Beweis. Nur die Dreiecksungleichung erfordert unsere Aufmerksamkeit. Man überzeugt sich zunächst von $\mathrm{Re}(z) \leq |z|$ und folgert

$$z\cdot\overline{w} + \overline{z}\cdot w = z\cdot\overline{w} + \overline{z\cdot\overline{w}} = 2\cdot\mathrm{Re}(z\cdot\overline{w}) \leq 2\cdot|z\cdot\overline{w}| = 2\cdot|z|\cdot|\overline{w}| = 2\cdot|z|\cdot|w|.$$

Es ergibt sich

$$\begin{aligned}|z+w|^2 &= (z+w)\cdot\overline{z+w} = (z+w)\cdot(\overline{z}+\overline{w})\\ &= z\cdot\overline{z} + z\cdot\overline{w} + \overline{z}\cdot w + w\cdot\overline{w} \leq |z|^2 + 2\cdot|z|\cdot|w| + |w|^2 = (|z|+|w|)^2.\end{aligned}$$

Das Ergebnis folgt durch Wurzelziehen. □

Das wichtigste Ergebnis an dieser Stelle ist sicher der sogenannte Fundamentalsatz der Algebra, welcher besagt, dass $\mathbb{C}$ algebraisch abgeschlossen ist. Dies werden in Abschnitt 5.4 beweisen, wobei wir hier noch das folgende Ergebnis festhalten.

Lemma 4.3.27. *Für jedes $z \in \mathbb{C}$ gibt es ein $w \in \mathbb{C}$ mit $w^2 = z$.*

Beweis. Wir schreiben $z = x + i\cdot y$ und setzen

$$w = \sqrt{\frac{|z|+x}{2}} + i\cdot\sqrt{\frac{|z|-x}{2}}.$$

Mit $(|z|+x)\cdot(|z|-x) = |z|^2 - x^2 = y^2$ rechnet man $w^2 = z$. □

Zur Übung möge man die Formel aus dem vorigen Beweis herleiten, indem man die Gleichung $(a+i\cdot b)^2 = x + i\cdot y$ löst (wobei man zwei quadratische Gleichungen für Real- und Imaginärteil erhält). Einen noch transparenteren Beweis erhält man über die sogenannte Polardarstellung, die in der Analysis besprochen wird.

4.4. **Summen und Produkte.** In Definition 2.4.7 haben wir Summen von Untervektorräumen erklärt, Wir haben dabei versprochen, dass die Summe $U \oplus W$ im Fall $U \cap W = \emptyset$ nicht vom umgebenden Raum abhängt. Insbesondere dieses Versprechen wollen wir im vorliegenden Teilkapitel einlösen.

Um einen systematischen Ansatz zu entwickeln, beschäftigen wir uns zunächst mit Summen und Produkten von Mengen. Bei letzteren denkt man wohl zuerst an das kartesische Produkt mit

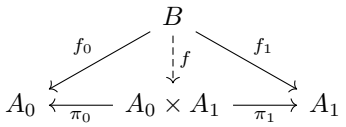
$$A_0 \times A_1 = \{(a_0, a_1) : a_i \in A_i\}.$$

Die sogenannten Projektionen sind gegeben durch

$$\pi_i : A_0 \times A_1 \to A_i \quad \text{mit} \quad \pi_i\big((a_0, a_1)\big) = a_i.$$

Für unseren systematischen Zugang ist entscheidend, dass sich das Produkt durch eine sogenannte universelle Eigenschaft charakterisieren lässt:

Proposition 4.4.1. *Für beliebige Funktionen $f_i : B \to A_i$ gibt es genau eine Funktion $f : B \to A_0 \times A_1$, die das Diagramm*



kommutativ macht, das heißt, für die jeweils $f_i = \pi_i \circ f$ gilt.

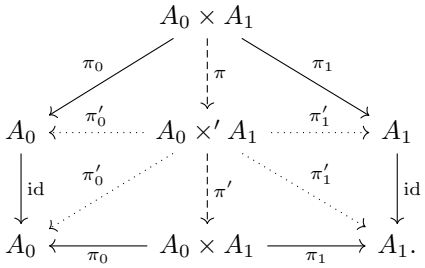
Beweis. Man sieht, dass

$$f(b) = \big(f_0(b), f_1(b)\big)$$

die einzig mögliche Wahl ist. □

Wir überzeugen uns, dass die universelle Eigenschaft tatsächlich eine im Wesentlichen eindeutige Charakterisierung liefert:

Bemerkung 4.4.2. Wir nehmen an, dass ein alternatives Produkt $A_0 \times' A_1$ mit Projektionen π_i' ebenfalls die universelle Eigenschaft erfüllt. Man erhält dann ein kommutatives Diagramm



Wenn wir die universelle Eigenschaft auf die durchgezogenen Pfeile anwenden, sagt uns die Eindeutigkeit, dass $\pi' \circ \pi = \mathrm{id}$ gelten muss. Genauso sieht man $\pi \circ \pi' = \mathrm{id}$. Dann ist also π eine Bijektion mit inverser Abbildung π'. Da $A_0 \times A_1$ und $A_0 \times' A_1$ als Mengen ohne weitere Struktur vorliegen, sind sie unter der Bijektion im Wesentlichen gleich. Genauer ist die einzig relevante Struktur durch die Projektionen π_i und π_i' gegeben, welche laut dem obigen Diagramm mit der Bijektion verträglich sind. Man beachte insbesondere, dass die Funktionen π_i per Konstruktion surjektiv sind. Mittels unserer Bijektion überträgt sich diese Eigenschaft auf die Funktionen π_i'. Kurz gesagt ist das Produkt durch die universelle Eigenschaft im Wesentlichen eindeutig festgelegt. Gleiches gilt für alle anderen universellen Eigenschaften, wie man in jedem Einzelfall prüfen oder im Rahmen der Kategorientheorie allgemein etablieren kann (was aber den Rahmen dieses Kurses sprengen würde).

Ein vernünftiger Kandidat für die Summe von zwei Mengen A_0 und A_1 ist ihre disjunkte Vereinigung[94]

$$A_0 \sqcup A_1 := \{(i, a) : i \in \{0, 1\} \text{ und } a \in A_i\}.$$

[94]Man schreibt manchmal auch $A_0 \dot\cup A_1$ anstelle von $A_0 \sqcup A_1$.

Gilt $A_0 \cap A_1 = \emptyset$, so steht $A_0 \sqcup A_1$ in Bijektion zu $A_0 \cup A_1$. Den allgemeinen Fall kann man so verstehen, dass A_i zunächst durch $A_i' = \{(i,a) : a \in A_i\}$ ersetzt wird (was bis auf eine Bijektion keinen Unterschied macht). Man hat dann $A_0' \cap A_1' = \emptyset$ und $A_0 \sqcup A_1 = A_0' \cup A_1'$. Die disjunkte Vereinigung hat den Vorzug, dass $A_0 \sqcup A_1$ zu $B_0 \sqcup B_1$ isomorph ist, wenn jeweils A_i zu B_i isomorph ist. Dies gilt nicht für die einfache Vereinigung, wie man etwa anhand von

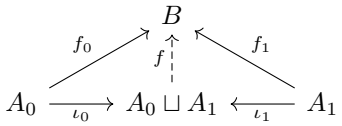
$$|\{0,1\} \cup \{1,2\}| = 3 \neq 4 = |\{0,1\} \cup \{2,3\}|$$

sehen kann. Weiter spricht für die disjunkte Vereinigung als Summe, dass sie eine universelle Eigenschaft erfüllt (welche zur Eigenschaft des Produktes dual ist). Wir betrachten dazu die injektiven Funktionen

$$\iota_i : A_i \to A_0 \sqcup A_1 \quad \text{mit} \quad \iota_i(a) = (i,a).$$

Mit diesen lässt sich das folgende Ergebnis formulieren.

Proposition 4.4.3. *Für beliebige Funktionen $f_i : A_i \to B$ gibt es genau eine Funktion $f : A_0 \sqcup A_1 \to B$, die das Diagramm*



kommutativ macht, das heißt, für die jeweils $f_i = f \circ \iota_i$ gilt.

Beweis. Die einzig mögliche Definition ist $f((i,a)) = f_i(a)$. □

Ein wichtiger Vorzug des systematischen Ansatzes ist, dass wir nun ein Kriterium für die ‚richtige' Definition von Summe und Produkt auf Vektorräumen haben: Gesucht sind die im Wesentlichen eindeutigen Konstruktionen, welche die universellen Eigenschaften von oben erfüllen, wobei alle Abbildungen Morphismen und also in diesem Fall linear sein sollen.

Definition 4.4.4. Für K-Vektorräume U und W setzen wir

$$U \oplus W := \{(u,w) : u \in U \text{ und } w \in W\},$$

wobei Skalarmultiplikation und Addition durch

$$\lambda \cdot (u,w) = (\lambda \cdot u, \lambda \cdot w) \quad \text{und} \quad (u,w) + (u',w') = (u+u', w+w')$$

erklärt sind.

Wir werden später sehen, dass die Notation nicht mit Definition 2.4.7 in Konflikt steht. Zunächst überzeuge man sich von der folgenden Tatsache.

Lemma 4.4.5. *Ist U und W jeweils ein K-Vektorraum, so gilt dies auch für $U \oplus W$.*

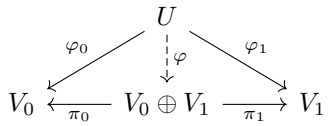
Bezüglich der Notation $\oplus$ lässt sich fragen, ob die Definition von $U \oplus W$ nicht eher ein Produkt als eine Summe nahelegt. Es mag zunächst erstaunen, dass endliche Summen und Produkte im Fall von Vektorräumen zusammenfallen. Genauer erfüllt $\oplus$ beide universelle Eigenschaften. Wir betrachten dazu

$$\pi_i : V_0 \times V_1 \to V_i \quad \text{mit} \quad \pi_i\big((v_0,v_1)\big) = v_i,$$

$$\iota_i : V_i \to V_0 \oplus V_1 \quad \text{mit} \quad \iota(v) = \begin{cases} (v,0) & \text{für } i = 0, \\ (0,v) & \text{für } i = 1. \end{cases}$$

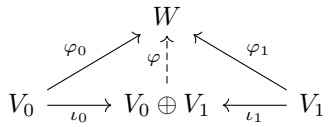
Man überzeuge sich, dass die Abbildungen π_i und ι_i linear sind.

Proposition 4.4.6. *(a) Für lineare Abbildungen $\varphi_i : U \to V_i$ gibt es genau eine lineare Abbildung $\varphi : U \to V_0 \oplus V_1$, die das Diagramm*



kommutativ macht, das heißt, für die jeweils $\varphi_i = \pi_i \circ \varphi$ gilt.

(b) Für lineare Abbildungen $\varphi_i : V_i \to W$ gibt es genau eine lineare Abbildung $\varphi : V_0 \oplus V_1 \to W$, die das Diagramm



kommutativ macht, das heißt, für die jeweils $\varphi_i = \varphi \circ \iota_i$ gilt.

Beweis. (a) Wie im Fall von Mengen ist $\varphi(u) = (\varphi_0(u), \varphi_1(u))$.

(b) Wir setzen

$$\varphi\big((v_0, v_1)\big) = \varphi_0(v_0) + \varphi_1(v_1)$$

und erhalten eine lineare Abbildung mit

$$\varphi \circ \iota_0(v_0) = \varphi\big((v_0, 0)\big) = \varphi_0(v_0) + \varphi_1(0) = \varphi_0(v_0) + 0 = \varphi_0(v_0).$$

Nun nehmen wir an, dass das Diagramm mit $\varphi' : V_0 \oplus V_1 \to W$ an der Stelle von φ ebenfalls kommutiert. Es gilt, $\varphi = \varphi'$ zu zeigen. Dazu wählt man jeweils eine Basis B_i von V_i. Wie man prüfen möge, erhält man eine Basis von $V_0 \oplus V_1$ als

$$B := \{(b, 0) : b \in B_0\} \cup \{(0, b) : b \in B_1\}.$$

Man rechnet etwa

$$\varphi'\big((b, 0)\big) = \varphi' \circ \iota_0(b) = \varphi_0(b) = \varphi\big((b, 0)\big),$$

sodass φ und φ' auf allen Basisvektoren übereinstimmen. □

Im folgenden alternativen Beweis für Teil (b) ist zwar die Definition von φ nicht so transparent. Dafür sieht man, wie sich endliche Produkte von Vektorräumen in disjunkte Vereinigungen von Basen und damit in Summen von Mengen übersetzen.

Bemerkung 4.4.7. In Teil (b) des vorigen Beweises steht B in Bijektion zur Summe $B_0 \sqcup B_1$. Genauer erfüllt B zusammen mit den Funktionen $\iota_i' : B_i \to B$ mit $\iota_0'(b) = (b, 0)$ und $\iota_1'(b) = (0, b)$ die universelle Eigenschaft für die Summe von Mengen. Wir können ι_i' auch jeweils als Funktion mit Wertebereich W auffassen. Es gilt dann $\iota_i = \mathrm{Abb}(\iota_i')$, da die linearen Abbildungen auf der Basis B_i übereinstimmen. Schreibt man weiter $\varphi_i = \mathrm{Abb}(f_i)$ mit $f_i : B_i \to W$, so erfüllt φ genau dann die Bedingung $\varphi_i = \varphi \circ \iota_i$, wenn $\varphi = \mathrm{Abb}(f)$ für $f : B \to W$ mit $f_i = f \circ \iota_i'$ gilt. Gemäß der universellen Eigenschaft für die Summe von Mengen hat man eine eindeutige Lösung für f und also auch für φ.

Wir Stellen nun die Verbindung mit einer bereits bekannten Notation her.

Bemerkung 4.4.8. In Definition 2.4.7 hatten wir für Untervektorräume $V_i \subseteq V$ schon eine Summe $V_0 \oplus^V V_1 \subseteq V$ erklärt. Diese hängt im Allgemeinen vom umgebenden Raum V ab. Genauer ist damit gemeint, dass aus $V_i \cong V_i'$ nicht unbedingt auch $V_0 \oplus^V V_1 \cong V_0' \oplus^V V_1'$ folgt (aber $V_0 \oplus V_1 \cong V_0' \oplus V_1'$ gemäß Definition 4.4.4). Hat man allerdings $V_0 \cap V_1 = \{0\}$, so ist $V_0 \oplus^V V_1$ zu dem in Definition 4.4.4 erklärten Raum $V_0 \oplus V_1$ isomorph (weswegen wir in diesem Spezialfall auch zuvor schon $\oplus$ statt $\oplus^V$ geschrieben haben). Man kann dies auf zwei Arten sehen: Zum einen ist $V_0 \oplus V_1 \ni (v_0, v_1) \mapsto v_0 + v_1 \in V_0 \oplus^V V_1$ ein explizit anzugebender Isomorphismus. Zum anderen erfüllt $V_0 \oplus^V V_1$ zusammen mit den Inklusionen $\iota_i : V_i \to V_0 \oplus^V V_1$ (also $\iota_i(v) = v$) die universelle Eigenschaft aus Teil (b) von Proposition 4.4.6, wobei $\varphi(v_0 + v_1) = \varphi_0(v_0) + \varphi_1(v_1)$ für $v_i \in V_i$ gilt. Zumindest a posteriori lässt sich umgekehrt auch $\oplus$ auf $\oplus^V$ reduzieren. Über die Injektionen $\iota_i : V_i \to V_0 \oplus V_1 =: V$ sind nämlich die V_i isomorph zu Untervektorräumen $U_i := \operatorname{img}(\iota_i)$ von V, für die man $U_0 \oplus^V U_1 = V = V_0 \oplus V_1$ erhält.

Interessanterweise fallen Summen und Produkte über unendliche Familien von Vektorräumen nicht mehr zusammen. Wir weisen zunächst darauf hin, dass ein Tupel (v_0, v_1) mit $v_i \in V_i$ als Funktion $f : \{0, 1\} \to V_0 \cup V_1$ mit $f(i) = v_i$ aufgefasst werden kann. Dies inspiriert die folgende Verallgemeinerung von Tupeln auf unendliche Indexmengen.

Definition 4.4.9. Für K-Vektorräume V_i mit $i \in I$ erklären wir

$$\bigotimes_{i\in I} V_i = \left\{ f : I \to \bigcup_{i\in I} V_i \;\middle|\; f(i) \in V_i \text{ für alle } i \in I \right\}.$$

Wir setzen dabei $\operatorname{supp}(f) = \{i \in I : f(i) \neq 0\}$ und definieren

$$\bigoplus_{i\in I} V_i = \left\{ f \in \bigotimes_{i\in I} V_i \;\middle|\; \operatorname{supp}(f) \text{ ist endlich} \right\}.$$

Skalarmultiplikation und Addition sind jeweils gegeben durch

$$(\lambda \cdot f)(i) = \lambda \cdot f(i) \quad \text{und} \quad (f + g)(i) = f(i) + g(i).$$

Wir betrachten die Funktionen

$$\pi_j : \bigotimes_{i\in I} V_i \to V_j \quad \text{mit} \quad \pi_j(f) = f(j),$$

$$\iota_j : V_j \to \bigoplus_{i\in I} V_i \quad \text{mit} \quad \iota_j(v)(i) = \begin{cases} v & \text{für } j = i, \\ 0 & \text{sonst.} \end{cases}$$

Ist I endlich, so gilt natürlich $\bigoplus_{i\in I} V_i = \bigotimes_{i\in I} V_i$, was weiteres Licht auf Proposition 4.4.6 wirft. Man überzeuge sich auch hier von dem folgenden Ergebnis.

Lemma 4.4.10. *Ist V_i für jeden Index $i \in I$ ein K-Vektorraum, so gilt dasselbe auch für $\bigotimes_{i\in I} V_i$ und für $\bigoplus_{i\in I} V_i$. Die Abbildungen π_j und ι_j sind jeweils linear.*

Die universellen Eigenschaften verallgemeinern sich aus dem Endlichen:

Proposition 4.4.11. *(a) Für lineare Abbildungen $\varphi_i : U \to V_i$ gibt es genau eine lineare Abbildung $\varphi : U \to \bigotimes_{i\in I} V_i$ mit $\varphi_i = \pi_i \circ \varphi$ für alle $i \in I$.*

(b) Für lineare Abbildungen $\varphi_i : V_i \to W$ gibt es genau eine lineare Abbildung $\varphi : \bigoplus_{i\in I} V_i \to W$ mit $\varphi_i = \varphi \circ \iota_i$ für alle $i \in I$.

Beweis. (a) Die Funktion $\varphi(u) : I \to \bigcup_{i\in I} V_i$ kann nur durch $\varphi(u)(i) = \varphi_i(u)$ gegeben sein (wegen $\pi_i \circ \varphi(u) = \pi_i(\varphi(u)) = \varphi(u)(i)$).

(b) Man erhält eine lineare Abbildung durch

$$\varphi(f) = \sum_{i\in\mathrm{supp}(f)} \varphi_i(f(i)) = \sum_{j=1}^{n} \varphi_{i(j)}(f(i(j))) \quad \text{für} \quad \mathrm{supp}(f) = \{i(1), \dots, i(n)\},$$

wobei vorausgesetzt ist, dass die $i(j)$ paarweise verschieden sind. Für $v \in V_i$ hat man jeweils $\mathrm{supp}(\iota_i(v)) \subseteq \{i\}$ (mit Gleichheit genau für $v \neq 0$) und also

$$\varphi \circ \iota_i(v) = \varphi\big(\iota_i(v)\big) = \varphi_i\big(\iota_i(v)(i)\big) = \varphi_i(v).$$

Um die Eindeutigkeit nachzuweisen, nehmen wir an, dass auch $\varphi_i = \varphi' \circ \iota_i$ für alle $i \in I$ gilt. Ist jeweils B_i eine Basis von V_i, so ist

$$B = \{\iota_i(b) : i \in I \text{ und } b \in B_i\}$$

eine Basis von $\bigoplus_{i\in I} V_i$ (aber im Allgemeinen nicht von $\bigotimes_{i\in I} V_i$). Man erhält jeweils

$$\varphi\big(\iota_i(b)\big) = \varphi_i(b) = \varphi'\big(\iota_i(b)\big).$$

Damit stimmen φ und φ' auf Basisvektoren und somit insgesamt überein. □

Die Basis B aus dem vorigen Beweis steht mittels $\iota_i(b) \mapsto (i, b)$ in Bijektion zur disjunkten Vereinigung

$$\coprod_{i\in I} B_i = \{(i, b) : i \in I \text{ und } b \in B_i\}$$

von Mengen. Auch für unendliche Produkte erhält man damit einen alternativen Beweis analog zu Bemerkung 4.4.7. Wir stellen eine weitere Verbindung mit einer bereits bekannten Konstruktion her:

Bemerkung 4.4.12. Gilt $V_i = K$ für alle $i \in I$, so stimmt $\bigoplus_{i\in I} V_i$ mit dem Vektorraum $\bigoplus_I K$ aus Definition 2.3.10 überein. Hat I endliche Kardinalität $|I| = n$, so stimmt also auch $\otimes_{i\in I} K$ mit K^n überein. Wie beim Rechnen mit Zahlen schreibt man also einen Exponenten, um ein mehrfaches Produkt anzuzeigen.

Die folgende Notation (welche sich noch allgemeiner und auch für Produkte definieren ließe) wird uns insbesondere für die Analyse von Matrizen helfen.

Definition 4.4.13. Für lineare Abbildungen $\varphi_i : V_i \to W$ bezeichnet $[\varphi_i]_{i\in I}$ die eindeutig bestimmte Abbildung $\varphi : \bigoplus_{i\in I} V_i \to W$ aus Proposition 4.4.11(b). Ist speziell $W = \bigoplus_{i\in I} V_i$ und jeweils $\varphi_i = \iota_i \circ \psi_i$ mit $\psi_i : V_i \to V_i$, so setzen wir

$$\bigoplus_{i\in I} \psi_i = [\iota_i \circ \psi_i]_{i\in I} : \bigoplus_{i\in I} V_i \to \bigoplus_{i\in I} V_i.$$

Für endliches $I = \{1, \dots, n\}$ schreiben wir auch $\psi_1 \oplus \dots \oplus \psi_n$ statt $\bigoplus_{i\in I} \psi_i$.

Die Definition vereinfacht sich für Summen von Untervektorräumen. Wir konzentrieren uns auf den Fall mit endlich vielen Summanden:

Bemerkung 4.4.14. Wir nennen $V = V_1 \oplus^V \dots \oplus^V V_n$ mit $V_i \neq \{0\}$ für $1 \leq i \leq n$ eine direkte Summe, wenn $\{v_1, \dots, v_n\}$ für beliebige $v_i \in V_i \backslash \{0\}$ linear unabhängig ist. Die Summe $V_1 \oplus \dots \oplus V_n$ aus Definition 4.4.9 ist dann mittels $f \mapsto \sum_{i=1}^n f(i)$ isomorph zu V (wie in Bemerkung 4.4.8). Lässt man den Isomorphismus und die

Inklusionsabbildungen von V_i nach V implizit, so ist $\psi_1 \oplus \cdots \oplus \psi_n : V \to V$ für lineare Abbildungen $\psi_i : V_i \to V_i$ gegeben durch

$$(\psi_1 \oplus \cdots \oplus \psi_n) \left(\sum_{i=1}^{n} v_i \right) = \sum_{i=1}^{n} \psi_i (v_i) \quad \text{für} \quad v_i \in V_i.$$

Dies folgt aus der Konstruktion von φ im Beweis von Proposition 4.4.11(b). Hat man weitere Abbildungen $\varphi_i : V_i \to V_i$, so folgt

$$(\psi_1 \oplus \cdots \oplus \psi_n) \circ (\varphi_1 \oplus \cdots \oplus \varphi_n) = (\psi_1 \circ \varphi_1) \oplus \cdots \oplus (\psi_n \circ \varphi_n).$$

Ein analoges Resultat lässt sich für unendliche Summen zeigen.

Mit Hilfe des folgenden Kriteriums werden wir lineare Abbildungen in einfachere Bestandteile zerlegen können.

Proposition 4.4.15. *Wir betrachten eine lineare Abbildung $\psi : V \to V$. Ist der Vektorraum $V = V_1 \oplus^V \cdots \oplus^V V_n$ als direkte Summe gegeben, so sind äquivalent:*

(i) Man findet lineare Abbildungen $\psi_i : V_i \to V_i$ mit $\psi = \psi_1 \oplus \cdots \oplus \psi_n$.
(ii) Es gilt jeweils $\psi(v) \in V_i$ für $v \in V_i$.

Beweis. Gilt (i), so liefert die vorige Bemerkung $\psi(v) = \psi_i(v) \in V_i$ für $v \in V_i$. Hat man (ii), so erhält man lineare Abbildungen $\psi_i : V_i \to V_i$ mit $\psi_i(v) = \psi(v)$. Wieder mit der vorigen Bemerkung ergibt sich

$$(\psi_1 \oplus \cdots \oplus \psi_n) \left(\sum_{i=1}^{n} v_i \right) = \sum_{i=1}^{n} \psi(v_i) = \psi \left(\sum_{i=1}^{n} v_i \right)$$

für $v_i \in V_i$, sodass (i) gilt. □

4.5. **Anwendung: Basiswechsel und Matrizen in Blockdiagonalform.** Will man eine lineare Abbildung $\varphi : V \to W$ zwischen K-Vektorräumen mit Dimensionen $m, n \in \mathbb{N}$ durch eine Matrix darstellen, so muss man Isomorphismen $V \cong K^m$ und $W \cong K^n$ fixieren. Ein Beispiel haben wir in Definition 4.1.17 gesehen. Sind V und W bereits gleich K^m und K^n, so bietet es sich natürlich an, als Isomorphismen die Identitätsabbildung zu wählen. Diese offensichtliche Wahl ist aber keineswegs verbindlich. Oft lässt sich die darstellende Matrix entscheidend vereinfachen, indem man andere Isomorphismen $K^m \cong K^m$ und $K^n \cong K^n$ verwendet.[95] Im vorliegenden Teilkapitel legen wir die Grundlagen für solche ‚Basiswechsel'. Konkrete Verfahren für die geschickte Wahl von Automorphismen werden dann im nächsten Kapitel besprochen.

Definition 4.5.1. Es seien $\sigma : V' \to V$ und $\tau : W' \to W$ zwei Isomorphismen. Für eine lineare Abbildung $\varphi : V \to W$ definieren wir $\varphi^\sigma_\tau : V' \to W'$ als die eindeutige lineare Abbildung, für die das Diagramm

$$\begin{array}{ccc} V' & \xrightarrow{\varphi^\sigma_\tau} & W' \\ {\scriptstyle\sigma}\downarrow & & \downarrow{\scriptstyle\tau} \\ V & \xrightarrow{\varphi} & W \end{array}$$

[95] Ein Isomorphismus $V \cong V$ mit gleichem Definitions- und Wertebereich heißt auch Automorphismus von V.

kommutiert, die also durch $\varphi_\tau^\sigma = \tau^{-1} \circ \varphi \circ \sigma$ gegeben ist. Ist speziell $V = V' = K^n$ und $W = W' = K^m$ sowie $\varphi = \mathrm{Abb}(M)$ mit $M \in K^{m\times n}$, so definieren wir die Matrix $M_\tau^\sigma \in K^{m\times n}$ durch

$$\mathrm{Abb}\,(M_\tau^\sigma) = \mathrm{Abb}(M)_\tau^\sigma.$$

Gilt dabei weiter $\sigma = \mathrm{Abb}(S)$ und $\tau = \mathrm{Abb}(T)$ mit $S \in K^{n\times n}$ und $T \in K^{m\times m}$, so schreibt man auch M_T^S für M_τ^σ.

Wir halten die folgende grundlegende Eigenschaft fest.

Lemma 4.5.2. *Es gilt $M_T^S = T^{-1} \cdot M \cdot S$.*

Beweis. Dies folgt aus

$$\begin{aligned}\mathrm{Abb}(M_T^S) &= \mathrm{Abb}(M)_{\mathrm{Abb}(T)}^{\mathrm{Abb}(S)} \\ &= \mathrm{Abb}(T)^{-1} \circ \mathrm{Abb}(M) \circ \mathrm{Abb}(S) = \mathrm{Abb}(T^{-1} \cdot M \cdot S).\end{aligned}$$ □

Der Begriff des Basiswechsels erhellt sich wie folgt.

Bemerkung 4.5.3. Ist $\mathrm{Abb}(S)$ mit $S \in K^{n\times n}$ ein Isomorphismus, so bilden die Spalten $s_i := S \cdot e_i$ eine Basis des K^n. Genauer bilden sie eine geordnete Basis, für die also eine Reihenfolge der Basisvektoren $s_1, \dots, s_n$ festgelegt ist. Umgekehrt ist S durch diese geordnete Basis bestimmt. Genauso entspricht T einer geordneten Basis $t_1, \dots, t_m$. Wir nehmen nun an, dass $\varphi = \mathrm{Abb}(M)$ gegeben ist durch

$$\varphi(s_j) = \sum_{i=1}^m \lambda_{ij} \cdot t_i.$$

Man hat dann $M_T^S = (\lambda_{ij})$. Dies gilt wegen dem vorigen Lemma und

$$T \cdot (\lambda_{ij}) \cdot e_k = T \cdot \begin{pmatrix} \lambda_{1k} \\ \vdots \\ \lambda_{mk} \end{pmatrix} = \sum_{i=1}^m \lambda_{ik} \cdot t_i = M \cdot s_k = M \cdot S \cdot e_k.$$

In diesem konkreten Sinn ist also M_T^S die Darstellungsmatrix von φ bezüglich der Basen $s_1, \dots, s_n$ und $t_1, \dots, t_m$.[96] Angenommen nun, wir haben weitere Matrizen $P \in K^{n\times n}$ und $Q \in K^{m\times m}$ gegeben, wobei die Spalten $p_i = P \cdot e_i$ und $q_i = Q \cdot e_i$ jeweils eine Basis bilden. Unser Ziel ist, eine Darstellung

$$\varphi(p_i) = \sum_{j=1}^m \mu_{ij} \cdot q_j$$

bezüglich dieser Basen zu bestimmen. Mit

$$M = T \cdot M_T^S \cdot S^{-1} = T \cdot (\lambda_{ij}) \cdot S^{-1}$$

ergibt sich

$$(\mu_{ij}) = M_Q^P = Q^{-1} \cdot M \cdot P = Q^{-1} \cdot T \cdot (\lambda_{ij}) \cdot S^{-1} \cdot P.$$

Aus den Matrizen P, Q, S, T – deren Spalten ja die Vektoren der beteiligten Basen sind – können wir also die gewünschten Koeffizienten μ_{ij} berechnen.

[96] Speziell für den Fall der Standardbasen – in dem S und T jeweils die Einheitsmatrix sind – hat man $M_T^S = M$. Viele Lehrbücher führen von Anfang an Darstellungsmatrizen M_T^S bezüglich beliebiger Basen ein. Der Autor des vorliegenden Skriptums hat sich entschieden, den Fokus zunächst auf die Standardbasen zu legen, weil man den allgemeinen Fall wie gezeigt darauf zurückführen kann (und dieser also nur scheinbar allgemeiner ist).

Das folgende Ergebnis zeigt, dass fast alle Information unter Basiswechseln verloren gehen kann.

Proposition 4.5.4. *Für lineare Abbildungen $\varphi, \psi : V \to W$ zwischen zwei endlichdimensionalen Vektorräumen sind äquivalent:*

(i) Es gibt Isomorphismen $\sigma : V \to V$ und $\tau : W \to W$ mit $\psi = \varphi^\sigma_\tau$.
(ii) Man hat $\operatorname{rank}(\varphi) = \operatorname{rank}(\psi)$.

Beweis. Wir nehmen zunächst (i) an und folgern (ii). Dazu zeigen wir, dass

$$\operatorname{img}(\varphi) \ni w \mapsto \tau(w) \in \operatorname{img}(\psi)$$

ein Isomorphismus ist. Für $w = \varphi(v) \in \operatorname{img}(\varphi)$ liegt $\tau(w) = \tau \circ \varphi(v) = \psi \circ \sigma(v)$ tatsächlich in $\operatorname{img}(\psi)$. Umgekehrt lässt sich ein beliebiges $w = \psi(v) \in \operatorname{img}(\psi)$ wegen der Surjektivität von σ als $w = \psi \circ \sigma(v') = \tau \circ \varphi(v) = \tau(w')$ mit $w' = \varphi(v) \in \operatorname{img}(\varphi)$ schreiben. Nun nehmen wir (ii) an, wo mit dem Dimensionssatz auch

$$\dim(\ker(\varphi)) = \dim(\ker(\psi))$$

gilt. Wir finden daher einen Isomorphismus $\sigma : V \to V$ mit $\ker(\varphi) = \sigma[\ker(\psi)]$ oder äquivalent $\ker(\varphi \circ \sigma) = \ker(\psi)$. Der Isomorphiesatz liefert kommutative Diagramme

$$\begin{array}{ccc} V & \xrightarrow{\varphi\circ\sigma} & W \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle \overline{\varphi\circ\sigma}} & \\ V/\ker(\psi) & & \end{array} \qquad \text{und} \qquad \begin{array}{ccc} V & \xrightarrow{\psi} & W, \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle \overline{\psi}} & \\ V/\ker(\psi) & & \end{array}$$

wobei die linearen Abbildungen $\overline{\varphi \circ \sigma}$ und $\overline{\psi}$ injektiv sind und gleiches Bild wie φ beziehungsweise ψ haben. Es gibt dann einen Isomorphismus $\tau : W \to W$, für welchen man $\tau \circ \overline{\psi} = \overline{\varphi \circ \sigma}$ hat, wie man als Übung zeige (wobei man $\overline{\psi} = \operatorname{Abb}(f)$ und $\overline{\varphi \circ \sigma} = \operatorname{Abb}(g)$ mit $f, g : B \to W$ für eine Basis B von $V/\ker(\psi)$ schreiben mag). Dann gilt aber auch $\tau \circ \psi = \tau \circ \overline{\psi} \circ \pi = \overline{\varphi \circ \sigma} \circ \pi = \varphi \circ \sigma$. □

Auf der Ebene von Matrizen hat die Proposition die folgende Anwendung: Für gegebenes $M \in K^{n\times n}$ mit $\operatorname{rank}(M) = k$ wählt man $N \in K^{n\times n}$ so, dass die ersten k Diagonaleinträge von N gleich Eins und alle anderen Einträge gleich Null sind. Da dann auch $\operatorname{rank}(N) = k$ gilt, liefert die Proposition nun invertierbare Matrizen $S, T \in K^{n\times n}$ mit $N = M^S_T = T^{-1} \cdot M \cdot S$. Man kann also eine beliebige Matrix M durch Basiswechsel in eine sehr einfache Form bringen. Dies ist aber nur bedingt fruchtbar, weil fast die gesamte Komplexität von M in die Matrizen S und T übergegangen ist. Einen sinnvolleren Begriff erhält man, wenn man mit $S = T$ fordert, dass im Ausgangs- und im Zielraum derselbe Basiswechsel erfolgt.

Definition 4.5.5. Zwei lineare Abbildungen $\varphi : V \to V$ und $\psi : W \to W$ heißen ähnlich, wenn es einen Isomorphismus $\sigma : V \to W$ mit $\varphi = \psi^\sigma_\sigma$ gibt. Matrizen M und N nennen wir ähnlich, wenn $\operatorname{Abb}(M)$ und $\operatorname{Abb}(N)$ ähnlich sind.

Gemäß Lemma 4.5.2 sind zwei Matrizen M und N aus $K^{n\times n}$ also ähnlich, wenn es eine invertierbare Matrix $S \in K^{n\times n}$ mit $M = S^{-1} \cdot N \cdot S$ gibt. Als Übung möge man verifizieren:

Lemma 4.5.6. *Ähnlichkeit ist eine Äquivalenzrelation auf der Klasse der linearen Abbildungen.*

Im Kontrast mit Proposition 4.5.4 zeigt das folgende Ergebnis exemplarisch, dass relevante Eigenschaften von Matrizen unter Ähnlichkeit erhalten bleiben.

Lemma 4.5.7. *Ähnliche Matrizen haben gleiche Determinante.*

Beweis. Theorem 3.2.7 und sein Korollar liefern

$$\det\left(S^{-1}\cdot M\cdot S\right)=\frac{1}{\det(S)}\cdot\det(M)\cdot\det(S)=\det(M).\qquad\square$$

Im vorigen Teilkapitel haben wir Abbildungen der Form $\varphi_1\oplus\cdots\oplus\varphi_n$ betrachtet. Wie wir nun zeigen, sind diese per Basiswechsel durch spezielle Matrizen darstellbar.

Definition 4.5.8. Für Matrizen $M_i=(\lambda^i_{jk})\in K^{n(i)\times n(i)}$ setzen wir $N=\sum_{i=1}^n n(i)$ und definieren $M_1\oplus\cdots\oplus M_n=(\lambda_{JK})\in K^{N\times N}$ wie folgt: Man hat jeweils

$$\lambda_{I+j,I+k}=\lambda^i_{jk}\quad\text{für}\quad I=\sum_{l=1}^{i-1}n(l)\text{ und }1\leq j,k\leq n(i)$$

sowie $\lambda_{JK}=0$ für alle übrigen Einträge.

Man spricht von Matrizen in Blockdiagonalform.

Beispiel 4.5.9. Es gilt

$$\begin{pmatrix}\lambda_{11}&\lambda_{12}\\\lambda_{21}&\lambda_{22}\end{pmatrix}\oplus\begin{pmatrix}\mu_{11}&\mu_{12}\\\mu_{21}&\mu_{22}\end{pmatrix}=\begin{pmatrix}\lambda_{11}&\lambda_{12}&0&0\\\lambda_{21}&\lambda_{22}&0&0\\0&0&\mu_{11}&\mu_{12}\\0&0&\mu_{21}&\mu_{22}\end{pmatrix}.$$

Das folgende wichtige Kriterium sagt uns (in Verbindung mit Proposition 4.4.15), wann wir Matrizen durch Basiswechsel in Blockdiagonalform bringen können.

Proposition 4.5.10. *Für eine Matrix $M\in K^{n\times n}$ sind äquivalent:*

(i) Die Abbildung $\mathrm{Abb}(M)$ *ist ähnlich zu einer Summe $\varphi_1\oplus\cdots\oplus\varphi_k$ für lineare Abbildungen $\varphi_i:V_i\to V_i$ mit* $\dim(V_i)=n(i)$.

(ii) Man hat $M=S\cdot(M_1\oplus\cdots\oplus M_k)\cdot S^{-1}$ für Matrizen $M_i\in K^{n(i)\times n(i)}$ und invertierbares $S\in K^{n\times n}$.

Beweis. Wir nehmen zunächst (ii) an und leiten (i) ab. Sei jeweils

$$V_i=\mathrm{Span}\left(e_{I+1},\ldots,e_{I+n(i)}\right)\quad\text{mit}\quad I=\sum_{l=0}^{i-1}n(l).$$

Für $M_i=(\lambda^i_{jk})$ und für I wie eben hat man

$$(M_1\oplus\cdots\oplus M_k)\cdot v=\sum_{j=1}^{n(i)}\sum_{k=1}^{n(i)}\lambda^i_{jk}\cdot\mu_k\cdot e_{I+j}\in V_i\quad\text{für}\quad v=\sum_{j=1}^{n(i)}\mu_j\cdot e_{I+j}\in V_i.$$

Gemäß Proposition 4.4.15 ist also $\mathrm{Abb}(M_1\oplus\cdots\oplus M_k)$ von der Form $\varphi_1\oplus\cdots\oplus\varphi_k$ für lineare Abbildungen φ_i auf unseren V_i (wobei φ_i bis auf den offensichtlichen Isomorphismus $V_i\cong K^{n(i)}$ gleich $\mathrm{Abb}(M_i)$ ist).

Wir setzen nun (i) voraus und folgern (ii). Bis auf Isomorphie können wir annehmen, dass die V_i Untervektorräume des K^n sind, für die

$$K^n=V_1\oplus^V\cdots\oplus^V V_k\quad\text{und}\quad\mathrm{Abb}(M)=\varphi_1\oplus\cdots\oplus\varphi_k$$

gilt. Wähle eine Basis $b_1, \dots, b_n$ des K^n so, dass $B_i = \{b_{I+1}, \dots, b_{I+n(i)}\}$ mit I wie oben jeweils eine Basis von V_i ist. Sei $M_i \in K^{n(i)\times n(i)}$ die darstellende Matrix von φ_i bezüglich B_i. Explizit ist also $M_i = (\lambda^i_{jk})$ charakterisiert durch

$$M \cdot b_{I+j} = \varphi_i(b_{I+j}) = \sum_{k=1}^{n(i)} \lambda^i_{jk} \cdot b_{I+k},$$

wobei die erste Gleichheit auf Bemerkung 4.4.14 zurückgeht. Sei weiter S die Matrix mit Spalten $S \cdot e_j = b_j$. Mit Bemerkung 4.5.3 ist dann $S^{-1} \cdot M \cdot S = M_1 \oplus \dots \oplus M_k$, wie gewünscht. □

Das folgende Ergebnis ergibt sich abstrakt aus dem vorigen Beweis und Bemerkung 4.4.14, lässt sich aber auch gut explizit nachrechnen.

Lemma 4.5.11. *Für Matrizen $M_i, N_i \in K^{n(i)\times n(i)}$ gilt*

$$(M_1 \oplus \dots \oplus M_k) \cdot (N_1 \oplus \dots \oplus N_k) = (M_1 \cdot N_1) \oplus \dots \oplus (M_k \cdot N_k).$$

Zum Abschluss dieses Kapitels stellen wir fest, dass die vorigen Überlegungen einen klaren Nutzen für Berechnungen haben:

Beispiel 4.5.12. Wir wollen die Potenzen

$$M^k = \underbrace{M \cdot \ldots \cdot M}_{k \text{ Faktoren}} \quad \text{für} \quad M = \begin{pmatrix} 26 & -18 \\ 36 & -25 \end{pmatrix} \in \mathbb{R}^2$$

berechnen. Dazu beachten wir

$$M \cdot v_1 = 2 \cdot v_1 \text{ für } v_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix} \quad \text{und} \quad M \cdot v_2 = -v_2 \text{ für } v_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Im nächsten Kapitel werden über die Theorie der Eigenwerte systematisch untersuchen, wie man solche v_i findet. Hier setzen wir $\mathbb{R}^2 = V_1 \oplus V_2$ mit $V_i = \mathrm{Span}(v_i)$, sodass aus $w \in V_i$ jeweils $M \cdot w \in V_i$ folgt. Mit der obigen Proposition und ihrem Beweis erhalten wir

$$M = S \cdot (M_1 \oplus M_2) \cdot S^{-1} = \begin{pmatrix} 3 & 2 \\ 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 3 & -2 \\ -4 & 3 \end{pmatrix}.$$

Die Potenzen ergeben sich als

$$M^k = S \cdot \begin{pmatrix} 2^k & 0 \\ 0 & (-1)^k \end{pmatrix} \cdot S^{-1},$$

weil ‚zwischen' den Faktoren jeweils $S^{-1} \cdot S = \mathbb{1}$ gilt (wie man für $k = 2$ explizit machen möge). Dies ist deshalb so, weil wir im Ausgangs- und Zielraum denselben Basiswechsel vollziehen (und also M in der Form $S \cdot M' \cdot S^{-1}$ statt $S \cdot M' \cdot T^{-1}$ mit $S \neq T$ schreiben). Die Vorzüge von Definition 4.5.5 gegenüber Proposition 4.5.4 manifestieren sich also auch auf rechnerischer Ebene.

5. Eigenwerttheorie

Im vorigen Kapitel haben wir gesehen, dass man eine Matrix in eine besonders gut handhabbare Form bringen kann, wenn man Untervektorräume findet, die unter der zugehörigen Abbildung abgeschlossen sind (Propositionen 4.4.15 und 4.5.10). Das vorliegende Kapitel untersucht, welche Normalformen für Matrizen konkret verfügbar sind. Weiter behandeln wir das Skalarprodukt und seine Implikationen für geometrische Fragen wie auch für Normalformen.

5.1. **Diagonalisierbarkeit.** Wie oben erwähnt suchen wir Untervektorräume, die unter einer linearen Abbildung φ abgeschlossen sind. Im einfachsten Fall haben diese die Form $\mathrm{Span}(v)$ mit $\varphi(v) = \lambda \cdot v$ für einen Skalar λ.

Definition 5.1.1. Das Spektrum einer linearen Abbildung $\varphi : V \to V$ auf einem K-Vektorraum V ist gegeben durch

$$\mathrm{Spec}(\varphi) = \{\lambda \in K : \varphi(v) = \lambda \cdot v \text{ für ein } v \neq 0 \text{ aus } V\}.$$

Die Elemente von $\mathrm{Spec}(\varphi)$ heißen Eigenwerte von φ. Für $\lambda \in \mathrm{Spec}(\varphi)$ nennt man

$$E(\lambda) = E_\varphi(\lambda) = \{v \in V : \varphi(v) = \lambda \cdot v\}$$

den Eigenraum von λ. Die Elemente von $E(\lambda)\backslash\{0\}$ heißen Eigenvektoren.

Teil (a) des folgenden Ergebnisses rechtfertigt den Begriff Eigenraum.

Lemma 5.1.2. *(a) Für jedes $\lambda \in \mathrm{Spec}(\varphi)$ ist $E(\lambda)$ ein Untervektorraum.*
(b) Aus $v \in E_\varphi(\lambda)$ folgt $\varphi(v) \in E_\varphi(\lambda)$.

Beweis. (a) Man hat $E(\lambda) = \ker(\varphi - \lambda \cdot \mathrm{id})$.
(b) Es ergibt sich $\varphi(\varphi(v)) = \varphi(\lambda \cdot v) = \lambda \cdot \varphi(v)$. □

Das nächste Ergebnis zeigt insbesondere $E(\lambda) \cap E(\mu) = \{0\}$ für $\lambda \neq \mu$.

Proposition 5.1.3. *Für paarweise verschiedene Eigenwerte $\lambda_1, \dots, \lambda_k$ und Eigenvektoren $v_i \in E(\lambda_i)\backslash\{0\}$ ist die Menge $\{v_1, \dots, v_k\}$ linear unabhängig.*

Beweis. Wir argumentieren per Induktion über k. Aus $\sum_{i=1}^k \mu_i \cdot v_i = 0$ folgt

$$\sum_{i=1}^k \mu_i \lambda_1 \cdot v_i = 0 = \varphi\left(\sum_{i=1}^k \mu_i \cdot v_i\right) = \sum_{i=1}^k \mu_i \cdot \varphi(v_i) = \sum_{i=1}^k \mu_i \lambda_i \cdot v_i$$

und also

$$\sum_{i=2}^k \mu_i \cdot (\lambda_1 - \lambda_i) \cdot v_i = 0.$$

Induktiv ergibt sich $\mu_i \cdot (\lambda_1 - \lambda_i) = 0$ und also $\mu_i = 0$ für jedes $i \geq 2$. Schließlich erhält man $\mu_1 \cdot v_1 = 0$ und also $\mu_1 = 0$. □

Insbesondere erhalten wir eine Schranke für die Zahl der Eigenwerte.

Korollar 5.1.4. *Für jede lineare Abbildung $\varphi : V \to V$ gilt $|\mathrm{Spec}(\varphi)| \leq \dim(V)$.*

Proposition 5.1.3 sagt uns, dass $\bigoplus^V \{E_\varphi(\lambda) : \lambda \in \mathrm{Spec}(\varphi)\}$ eine direkte Summe im Sinn von Beispiel 4.4.8 ist. Ist die Eigenschaft aus der folgenden Definition erfüllt, so lässt sich die Abbildung φ gemäß Proposition 4.4.15 zerlegen.

Definition 5.1.5. Eine lineare Abbildung $\varphi : V \to V$ mit

$$V = \bigoplus^V \{E_\varphi(\lambda) : \lambda \in \mathrm{Spec}(\varphi)\}$$

heißt diagonalisierbar.

Wir werden sehen, dass eine lineare Abbildung auf einem endlich-dimensionalen Vektorraum genau dann die gegebene Bedingung erfüllt, wenn sie durch eine Diagonalmatrix darstellbar ist – also durch eine Matrix (μ_{ij}) mit $\mu_{ij} = 0$ für $i \neq j$. Zunächst führen wir den folgenden Begriff ein.

Definition 5.1.6. Die geometrische Vielfachheit von $\lambda \in \mathrm{Spec}(\varphi)$ ist gegeben durch

$$\gamma(\lambda) = \gamma_\varphi(\lambda) = \dim\left(E_\varphi(\lambda)\right).$$

Wir erhalten nun die folgende Charakterisierung.

Proposition 5.1.7. *Für eine lineare Abbildung $\varphi : V \to V$ auf einem endlich-dimensionalen Vektorraum V sind äquivalent:*

(i) Die Abbildung φ ist diagonalisierbar.
(ii) Es gilt $\sum_{\lambda\in\mathrm{Spec}(\varphi)} \gamma(\lambda) = \dim(V)$.
(iii) Der Vektorraum V hat eine Basis, die aus Eigenvektoren für φ besteht.

Wie der folgende Beweis zeigt, gilt die Äquivalenz von (i) und (iii) auch bei unendlicher Dimension.

Beweis. Mit Proposition 5.1.3 erhält man die verallgemeinerte Dimensionsformel

$$\dim\left(\bigoplus^V \{E(\lambda) : \lambda \in \mathrm{Spec}(\varphi)\}\right) = \textstyle\sum_{\lambda\in\mathrm{Spec}(\varphi)} \gamma(\lambda).$$

Die Äquivalenz zwischen (i) und (ii) folgt mit Lemma 2.4.2. Gilt (i), so erhält man (iii), indem man Basen der verschiedenen Eigenräume zusammenlegt. Umgekehrt kann man eine Basis aus Eigenvektoren auf die verschiedenen Eigenräume aufteilen, wobei man wegen Proposition 5.1.3 (oder im endlich-dimensionalen Fall auch wegen der Kardinalität) jeweils eine Basis erhält. □

Da stets $\gamma(\lambda) \geq 1$ gilt, erhält man die folgende hinreichende Bedingung.

Korollar 5.1.8. *Gilt $|\mathrm{Spec}(\varphi)| = \dim(V)$, so ist $\varphi : V \to V$ diagonalisierbar.*

Mit Blick auf die Matrixdarstellung betrachten wir zunächst den Fall eines einzelnen Eigenwerts. Für $M \in K^{n\times n}$ schreiben wir oft M anstelle von $\mathrm{Abb}(M)$. Beispielsweise sagen wir, dass M diagonalisierbar ist, wenn dies für $\mathrm{Abb}(M)$ gilt.

Lemma 5.1.9. *Hat man $M \in K^{n\times n}$ mit $\mathrm{Spec}(M) = \{\lambda\}$ und $\gamma(\lambda) = n$, so gilt schon $M = \lambda \cdot \mathbb{1}$ für die Einheitsmatrix $\mathbb{1} \in K^{n\times n}$.*

Beweis. Die vorige Proposition liefert eine Basis $s_1, \ldots, s_n$ mit $M \cdot s_i = \lambda \cdot s_i$. Es sei S die Matrix mit Spalten s_i. Gemäß Bemerkung 4.5.3 gilt $S^{-1} \cdot M \cdot S = \lambda \cdot \mathbb{1}$, was die Behauptung impliziert. □

Man kann im vorigen Beweis auch direkter argumentieren, dass $M \cdot v = \lambda \cdot v$ für alle $v \in K^n$ gilt. Die Wahl einer (beliebigen) Basis liefert aber eine nützliche Perspektive: Wir werden das Lemma nämlich nicht nur für den K^n sondern allgemeiner für n-dimensionale (Unter-) Vektorräume verwenden, welche in der Form $\mathrm{Span}(s_1, \ldots, s_n) \subseteq K^N$ gegeben sein mögen. Dabei wechselt man von der Standardbasis des K^N auf eine Basis, die $s_1, \ldots, s_n$ enthält. Mit den Propositionen 4.4.15 und 4.5.10 (beziehungsweise mit ihren Beweisen) ergibt sich das folgende Ergebnis.

Korollar 5.1.10. *Jede diagonalisierbare Matrix M mit* $\mathrm{Spec}(M) = \{\lambda_1, \dots, \lambda_k\}$ *ist ähnlich zur Matrix $M_1 \oplus \ldots \oplus M_k$ mit $M_i = \lambda \cdot \mathbb{1} \in K^{n(i)\times n(i)}$ für $n(i) = \gamma(\lambda_i)$.*

Man beachte, dass $M_1 \oplus \ldots \oplus M_k$ hier eine Diagonalmatrix ist. Umgekehrt ist jede Diagonalmatrix im Sinn von Definition 5.1.5 diagonalisierbar (weil in diesem Fall die Standardbasis aus Eigenvektoren besteht). Dasselbe gilt dann für jede zu einer Diagonalmatrix ähnliche Matrix (weil ein Isomorphismus Eigenvektoren erhält).

Wie man die Diagonalform bestimmt, ist in den zuvor erwähnten Beweisen implizit. Um es explizit zu machen, diskutieren wir einen konkreten Fall.

Beispiel 5.1.11. Wir betrachten die Matrix

$$M = \begin{pmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Mit einem später zu besprechenden Verfahren bestimmt man $\mathrm{Spec}(M) = \{-2, 1\}$. Die Bedingung $M \cdot v = -2 \cdot v$ ist äquivalent zu $(M + 2 \cdot \mathbb{1}) \cdot v = 0$ (wie schon im Beweis von Lemma 5.1.2 verwendet). Durch Lösen des Gleichungssystems erhält man $E(-2) = \mathrm{Span}(s_1, s_2)$ und $E(1) = \mathrm{Span}(s_3)$ mit

$$s_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad s_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad \text{und} \quad s_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Es ergibt sich also

$$S^{-1} \cdot M \cdot S = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{für} \quad S = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Im Beispiel haben wir $v \in E_M(\lambda)$ durch Lösen der Gleichung $(M - \lambda \cdot \mathbb{1}) \cdot v = 0$ bestimmt. Dies motiviert die folgende Konstruktion.

Definition 5.1.12. Das charakteristische Polynom χ_M einer Matrix $M \in K^{n\times n}$ ist gegeben durch

$$\chi_M(x) = \det(M - x \cdot \mathbb{1}).$$

Im Folgenden überzeugen wir uns, dass wir es hier tatsächlich mit einem Polynom zu tun haben.

Bemerkung 5.1.13. Um die Definition zu präzisieren, kann man $\chi_M(x)$ formal über die Leibniz-Formel definieren. Alternativ kann man den Integritätsring $K[x]$ der Polynome in seinen Quotientenkörper $K(x)$ einbetten. Dieser besteht – analog zur Konstruktion von $\mathbb{Q}$ aus $\mathbb{Z}$ – aus gekürzten Ausdrücken p/q für $p, q \in K[x]$ mit $q \neq 0$ (welche den sogenannten rationalen Funktionen entsprechen). Dies hat den Vorteil, dass man $M - x \cdot \mathbb{1}$ als Matrix über dem Körper $K(x)$ auffassen kann, sodass alle Ergebnisse über Determinanten auch in Gegenwart der 'Unbekannten' x verwendbar sind.

Die Eigenwerte ergeben sich nun als Nullstellen.

Proposition 5.1.14. *Für $M \in K^{n\times n}$ gilt*

$$\mathrm{Spec}(M) = \{\lambda \in K : \chi_M(\lambda) = 0\}.$$

Beweis. Hat man $\lambda \in \mathrm{Spec}(M)$, so findet man ein $v \neq 0$ mit $M \cdot v = \lambda \cdot v$ und dann also $(M - \lambda \cdot \mathbb{1}) \cdot v = 0$. Hier ist $M - \lambda \cdot \mathbb{1}$ nicht invertierbar und es gilt $\chi_M(\lambda) = 0$. Ist umgekehrt $M - \lambda \cdot \mathbb{1}$ nicht invertierbar, so ist die zugehörige Abbildung nicht bijektiv und also nicht injektiv. Sie hat also nicht-trivialen Kern, was einen Eigenvektor zum Eigenwert λ liefert. □

Wir liefern nun nach, wie im vorigen Beispiel die Eigenwerte bestimmt wurden.

Beispiel 5.1.15. Für die Matrix M aus Beispiel 5.1.11 liefert die Regel von Sarrus (als Spezialfall der Leibniz-Formel) das charakteristische Polynom

$$\chi_M(x) = \begin{vmatrix} 1-x & 3 & 3 \\ -3 & -5-x & -3 \\ 3 & 3 & 1-x \end{vmatrix} = -x^3 - 3 \cdot x^2 + 4 = -(x+2)^2 \cdot (x-1)$$

und also in der Tat $\mathrm{Spec}(M) = \{-2, 1\}$. Um das charakteristische Polynom zu faktorisieren, hat man hier zwei Möglichkeiten: Entweder klammert man bei der Berechnung von χ_M von Vornherein den Faktor $x-1$ aus, oder man rät die Nullstelle $x = 1$ und wendet Polynomdivision durch $x - 1$ an.

Mit unserer Definition des charakteristischen Polynoms erhält man bei ungerader Dimension stets Leitkoeffizient -1 (vgl. den Summanden $-x^3$ im Beispiel). Manche Autoren definieren $\chi_M(x)$ als $\det(x \cdot \mathbb{1} - M)$, um ein normiertes Polynom zu erhalten. In jedem Fall kann man sich von dem folgenden Ergebnis überzeugen.

Lemma 5.1.16. *Für $M \in K^{n \times n}$ hat man* $\deg(\chi_M) = n$.

Die Theorie der Polynome liefert einen weiteren Begriff von Vielfachheit.

Definition 5.1.17. Die algebraische Vielfachheit $\alpha(\lambda)$ von $\lambda \in \mathrm{Spec}(M)$ ist die Vielfachheit von λ als Nullstelle des Polynoms χ_M (im Sinn von Definition 4.3.11).

Um algebraische und geometrische Vielfachheit in Relation zu setzen, werden wir das folgende Ergebnis benutzen. Dieses erlaubt uns auch die allgemeinere Definition eines charakteristischen Polynoms χ_φ für eine lineare Abbildung φ auf einem endlich-dimensionalen Vektorraum: Man erkärt $\chi_\varphi = \chi_M$ für eine und also jede Matrix M mit der Eigenschaft, dass φ und $\mathrm{Abb}(M)$ ähnlich sind. Mit anderen Worten hängt das charakteristische Polynom nicht von der Wahl einer Basis ab.

Proposition 5.1.18. *Für ähnliche Matrizen M und N hat man $\chi_M = \chi_N$ und damit auch* $\mathrm{Spec}(M) = \mathrm{Spec}(N)$.

Beweis. Für $N = S^{-1} \cdot M \cdot S$ und mit $x \cdot \mathbb{1} = x \cdot S^{-1} \cdot S = S^{-1} \cdot x \cdot \mathbb{1} \cdot S$ ist

$$\begin{aligned} \chi_N(x) &= \det\left(S^{-1} \cdot M \cdot S - S^{-1} \cdot x \cdot \mathbb{1} \cdot S\right) \\ &= \det\left(S^{-1}\right) \cdot \det\left(M - x \cdot \mathbb{1}\right) \cdot \det(S) = \chi_M(x). \end{aligned}$$

Mit Proposition 5.1.14 stimmen dann auch die Eigenwerte überein (wobei sich letzteres auch direkt zeigen lässt). □

Wie versprochen erhalten wir eine Beziehung zwischen den Vielfachheiten.

Proposition 5.1.19. *Für jedes $\lambda \in \mathrm{Spec}(M)$ hat man $\gamma(\lambda) \leq \alpha(\lambda)$.*

Beweis. Wir wählen eine Basis $s_1, \dots, s_{\gamma(\lambda)}$ des Eigenraums $E(\lambda)$ und ergänzen sie zu einer Basis $s_1, \dots, s_n$ des K^n (wobei $M \in K^{n\times n}$). Wegen $M \cdot s_i = \lambda \cdot s_i$ ist M gemäß Bemerkung 4.5.3 ähnlich zu einer Matrix $N = (\nu_{ij})$ mit

$$\nu_{ij} = \begin{cases} \lambda & \text{wenn } i = j, \\ 0 & \text{sonst} \end{cases}$$

für $j \leq \gamma(\lambda)$. Mit anderen Worten sind in den ersten $\gamma(\lambda)$ Spalten von N allenfalls die Diagonaleinträge λ von Null verschieden. Durch Entwicklung nach diesen Spalten erhält man damit

$$\chi_M(x) = \chi_N(x) = (\lambda - x)^{\gamma(\lambda)} \cdot p(x)$$

für ein Polynom p. Die Vielfachheit von λ als Nullstelle des charakteristischen Polynoms χ_M ist also mindestens $\gamma(\lambda)$. □

Wir erhalten nun ein weiteres Kriterium für Diagonalisierbarkeit. Die Bedingung, dass das charakteristische Polynom in Linearfaktoren zerfällt (also $\sum_{\lambda\in\mathrm{Spec}(M)} \alpha(\lambda)$ gleich $\deg(\chi_M)$ ist) ist dabei automatisch, wenn K algebraisch abgeschlossen ist.

Proposition 5.1.20. *Eine Matrix $M \in K^{n\times n}$ ist genau dann diagonalisierbar, wenn ihr charakteristisches Polynom in Linearfaktoren zerfällt und $\alpha(\lambda) = \gamma(\lambda)$ für alle $\lambda \in \mathrm{Spec}(M)$ gilt.*

Beweis. Die vorige Proposition liefert

$$\sum_{\lambda\in\mathrm{Spec}(M)} \gamma(\lambda) \leq \sum_{\lambda\in\mathrm{Spec}(M)} \alpha(\lambda) \leq n$$

mit Gleichheit genau unter den Bedingungen aus der Proposition. Die Behauptung folgt mit Proposition 5.1.7. □

Wir zeigen nun, wie Diagonalisierung fehlschlagen kann.

Beispiel 5.1.21. Die Matrix

$$M = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

ist selbst über dem algebraisch abgeschlossenen Körper $\mathbb{C}$ nicht diagonalisierbar. Es gilt nämlich

$$\chi_M(x) = (2 - x)^2$$

und also $\mathrm{Spec}(M) = \{2\}$ mit $\alpha(2) = 2$. Andererseits hat man

$$M \cdot \begin{pmatrix} x \\ y \end{pmatrix} = 2 \cdot \begin{pmatrix} x \\ y \end{pmatrix} \quad \Leftrightarrow \quad 2x + y = 2x \quad \Leftrightarrow \quad y = 0.$$

Dies ergibt $E(2) = \mathrm{Span}(e_1)$ und also $\gamma(2) = 1$.

Zum Abschluss dieses Teilkapitels untersuchen wir die Frage, wann sich mehrere Matrizen simultan diagonalisieren lassen.

Proposition 5.1.22. *Für diagonalisierbare Matrizen $M, N \in K^{n\times n}$ sind äquivalent:*

(i) Man hat $M \cdot N = N \cdot M$.

(ii) Es gibt eine invertierbare Matrix $S \in K^{n\times n}$, sodass $S^{-1}\cdot M\cdot S$ und $S^{-1}\cdot N\cdot S$ beide Diagonalform haben.

Beweis. Man kann sich überzeugen, dass Matrizen in Diagonalform stets kommutieren. Hat man (ii), so erhält man also

$$\begin{aligned} M \cdot N &= S \cdot \big(S^{-1} \cdot M \cdot S\big) \cdot \big(S^{-1} \cdot N \cdot S\big) \cdot S^{-1} \\ &= S \cdot \big(S^{-1} \cdot N \cdot S\big) \cdot \big(S^{-1} \cdot M \cdot S\big) \cdot S^{-1} = N \cdot M. \end{aligned}$$

Für die Gegenrichtung brauchen wir wegen der Diagonalisierbarkeit von M nur zu zeigen, dass für $\lambda \in \operatorname{Spec}(M)$ jeweils

$$E_M(\lambda) = \bigoplus \{U_\mu : \mu \in \operatorname{Spec}(N)\} \quad \text{mit} \quad U_\mu = E_M(\lambda) \cap E_N(\mu)$$

gilt. Ein beliebiges $w \in E_M(\lambda)$ lässt sich wegen der Diagonalisierbarkeit von N eindeutig schreiben als

$$w = \textstyle\sum_{\mu \in \operatorname{Spec}(N)} w_\mu \quad \text{mit} \quad w_\mu \in E_N(\mu).$$

Gilt nun (i), so erhält man jeweils

$$N \cdot M \cdot w_\mu = M \cdot N \cdot w_\mu = M \cdot \mu \cdot w_\mu = \mu \cdot M \cdot w_\mu$$

und also $M \cdot w_\mu \in E_N(\mu)$. In Anbetracht von

$$\textstyle\sum_{\mu \in \operatorname{Spec}(N)} M \cdot w_\mu = M \cdot w = \lambda \cdot w = \sum_{\mu \in \operatorname{Spec}(N)} \lambda \cdot w_\mu$$

liefert die Eindeutigkeit jeweils $M \cdot w_\mu = \lambda \cdot w_\mu$ und also $w_\mu \in U_\mu$. □

5.2. **Die Jordansche Normalform.** Im vorigen Teilkapitel haben wir gesehen, dass sich selbst über algebraisch abgeschlossenen Körpern nicht alle Matrizen diagonalisieren lassen. Wir untersuchen deshalb im Folgenden, inwieweit man jede quadratische Matrix in eine Form bringen kann, die etwas aber nicht allzu viel komplizierter ist als eine Diagonalmatrix.

Die Idee der folgenden Definition ist es, die Eigenräume etwas zu vergrößern. Wir rufen in Erinnerung, dass die Iterationen einer Funktion $f : M \to M$ durch $f^0 = \mathrm{id}_M$ und $f^{n+1} = f \circ f^n$ gegeben sind.

Definition 5.2.1. Wir betrachten eine lineare Abbildung $\varphi : V \to V$. Der Hauptraum (oder verallgemeinerte Eigenraum) von $\lambda \in \operatorname{Spec}(\varphi)$ ist gegeben durch

$$E^*(\lambda) = E^*_\varphi(\lambda) = \bigcup_{n \in \mathbb{N}} E^n_\varphi(\lambda) \qquad \text{für} \qquad E^n(\lambda) = \ker\big((\varphi - \lambda \cdot \mathrm{id}_V)^n\big).$$

Wir halten fest:

Korollar 5.2.2. *Es gilt $E(\lambda) = E^1(\lambda) \subseteq E^*(\lambda)$.*

Die folgenden Ergebnisse sind analog zu solchen aus dem vorigen Kapitel.

Lemma 5.2.3. *(a) Für jedes $\lambda \in \operatorname{Spec}(\varphi)$ ist $E^*(\lambda)$ ein Untervektorraum.*
(b) Aus $v \in E^*(\lambda)$ folgt $\varphi(v) \in E^*(\lambda)$.

Beweis. (a) Dies folgt aus der Tatsache, dass $E^0(\lambda) \subseteq E^1(\lambda) \subseteq \ldots$ eine Kette von wachsenden Untervektorräumen ist.

(b) Man hat

$$(\varphi - \lambda \cdot \mathrm{id}) \circ \varphi(w) = \varphi(\varphi(w)) - \lambda \cdot \varphi(w) = \varphi\big(\varphi(w) - \lambda \cdot w\big) = \varphi \circ (\varphi - \lambda \cdot \mathrm{id})(w)$$

und also induktiv $(\varphi - \lambda \cdot \mathrm{id})^n \circ \varphi = \varphi \circ (\varphi - \lambda \cdot \mathrm{id})^n$. Aus $v \in E^n(\lambda)$ folgt somit

$$(\varphi - \lambda \cdot \mathrm{id})^n(\varphi(v)) = \varphi(0) = 0$$

und daher $\varphi(v) \in E^n(\lambda)$. □

Um Proposition 5.1.3 zu verallgemeinern, zeigen wir zunächst das folgende Lemma. Es ergibt sich insbesondere, dass der Hauptraum $E^*(\lambda)$ keinen anderen Eigenraum $E(\mu)$ schneidet. Wir haben die Eigenräume also nicht über Gebühr vergrößert.

Lemma 5.2.4. *Für Eigenwerte $\lambda \neq \mu$ gilt $E^*(\lambda) \cap E^*(\mu) = \{0\}$.*

Beweis. Wir nehmen an, es gilt

$$(\varphi - \lambda \cdot \mathrm{id})^m(v) = 0 = (\varphi - \mu \cdot \mathrm{id})^n(v).$$

Per Induktion über n folgern wir $v = 0$. Hat man $n = 1$, so ergibt sich $\varphi(v) = \mu \cdot v$ und man zeigt induktiv

$$(\varphi - \lambda \cdot \mathrm{id})^m(v) = (\varphi - \lambda \cdot \mathrm{id})\big((\mu - \lambda)^{m-1} \cdot v\big) = (\mu - \lambda)^m \cdot v.$$

Ist dies gleich Null, so folgt $v = 0$ wegen $\lambda \neq \mu$. Im Induktionsschritt beachtet man

$$(\varphi - \mu \cdot \mathrm{id})^{n-1}(w) = 0 \qquad \text{für} \qquad w = (\varphi - \mu \cdot \mathrm{id})(v).$$

In Anbetracht von

$$\begin{aligned}(\varphi - \lambda \cdot \mathrm{id}) \circ (\varphi - \mu \cdot \mathrm{id})(u) &= \varphi^2(u) - \mu \cdot \varphi(u) - \lambda \cdot \varphi(u) + \lambda \cdot \mu \cdot u \\ &= (\varphi - \mu \cdot \mathrm{id}) \circ (\varphi - \lambda \cdot \mathrm{id})(u)\end{aligned}$$

ergibt sich außerdem

$$(\varphi - \lambda \cdot \mathrm{id})^m(w) = (\varphi - \mu \cdot \mathrm{id}) \circ (\varphi - \lambda \cdot \mathrm{id})^m(v) = (\varphi - \mu \cdot \mathrm{id})(0) = 0.$$

Induktiv erhält man damit $w = 0$ und dann mit dem Basisfall $v = 0$. □

Wie versprochen können wir nun das Folgende zeigen.

Proposition 5.2.5. *Für paarweise verschiedene Eigenwerte $\lambda_1, \dots, \lambda_k$ und Vektoren $v_i \in E^*(\lambda_i)\backslash\{0\}$ ist $\{v_1, \dots, v_k\}$ linear unabhängig.*

Beweis. Wir argumentieren per Induktion über k. Es gelte dafür $\sum_{i=1}^k \mu_i \cdot v_i = 0$ sowie $(\varphi - \lambda_1 \cdot \mathrm{id})^n(v_1) = 0$. Man hat dann

$$\sum_{i=2}^{k} \mu_i \cdot (\varphi - \lambda_1 \cdot \mathrm{id})^n(v_i) = 0.$$

Wegen Lemma 5.2.3(b) und Lemma 5.2.4 hat man für $i \geq 2$ jeweils

$$(\varphi - \lambda_1 \cdot \mathrm{id})^n(v_i) \in E^*(\lambda_i)\backslash\{0\}.$$

Induktiv ergibt sich also $\mu_2 = \ldots = \mu_k = 0$ sowie dann auch $\mu_1 = 0$. □

Für algebraisch abgeschlossene Körper werden wir zeigen, dass sich der gesamte Vektorraum als direkte Summe der Haupträume ergibt. Zunächst untersuchen wir die Struktur eines einzelnen Hauptraums, wobei wir uns auf den endlichdimensionalen Fall beschränken.

Lemma 5.2.6. *Sei $\varphi : V \to V$ linear. Ist $\dim(V) = n$ endlich, so gilt*

$$E^*_\varphi(\lambda) = E^n_\varphi(\lambda) \quad \textit{für alle} \quad \lambda \in \mathrm{Spec}(\varphi).$$

Beweis. Gilt $\dim(E^n(\lambda)) < n$, so hat man ein $k < n$ mit $E^k(\lambda) = E^{k+1}(\lambda)$. Für beliebiges $v \in E^*(\lambda)$ wählen wir l minimal mit $\psi^l(v) = 0$ für $\psi = \varphi - \lambda \cdot \mathrm{id}$. Man hat dann $l \leq k$, da sonst $\psi^{l-k-1}(v) \in E^{k+1}(\lambda) = E^k(\lambda)$ und also

$$\psi^{l-1}(v) = \psi^k \circ \psi^{l-k-1}(v) = 0$$

folgt. Es ergibt sich $E^*(\lambda) \subseteq E^k(\lambda) \subseteq E^n(\lambda) \subseteq E^*(\lambda)$. □

Sind φ und $\lambda \in \mathrm{Spec}(\varphi)$ aus dem Kontext klar, so schreiben wir $E^i = E^i_\varphi(\lambda)$. Mit Proposition 4.2.15 (und ihrem Beweis; vgl. auch die Dimensionsformel) haben wir jeweils $E^{i+1} \cong E^i \oplus E^{i+1}/E^i$ und für $\varphi : V \to V$ mit $\dim(V) = n$ also

$$E^*_\varphi(\lambda) \cong E^1 \oplus E^2/E^1 \oplus \ldots \oplus E^n/E^{n-1}.$$

Wir können eine Basis des Hauptraums also aus Basen der Faktoren E^{i+1}/E^i zusammensetzen (mit $E^1 \cong E^1/E^0$ wegen $E^0 = \{0\}$), wobei eine geschickte Wahl der Basen entscheidend ist. Für $v \in E^{i+1}$ hat man $(\varphi - \lambda \cdot \mathrm{id})(v) \in E^i$. Gilt $i > 0$, so haben wir also eine Funktion

$$\psi_i : E^{i+1} \to E^i/E^{i-1} \qquad \text{mit} \qquad \psi_i(v) = [\varphi(v) - \lambda \cdot v].$$

Es gilt dabei

$$\psi_i(v) = 0 \quad \Leftrightarrow \quad (\varphi - \lambda \cdot \mathrm{id})(v) \in E^{i-1} \quad \Leftrightarrow \quad v \in E^i,$$

also $\ker(\psi_i) = E_i$. Der Homomorphiesatz liefert eine injektive lineare Abbildung

$$\overline{\psi}_i : E^{i+1}/E^i \to E^i/E^{i-1} \qquad \text{mit} \qquad \overline{\psi}_i([v]) = \psi_i(v).$$

Wegen der Injektivität haben wir das folgende Ergebnis.

Lemma 5.2.7. *Ist $[v_1], \ldots, [v_k]$ eine Basis von E^{i+1}/E^i, so sind die Elemente*

$$[\varphi(v_1) - \lambda \cdot v_1], \ldots, [\varphi(v_k) - \lambda \cdot v_k] \in E^i/E^{i-1}$$

paarweise verschieden und in E^i/E^{i-1} linear unabhängig.

Wie oben angekündigt erhalten wir nun eine Basis mit guten Eigenschaften.

Proposition 5.2.8. *Sei φ eine lineare Abbildung auf einem endlich-dimensionalen Vektorraum. Für jeden Eigenwert $\lambda \in \mathrm{Spec}(\varphi)$ hat der Hauptraum $E^*_\varphi(\lambda)$ eine Basis B mit der Eigenschaft, dass für $\psi = \varphi - \lambda \cdot \mathrm{id}$ gilt:*

(a) Hat man $v \in B$, so ist $\psi(v) \in B \cup \{0\}$.
(b) Für $v, w \in B$ folgt aus $\psi(v) = \psi(w) \neq 0$ stets $v = w$.
(c) Die Menge $B \cap E_\varphi(\lambda)$ ist eine Basis des Eigenraums zu λ.

Beweis. Lemma 5.2.6 liefert $E^*_\varphi(\lambda) = E^n_\varphi(\lambda) = E^n$ für ein geeignetes $n \in \mathbb{N}$. Wähle zunächst $B_n \subseteq E^n$ so, dass $\{[v] : v \in B_n\}$ eine Basis von E^n/E^{n-1} ist. Induktiv nehmen wir an, dass $\{[v] : v \in B_{i+1}\}$ mit $0 < i < n$ eine Basis von E^{i+1}/E^i ist. Das vorige Lemma liefert uns eine Basis $\{[v] : v \in B_i\}$ von E^i/E^{i-1} mit $\psi(v) \in B_i$ für alle $v \in B_{i+1}$. Wir setzen $B = \bigcup_{i=1}^n B_i$ und bemerken, dass diese Vereinigung disjunkt ist: Hat man $i < j$ und $w \in B_i \subseteq E^{j-1}$, so kann $[w] = 0 \in E^j \backslash E^{j-1}$ kein Element der Basis $\{[v] : v \in B_j\}$ sein. Es folgt

$$|B| = \sum_{i=1}^n |B_i| = \sum_{i=1}^n \dim\left(E^i/E^{i-1}\right) = \dim\left(E^*_\varphi(\lambda)\right),$$

wobei die letzte Gleichheit auf dem Absatz nach Lemma 5.2.6 beruht. Wir zeigen, dass $B \subseteq E^*_\varphi(\lambda)$ linear unabhängig und also eine Basis ist. Man betrachte dazu eine Linearkombination

$$0 = \sum_{v \in B} \lambda_v \cdot v = \sum_{i=1}^n \sum_{v \in B_i} \lambda_v \cdot v.$$

Indem wir Äquivalenzklassen in E^n/E^{n-1} nehmen (wo wie oben $[v] = 0$ für alle $v \in B_i$ mit $i < n$ gilt), erhalten wir

$$\sum_{v \in B_n} \lambda_v \cdot [v] = 0.$$

Es folgt $\lambda_v = 0$ für alle $v \in B_n$, weil $\{[v] : v \in B_n\}$ als Basis von E^n/E^{n-1} gewählt war. Nimmt man als nächstes Äquivalenzklassen in E^{n-1}/E^{n-2}, so erhält man auch $\lambda_v = 0$ für alle $v \in B_{n-1}$ und induktiv für alle $v \in B$.

Eigenschaft (a) ist erfüllt, da aus $v \in B_{i+1}$ mit $i > 0$ per Konstruktion $\psi(v) \in B_i$ sowie für $v \in B_1 \subseteq E^1 = E_\varphi(\lambda)$ stets $\psi(v) = 0$ gilt. Für (b) betrachten wir $v \in B_{i+1}$ und $w \in B_{j+1}$ mit $\psi(v) = \psi(w) \in B_i \cap B_j$. Wie oben gesehen gilt dann $i = j$ und mit dem vorigen Lemma also $v = w$. Schließlich bemerken wir für (c), dass $\{[v] : v \in B_1\}$ eine Basis von $E^1/E^0 \cong E^1$ und also $B \cap E_\varphi(\lambda) = B_1$ eine Basis des Eigenraums $E^1 = E_\varphi(\lambda)$ ist. □

Wir können unsere Basis nun in Ketten zerlegen, die in Eigenvektoren wurzeln.

Definition 5.2.9. Sei B eine Basis wie in der vorigen Proposition. Für jeden Eigenvektor $v \in E_\varphi(\lambda) \cap B$ definieren wir

$$J^B(v) = \big\{w \in E^*_\varphi(\lambda) : \psi^i(w) = v \text{ für ein } i \in \mathbb{N}\big\} \quad \text{mit} \quad \psi = \varphi - \lambda \cdot \mathrm{id}\,.$$

Die Mengen $J^B(v)$ nennt man Jordan-Ketten. Es ist dringend zu beachten, dass unsere Definition nur anwendbar ist, wenn bereits eine geeignete Basis gegeben ist. Startet man mit einer beliebigen Basis des Eigenraums und wählt dann beliebige Ketten über den Basisvektoren, so erhält man im Allgemeinen keine Basis des Hauptraums, weil die Ketten in 'Sackgassen' geraten können (vgl. die Bemerkung am Ende von Beispiel 5.2.13).

Lemma 5.2.10. *Unter den Voraussetzungen der vorigen Proposition gilt:*

(a) Für jedes $v \in E_\varphi(\lambda) \cap B$ gibt es ein $w \in J^B(v)$ und ein $k \in \mathbb{N}$ mit $\psi^k(w) = v$ und $J^B(v) = \{\psi^0(w), \dots, \psi^k(w)\}$.

(b) Man erhält $B = \bigcup_{v \in E(\lambda) \cap B} J^B(v)$ als disjunkte Vereinigung.

Beweis. (a) Setze $v_0 = v$ und definiere induktiv $v_{i+1} \in B$ als den eindeutigen Vektor mit $\psi(v_{i+1}) = v_i$. Dies ist nur bis zu einer endlichen Schranke $i \leq k$ möglich: Wäre nämlich $v_i = v_j$ mit $i < j$, so hätte man $\psi^{j-i}(v_j) = v_i = v_j$ und also $\psi^l(v_j) \neq 0$ für alle $l \in \mathbb{N}$. Man setzt $w = v_k$ (also $\psi^i(w) = v_{k-i}$) und verifiziert die Behauptung.

(b) Für jedes $w \in B \subseteq E^*(\lambda)$ hat man ein kleinstes $k \in \mathbb{N}$ mit $\psi^{k+1}(w) = 0$ und also $\psi^k(w) =: v \in E(\lambda) \cap B$, was $w \in J^B(v)$ ergibt. Schneiden sich zwei Jordan-Ketten, so enden sie im selben Eigenvektor. □

Die Untervektorräume $\mathrm{Span}(J^B(v))$ heißen zyklisch. Man möge sich überzeugen, dass sie unter ψ und damit auch unter $\varphi = \psi + \lambda \cdot \mathrm{id}$ abgeschlossen sind. Wir diskutieren die Normalform der darstellenden Matrix zunächst für den Fall, dass der gesamte Vektorraum ein solch zyklischer Raum ist.

Lemma 5.2.11. *Wir betrachten eine Matrix $M \in K^{d \times d}$ und setzen $P = M - \lambda \cdot \mathbb{1}$. Gibt es einen Vektor $w \in K^d$ mit $P^d \cdot w = 0$ und*

$$K^d = \mathrm{Span}\left(\left\{P^0 \cdot w, \dots, P^{d-1} \cdot w\right\}\right),$$

so ist M ähnlich zur Matrix

$$\lambda \cdot \mathbb{1} + N_d \quad \textit{mit} \quad N_d = (\nu_{ij})_{1 \le i,j \le d} \textit{ für } \nu_{ij} = \begin{cases} 1 & \textit{wenn } j = i+1, \\ 0 & \textit{sonst.} \end{cases}$$

Beweis. Für jeden Basisvektor $P^i \cdot w$ mit $i < d$ hat man

$$M \cdot P^i \cdot w = (\lambda \cdot \mathbb{1} + P) \cdot P^i \cdot w = \lambda \cdot P^i \cdot w + P^{i+1} \cdot w$$

und insbesondere $M \cdot P^{d-1} \cdot w = \lambda \cdot P^{d-1} \cdot w$. Hat S Spalten $P^{d-1} \cdot w, \ldots, P^0 \cdot w$, so hat man also $S^{-1} \cdot M \cdot S = \lambda \cdot \mathbb{1} + N_d$ gemäß Bemerkung 4.5.3. □

Im nächsten Schritt verallgemeinern wir auf den Fall eines einzelnen Hauptraums, der aber in mehrere zyklische Unterräume zerfallen mag.

Korollar 5.2.12. *Jede Matrix $M \in K^{n \times n}$ mit nur einem Hauptraum $E_M^*(\lambda) = K^n$ ist ähnlich zu einer Matrix der Form $J_1 \oplus \cdots \oplus J_m$ mit $J_i = \lambda \cdot \mathbb{1} + N_{d(i)}$.*

Beweis. Wir wählen eine Basis B wie in Proposition 5.2.8 und zerlegen sie wie in Lemma 5.2.10 in Jordan-Ketten $J^B(v_i)$, wobei man also $E(\lambda) \cap B = \{v_1, \ldots, v_m\}$ hat. Für die zyklischen Unterräume $Z_i = \mathrm{Span}(J^B(v_i))$ ergibt sich dann

$$E_M^*(\lambda) = Z_1 \oplus^{K^n} \ldots \oplus^{K^n} Z_m$$

als direkte Summe. Gemäß der Bemerkung vor Lemma 5.2.11 folgt aus $w \in Z_i$ jeweils $M \cdot w \in Z_i$. Mit Proposition 4.4.15 und den Ergebnissen aus Abschnitt 4.5 zerfällt M in Blöcke J_i, die laut dem vorigen Lemma die gegebene Form haben (wobei jeweils $d(i) = \dim(Z_i) = |J^B(v_i)|$ gilt). □

Es ist höchste Zeit, einen konkreten Fall zu diskutieren.

Beispiel 5.2.13. Man betrachte die Matrix

$$M = \begin{pmatrix} 3 & -1 & -1 & 0 \\ 2 & 0 & 0 & -1 \\ -1 & 1 & 1 & 1 \\ -2 & 2 & -2 & 4 \end{pmatrix} \in \mathbb{R}^{4 \times 4}.$$

Das charakteristische Polynom kann man geschickt berechnen, indem man zunächst die zweite zur ersten Spalte addiert. Fährt man auf ähnliche Weise fort, erhält man

$$\chi_M(x) = \left| \begin{pmatrix} 3-x & -1 & -1 & 0 \\ 2 & -x & 0 & -1 \\ -1 & 1 & 1-x & 1 \\ -2 & 2 & -2 & 4-x \end{pmatrix} \right| = \left| \begin{pmatrix} 2-x & -1 & -1 & 0 \\ 2-x & -x & 0 & -1 \\ 0 & 1 & 1-x & 1 \\ 0 & 2 & -2 & 4-x \end{pmatrix} \right|$$

$$= (2-x) \cdot \left| \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & -x & 0 & -1 \\ 0 & 1 & 1-x & 1 \\ 0 & 2 & -2 & 4-x \end{pmatrix} \right| = \ldots = (2-x)^4,$$

womit es nur den einen Hauptraum $E_M^*(2)$ gibt. Durch Lösen von $(M - 2 \cdot \mathbb{1}) \cdot v = 0$ bestimmt man den Eigenraum

$$E^1 = E_M^1(2) = \mathrm{Span}(v_1, v_2) \quad \text{mit} \quad v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \text{ und } v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ -2 \end{pmatrix}.$$

Weiter rechnet man $(M-2\cdot\mathbb{1})^2=0$ und erhält $E^2=E^2_M(2)=E^*_M(2)=\mathbb{R}^4$. Gemäß dem Beweis von Proposition 5.2.8 suchen wir eine Basis $[w_1],[w_2]$ von E^2/E^1. Äquivalent möchte man zu einer Basis v_1,v_2,w_1,w_2 des $\mathbb{R}^4$ ergänzen. Man kann hier die Einheitsvektoren $w_1=e_1$ und $w_2=e_3$ wählen. Immer noch mit dem Beweis von Proposition 5.2.8 betrachten wir dann die Basis

$$B=\{(M-2\cdot\mathbb{1})\cdot e_1,e_1,(M-2\cdot\mathbb{1})\cdot e_3,e_3\},$$

wobei die ersten beiden und die letzten beiden Vektoren jeweils eine Jordan-Kette bilden. Die Vektoren aus B bilden die Spalten der folgenden Matrix S, die wir samt ihrer Inversen angeben als

$$S=\begin{pmatrix}1&1&-1&0\\2&0&0&0\\-1&0&-1&1\\-2&0&-2&0\end{pmatrix}\quad\text{und}\quad S^{-1}=\frac{1}{2}\cdot\begin{pmatrix}0&1&0&0\\2&-2&0&-1\\0&-1&0&-1\\0&0&2&-1\end{pmatrix}.$$

Gemäß dem Beweis von Korollar 5.2.12 ergibt sich

$$S^{-1}\cdot M\cdot S=\begin{pmatrix}2&1&0&0\\0&2&0&0\\0&0&2&1\\0&0&0&2\end{pmatrix}.$$

Man beachte, dass $(M-2\cdot\mathbb{1})\cdot e_1$ und $(M-2\cdot\mathbb{1})\cdot e_3$ eine Basis des Eigenraums bilden, die sich von unserer Ausgangsbasis aus v_1 und v_2 unterscheidet. Dies ist essenziell, da die Gleichung $(M-2\cdot\mathbb{1})\cdot u=v_2$ keine Lösung u hat.

Um zu erklären, warum die Struktur der Jordan-Blöcke $J=\lambda\cdot\mathbb{1}+N_d$ rechnerisch vorteilhaft ist, bemerken wir zunächst, dass die Summanden kommutieren, dass man also $\lambda\cdot\mathbb{1}\cdot N_d=N_d\cdot\lambda\cdot\mathbb{1}$ hat. Dies erlaubt es uns, J^n wie im folgenden Beispiel mit dem binomischen Lehrsatz

$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}\cdot x^{n-i}\cdot y^i\quad\text{für}\quad\binom{n}{i}=\frac{n!}{i!\cdot(n-i)!}$$

zu berechnen (welchen man per Induktion über n prüfen möge). Weiter ist dabei relevant, dass die Matrix N_d nilpotent ist, es also ein j mit ${N_d}^j=0$ gibt (nämlich $j=d$ wegen $N_d\cdot e_1=0$ und $N_d\cdot e_{i+1}=e_i$).

Beispiel 5.2.14. Die Matrix $J=\lambda\cdot\mathbb{1}+N_3\in\mathbb{R}^{3\times3}$ hat wegen ${N_3}^i=0$ für $i\geq3$ die Potenzen

$$J^n=\sum_{i=0}^{2}\binom{n}{i}\cdot\lambda^{n-i}\cdot\mathbb{1}\cdot{N_3}^i=\lambda^n\cdot\mathbb{1}+n\cdot\lambda^{n-1}\cdot N_3+\frac{n\cdot(n-1)}{2}\cdot\lambda^{n-2}\cdot{N_3}^2.$$

Es gilt also

$$\begin{pmatrix}\lambda&1&0\\0&\lambda&1\\0&0&\lambda\end{pmatrix}^n=\begin{pmatrix}\lambda^n&n\cdot\lambda^{n-1}&\frac{n}{2}\cdot(n-1)\cdot\lambda^{n-2}\\0&\lambda^n&n\cdot\lambda^{n-1}\\0&0&\lambda^n\end{pmatrix}.$$

Ist M ähnlich zu J vermöge $S^{-1}\cdot M\cdot S=J$, so rechnet man

$$M^n=\big(S\cdot J\cdot S^{-1}\big)^n=S\cdot J^n\cdot S^{-1},$$

weil die Produkte $S^{-1}\cdot S$ jeweils wegfallen.

Wir kommen nun zum allgemeinen Fall, in dem es mehrere Eigenwerte gibt.

Theorem 5.2.15. *Für eine lineare Abbildung $\varphi : V \to V$ auf einem endlichdimensionalen Vektorraum V sind äquivalent:*

(i) Das charakteristische Polynom von φ zerfällt in Linearfaktoren.
(ii) Der Raum V ergibt sich als direkte Summe

$$V = \bigoplus^V \left\{ E^*_\varphi(\lambda) : \lambda \in \mathrm{Spec}(\varphi) \right\}.$$

Insbesondere gilt (b) im algebraisch abgeschlossenen Fall.

Beweis. Gilt (b), so ist φ laut Korollar 5.2.12 ähnlich zu einer Abbildung $\mathrm{Abb}(M)$ für eine Summe $M = J_1 \oplus \ldots \oplus J_k$ von Jordan-Blöcken $J_i = \lambda_i \cdot \mathbb{1} + N_{d(i)}$. Das charakteristische Polynom ergibt sich mit Proposition 5.1.18 und Beispiel 3.2.6 (Determinante von Dreiecksmatrizen) als

$$\chi_\varphi(x) = \chi_M(x) = \prod_{i=1}^{k} (\lambda_i - x)^{d(i)},$$

sodass (a) erfüllt ist.

In der umgekehrten Richtung wissen wir aus Proposition 5.2.5, dass die Summe direkt ist. Zeigen müssen wir lediglich, dass sie ganz V aufspannt. Mit (a) ergibt sich $\dim(V)$ als Summe der algebraischen Vielfachheiten der Eigenwerte. Es genügt also, wenn wir

$$\dim\left(E^*(\lambda)\right) = \alpha(\lambda)$$

für alle $\lambda \in \mathrm{Spec}(\varphi)$ zeigen (vgl. Proposition 5.1.20). Schreiben wir $n = \dim(V)$ und $\psi = \varphi - \lambda \cdot \mathrm{id}$, so gilt $E^*(\lambda) = \ker(\psi^N)$ laut Lemma 5.2.6 für alle $N \geq n$. Wir zeigen, dass

$$V = \ker\left(\psi^n\right) \oplus^V \mathrm{img}\left(\psi^n\right)$$

eine direkte Summe ist. Wegen dem Rangsatz ist nur nachzuweisen, dass wir $v = 0$ für alle $v \in \ker(\psi^n) \cap \mathrm{img}(\psi^n)$ haben. Wir schreiben $v = \psi^n(w)$ und erhalten dann $0 = \psi^n(v) = \psi^{2n}(w)$, also $w \in E^*(\lambda) = \ker(\psi^n)$ und damit $v = 0$ wie gewünscht. Nun ist $\mathrm{img}(\psi^n)$ unter φ abgeschlossen: Wie im Beweis von Lemma 5.2.4 kommutieren nämlich φ und ψ, sodass aus $v = \psi^n(w) \in \mathrm{img}(\psi^n)$ auch

$$\varphi(v) = \psi^n(\varphi(w)) \in \mathrm{img}(\psi^n)$$

folgt. Bis auf Ähnlichkeit hat also φ eine darstellende Matrix $M \oplus N$, wobei der Summand $M = J_1 \oplus \ldots \oplus J_m$ wegen Korollar 5.2.12 in Jordan-Blöcke zu unserem fixen Eigenwert λ zerfällt. Mit $M \in K^{d \times d}$ für $d = \dim(E^*(\lambda))$ ergibt sich (durch Entwicklung nach den ersten d Zeilen) das charakteristische Polynom

$$\chi_\varphi(x) = \chi_M(x) \cdot \chi_N(x) = (\lambda - x)^d \cdot \chi_N(x).$$

Hierbei ist λ keine Nullstelle von χ_N. Sonst wäre nämlich λ ein Eigenwert von N und man hätte einen Vektor $v \in \mathrm{img}(\psi^n) \backslash \{0\}$ mit $\psi(v) = 0$ und also $v \in \ker(\psi^n)$. Es gilt also in der Tat $\dim(E^*(\lambda)) = d = \alpha(\lambda)$. □

Mit Lemma 5.2.3(b) und Korollar 5.2.12 ergibt sich das folgende Ergebnis.

Korollar 5.2.16 (Satz über die Jordan-Normalform)**.** *Zerfällt das charakteristische Polynom von $M \in K^{n \times n}$ in Linearfaktoren, so ist M ähnlich zu einer Matrix $J_1 \oplus \ldots \oplus J_k$ für Jordan-Blöcke $J_i = \lambda_i \cdot \mathbb{1} + N_{d(i)}$ mit* $\mathrm{Spec}(M) = \{\lambda_i : 1 \leq i \leq k\}$.

Man beachte, dass die Jordan-Normalform mehrere Blöcke zum selben Eigenwert enthalten mag, sodass im Korollar also $k > |\operatorname{Spec}(M)|$ gelten kann.

Bemerkung 5.2.17. Die Jordan-Normalform einer linearen Abbildung ist im Wesentlichen – nämlich bis auf die Reihenfolge der Jordan-Blöcke – eindeutig bestimmt und von der darstellenden Matrix unabhängig. Dies gilt nämlich zunächst für die Vielfachheit $\dim(E^*(\lambda)) = \alpha(\lambda)$, mit welcher der Eigenwert λ auf der Diagonalen auftritt (vgl. Proposition 5.1.18). Die Dimensionen der Jordan-Blöcke zu einem festen Eigenwert ergeben sich dann aus den Dimensionen der zugehörigen Quotienten E^{i+1}/E^i (vgl. den Abschnitt nach Lemma 5.2.6), welche nicht von der Wahl einer Basis abhängen (die also nur die spezifischen Jordan-Ketten aber nicht deren Längen beeinflusst). Es folgt, dass zwei Matrizen genau dann ähnlich sind, wenn sie dieselbe Jordan-Normalform haben.

Wir besprechen auch eine wichtige Abschwächung der Jordan-Normalform:

Bemerkung 5.2.18. Eine Matrix $M \in K^{N\times N}$ heißt trigonalisierbar, wenn sie ähnlich ist zu einer Matrix (μ_{ij}) in oberer Dreiecksform, wobei also $\mu_{ij} = 0$ für $j < i$ gilt. Dies ist genau dann der Fall, wenn χ_M in Linearfaktoren zerfällt. Man folgert dies aus Theorem 5.2.15 und seinem Beweis, in dem wir für die Richtung von (b) nach (a) nur Dreiecksform und nicht die stärkere Jordan-Normalform verwendet haben. Tatsächlich lassen sich Matrizen meist deutlich leichter trigonalisieren als in Jordan-Normalform bringen, nämlich durch das Gaußsche Eliminationsverfahren. Letzteres kann auch verwendet werden, um eine Matrix – gegebenenfalls nach Permutation von Zeilen oder Spalten – als Produkt $L \cdot R$ einer unteren und oberen Dreiecksmatrix zu schreiben (LR-Zerlegung oder englisch LU decomposition; siehe auch Proposition 5.4.5).

Als klassische Anwendung diskutieren wir, wie man mit der Jordan-Normalform eine gewöhnliche Differentialgleichung höherer Ordnung lösen kann (wie sie in der Physik etwa im Zusammenhang mit gedämpften Schwingungen relevant ist).[97]

Beispiel 5.2.19. Wir suchen nach einer Funktion $y : \mathbb{R} \to \mathbb{R}$, die zusammen mit ihren Ableitungen die Gleichung

$$y''' - 5 \cdot y'' + 8 \cdot y' - 4 \cdot y = 0$$

erfüllt. Hierzu substituieren wir $x_1 = y$ und $x_2 = y'$ sowie $x_3 = y''$. Dies führt auf das Gleichungssystem

$$x_1' = x_2 \quad \text{und} \quad x_2' = x_3 \quad \text{und} \quad x_3' = 4 \cdot x_1 - 8 \cdot x_2 + 5 \cdot x_3$$

mit Matrixdarstellung

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = M \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{mit} \quad M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -8 & 5 \end{pmatrix}.$$

Zur Bestimmung der Jordan-Normalform berechnen wir

$$\chi_M(x) = (x-2)^2 \cdot (x-1).$$

[97] Der Autor stellt in seinen Prüfungen zur linearen Algebra keine Fragen zu Differentialgleichungen. Hingegen ist die Berechnung der Jordan-Normalform selbstverständlich Prüfungsstoff.

Die Eigenräume ergeben sich als

$$E_M(2) = \mathrm{Span}(v_1) \text{ mit } v_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \quad \text{und} \quad E_M(1) = \mathrm{Span}(v_3) \text{ mit } v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Weiter bestimmen wir die Lösung

$$(M - 2 \cdot \mathbb{1}) \cdot v_2 = v_1 \quad \text{für} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix}.$$

Die Jordan-Normalform ergibt sich als

$$M = S^{-1} \cdot J \cdot S \quad \text{mit} \quad J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{und} \quad S = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 4 & 4 & 1 \end{pmatrix}.$$

Wegen der Linearität der Ableitung ist unser System von Differentialgleichungen äquivalent zu

$$J \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = S \cdot M \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = S \cdot \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} \quad \text{für} \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = S \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Wir haben es somit zu tun mit den Gleichungen

$$2 \cdot y_1 + y_2 = y_1' \quad \text{und} \quad 2 \cdot y_2 = y_2' \quad \text{und} \quad y_3 = y_3'.$$

Diese lassen sich lösen durch

$$y_1(t) = t \cdot e^{2t} \quad \text{und} \quad y_2(t) = e^{2t} \quad \text{und} \quad y_3(t) = e^t,$$

wobei allenfalls die Suche nach y_1 einiger Überlegung bedarf. Indem wir mit S^{-1} multiplizieren, erhalten wir insbesondere

$$y(t) = x_1(t) = (-3t + 4) \cdot e^{2t} - e^t.$$

Man möge prüfen, dass dies tatsächlich die Differentialgleichung vom Beginn des vorliegenden Beispiels löst.

Das folgende Ergebnis verallgemeinert eine Beobachtung, die wir im Abschnitt vor Beispiel 5.2.14 gemacht haben.

Proposition 5.2.20. *Ist K algebraisch abgeschlossen, so lässt sich $M \in K^{n \times n}$ stets als Summe $M = D + N$ einer diagonalisierbaren Matrix D und einer nilpotenten Matrix N mit $D \cdot N = N \cdot D$ schreiben.*

Beweis. Der Satz über die Jordan-Normalform liefert eine invertierbares S mit

$$S^{-1} \cdot M \cdot S = (\lambda_1 \cdot \mathbb{1} + N_{d(1)}) \oplus \ldots \oplus (\lambda_k \cdot \mathbb{1} + N_{d(k)}).$$

Wir setzen

$$D = S \cdot (\lambda_1 \cdot \mathbb{1} \oplus \ldots \oplus \lambda_k \cdot \mathbb{1}) \cdot S^{-1} \quad \text{und} \quad N = S \cdot (N_{d(1)} \oplus \ldots \oplus N_{d(k)}) \cdot S^{-1}.$$

Man überzeuge sich, dass die Eigenschaften aus der Proposition erfüllt sind. □

Da sich Matrizen in $K^{n \times n}$ untereinander und mit Skalaren aus K multiplizieren lassen, können wir sie in Polynome über K einsetzen. Dies lässt sich auch so verstehen, dass wir jeden Koeffizienten $\gamma \in K$ eines Polynoms mit der Matrix $\lambda \cdot \mathbb{1} \in K^{n \times n}$ identifizieren.

Theorem 5.2.21 (Satz von Cayley-Hamilton[98])**.** *Ist K algebraisch abgeschlossen, so hat man $\chi_M(M) = 0$ für jede Matrix $M \in K^{n\times n}$.*

Beweis. Schreibe $M = S^{-1} \cdot J \cdot S$ mit J in Jordan-Normalform. Man hat $\chi_M = \chi_J$ und somit

$$\chi_M(M) = \chi_J\big(S^{-1} \cdot J \cdot S\big) = S^{-1} \cdot \chi_J(J) \cdot S.$$

Es genügt also, $\chi_J(J) = 0$ zu zeigen, was sich wiederum auf $\chi_{J_i}(J_i) = 0$ für die einzelnen Jordan-Blöcke $J_i = \lambda_i \cdot \mathbb{1} + N_{d(i)}$ reduzieren lässt. Hier ergibt sich

$$\chi_{J_i}\big(J_i\big) = \big(\lambda_i \cdot \mathbb{1} - J_i\big)^{d(i)} = \big(-N_{d(i)}\big)^{d(i)} = 0,$$

wobei wir die letzte Gleichheit aus dem Paragraph vor Beispiel 5.2.14 nehmen. □

Der Satz hat neben theoretischen auch rechnerische Implikationen:

Bemerkung 5.2.22. Wir betrachten eine Matrix $M \in K^{n\times n}$ und schreiben

$$0 = \chi_M(M) = a_n \cdot M^n + \ldots + a_0 \cdot M^0.$$

Ist M invertierbar, so multiplizieren wir mit M^{-1} und erhalten

$$a_n \cdot M^{n-1} + \ldots + a_0 \cdot M^{-1} = 0.$$

In Anbetracht von $a_0 = \chi_M(0) = \det(M) \neq 0$ ergibt sich

$$M^{-1} = -\frac{1}{a_0} \cdot \big(a_n \cdot M^{n-1} + \ldots + a_1 \cdot M^0\big).$$

Wir haben also einen neuen Weg, inverse Matrizen zu berechnen. Für große $N \in \mathbb{N}$ erhalten wir durch Polynomdivision weiter

$$x^N = \chi_M(x) \cdot q(x) + r(x) \quad \text{mit} \quad \deg(r) < \deg(\chi_M) = n.$$

Es ergibt sich

$$M^N = \chi_M(M) \cdot q(M) + r(M) = r(M),$$

wobei wir auf der rechten Seite nur M^i für $i < n$ berechnen müssen.

Zum Abschluss dieses Kapitels besprechen wir kurz eine wichtige Variante des charakteristischen Polynoms:

Bemerkung 5.2.23. Für eine Matrix $M \in K^{n\times n}$ betrachten wir das Ideal

$$I_M = \{p \in K[x] : p(M) = 0\} \subseteq K[x]$$

mit $I_M \neq \emptyset$ nach Cayley-Hamilton. Das Polynom $\mu_M = \sum_{i\leq d} a_i \cdot x^i \in I_M \backslash \{0\}$ mit kleinstmöglichem Grad $d \in \mathbb{N}$ und Leitkoeffizient $a_d = 1$ heißt Minimalpolynom von M. Es gilt $I_M = (\mu_M)$. Für beliebiges $p \in I_M$ erhalten wir durch Polynomdivision nämlich $p = \mu_M \cdot q + r$ mit $\deg(r) < \deg(\mu_M)$, wobei man dann $r \in I_M$ und wegen der Minimalität von μ_M also $r = 0$ erhält. Insbesondere ist μ_M ein Teiler des charakteristischen Polynoms χ_M. Man kann zeigen, dass das Minimalpolynom einer Matrix M mit paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_m$ die Form

$$\mu_M(x) = (x - \lambda_1)^{d(1)} \cdot \ldots \cdot (x - \lambda_m)^{d(m)}$$

hat, wobei $d(i)$ die Dimension des größten Jordan-Blocks zum Eigenwert λ_i ist. Auch wenn μ_M im Vergleich mit χ_M manche Vorzüge hat, werden wir das Minimalpolynom in dieser Vorlesung nicht weiter brauchen.

[98]Dieses Ergebnis ist benannt nach Arthur Cayley und William Hamilton. Es gilt auch ohne die Voraussetzung, dass K algebraisch abgeschlossen ist, weil jeder Körper algebraischen Abschluss hat, was wir in dieser Vorlesung aber nicht zeigen.

5.3. **Intermezzo: Rund um das Skalarprodukt.** In diesem Abschnitt untersuchen wir Skalarprodukte (synonym: innere Produkte) und die geometrische Struktur, die sie auf reellen und komplexen Vektorräumen induzieren.

Die folgende Definition involviert die Konjugation auf den komplexen Zahlen (vgl. Definition 4.3.24), die im reellen Fall wegen $\overline{x} = x$ für $x \in \mathbb{R}$ entfällt.

Definition 5.3.1. Wir betrachten einen K-Vektorraum V mit $K = \mathbb{R}$ oder $K = \mathbb{C}$. Eine Abbildung $\langle \cdot, \cdot \rangle : V^2 \to K$ heißt Skalarprodukt, wenn gilt:

(Hermitesche[99]Symmetrie) $\langle v, w \rangle = \overline{\langle w, v \rangle}$,

(Sesquilinearität) $\langle \lambda \cdot v, w \rangle = \lambda \cdot \langle v, w \rangle$ und $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$,

(Positive Definitheit)
$$\langle v, v \rangle \in \{x \in \mathbb{R} : x \geq 0\} \quad \text{und} \quad \langle v, v \rangle = 0 \text{ nur für } v = 0.$$

Man nennt V zusammen mit dem Skalarprodukt einen Innenproduktraum. Im reellen beziehungsweise komplexen Fall spricht man auch von einem euklidischen beziehungsweise unitären Vektorraum.

Bezüglich der Sesquilinearität beachte man die abgeleiteten Eigenschaften

$$\langle v, \lambda \cdot w \rangle = \overline{\langle \lambda \cdot w, v \rangle} = \overline{\lambda \cdot \langle w, v \rangle} = \overline{\lambda} \cdot \overline{\langle w, v \rangle} = \overline{\lambda} \cdot \langle v, w \rangle,$$
$$\langle u, v + w \rangle = \overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle + \langle w, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} = \langle u, v \rangle + \langle u, w \rangle.$$

Als Motivation für die komplexe Konjugation mag gelten, dass sich $\langle v, v \rangle = \overline{\langle v, v \rangle}$ stets als reelle Zahl erweist. Da man nur auf $\mathbb{R}$ und nicht auf $\mathbb{C}$ eine Ordnung hat (und also von positiven Elementen sprechen kann), ist dies entscheidend für die positive Definitheit. Mit Blick auf letztere sei angemerkt, dass die Sesquilinearität schon $\langle 0, v \rangle = 0 = \langle v, 0 \rangle$ garantiert.

Beispiel 5.3.2. Das Standard-Skalarprodukt auf dem $\mathbb{C}$-Vektorraum $\mathbb{C}^n$ ist gegeben durch

$$\left\langle \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \right\rangle = \sum_{i=1}^{n} w_i \cdot \overline{z_i}.$$

Hermitesche Symmetrie und Sesquilinearität kann man direkt nachrechnen, während sich die Definitheit aus Proposition 4.3.26 ergibt. Das Standard-Skalarprodukt auf dem $\mathbb{R}$-Vektorraum $\mathbb{R}^n$ ist genauso definiert, wobei die Konjugation entfällt. Ein anderes wichtiges Beispiel betrifft den $\mathbb{R}$-Vektorraum der stetigen Funktionen von $[0, 1]$ nach $\mathbb{R}$. Mit ein wenig Analysis sieht man, dass durch

$$\langle f, g \rangle = \int_0^1 f(x) \cdot g(x)\, dx$$

ein Skalarprodukt gegeben ist.

Das folgende Ergebnis ist von fundamentaler Bedeutung.

Proposition 5.3.3 (Cauchy-Schwarzsche Ungleichung)**.** *Man hat*

$$|\langle v, w \rangle|^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle.$$

[99]Dieser Begriff ist benannt nach Charles Hermite. Im reellen Fall spricht man schlicht von Symmetrie sowie von Linearität statt Sesquilinearität.

Beweis. Wir nehmen zunächst an, dass $\langle v, w\rangle \in \mathbb{R}$ und also $\langle w, v\rangle = \langle v, w\rangle$ gilt. Für jedes $\lambda \in \mathbb{R}$ hat man hier

$$0 \leq \langle v + \lambda \cdot w, v + \lambda \cdot w\rangle = \langle v, v\rangle + 2\lambda \cdot \langle v, w\rangle + \lambda^2 \cdot \langle w, w\rangle.$$

Gilt $w = 0$, so ist die Aussage unmittelbar. Anderenfalls erhalten wir

$$0 \leq \langle v, v\rangle + 2 \cdot \left(-\frac{\langle v, w\rangle}{\langle w, w\rangle}\right) \cdot \langle v, w\rangle + \left(\frac{\langle v, w\rangle}{\langle w, w\rangle}\right)^2 \cdot \langle w, w\rangle = \langle v, v\rangle - \frac{\langle v, w\rangle^2}{\langle w, w\rangle}.$$

Die Behauptung folgt per Multiplikation mit $\langle w, w\rangle$.

Wir leiten das Ergebnis nun für den Fall ab, dass $\langle v, w\rangle \notin \mathbb{R}$ gilt. Man beachte

$$\langle \mu \cdot v, w\rangle = \overline{\langle v, w\rangle} \cdot \langle v, w\rangle = |\langle v, w\rangle|^2 \in \mathbb{R} \quad \text{für} \quad \mu = \langle w, v\rangle.$$

Mit dem zunächst behandelten Fall ergibt sich

$$\begin{aligned} |\mu|^2 \cdot |\langle v, w\rangle|^2 = |\langle \mu \dot{v}, w\rangle|^2 &\leq \langle \mu \cdot v, \mu \cdot v\rangle \cdot \langle w, w\rangle \\ &= \mu \cdot \overline{\mu} \cdot \langle v, v\rangle \cdot \langle w, w\rangle = |\mu|^2 \cdot \langle v, v\rangle \cdot \langle w, w\rangle. \end{aligned}$$

Die Behauptung folgt per Division durch $|\mu|^2 \neq 0$. □

Die Cauchy-Schwarzsche Ungleichung hat insbesondere die folgende Anwendung.

Korollar 5.3.4. *Ist $\langle \cdot, \cdot\rangle : V^2 \to K$ ein Skalarprodukt, wird durch $\|v\| = \sqrt{\langle v, v\rangle}$ eine Norm induziert (das heißt, es gelten die Eigenschaften aus Proposition 4.3.26).*

Beweis. Die positive Definitheit war für das Skalarprodukt explizit gefordert. Für die Homogenität beachten wir $\lambda \cdot \overline{\lambda} = |\lambda|^2$ und also

$$\|\lambda \cdot v\| = \sqrt{\langle \lambda \cdot v, \lambda \cdot v\rangle} = \sqrt{\lambda \cdot \overline{\lambda} \cdot \langle v, v\rangle} = |\lambda| \cdot \|v\|.$$

Die Dreiecksungleichung zeigt man analog zum Beweis von Proposition 4.3.26, wobei die dort verwendete Gleichung $|z \cdot \overline{w}| = |z| \cdot |w|$ ersetzt wird durch die Ungleichung $|\langle v, w\rangle| \leq \|v\| \cdot \|w\|$, die modulo Wurzelziehen mit der Cauchy-Schwarzschen Ungleichung übereinstimmt. □

Wir kommen nun zu einem wichtigen geometrischen Begriff.

Definition 5.3.5. Man man nennt v und w orthogonal (im rechten Winkel) und schreibt $v \perp w$, wenn $\langle v, w\rangle = 0$ gilt.

Dass man den Begriff des rechten Winkels angemessen eingefangen hat, kann man zunächst anhand von Beispielen im $\mathbb{R}^2$ testen. Eine weitere Rechtfertigung ergibt sich aus der nächsten Proposition und verwandten Ergebnissen im Folgenden.

Proposition 5.3.6 (Satz des Pythagoras)**.** *Man hat*

$$v \perp w \quad \Rightarrow \quad \|v\|^2 + \|w\|^2 = \|v - w\|^2.$$

Im euklidischen Fall gilt auch die umgekehrte Implikation.

Man beachte, dass $v - w$ der Hypotenuse zwischen w und v entspricht.

Beweis. Man hat

$$\|v - w\|^2 = \langle v - w, v - w\rangle = \|v\|^2 - \langle v, w\rangle - \langle w, v\rangle + \|w\|^2.$$

Die Implikation folgt, da mit $\langle v, w\rangle = 0$ auch $\langle w, v\rangle = 0$ (also $w \perp v$) gilt. Umgekehrt folgt (nur) im euklidischen Fall aus $0 = \langle v, w\rangle + \langle w, v\rangle = 2 \cdot \langle v, w\rangle$ schon $v \perp w$. □

Zusätzlich zur Gleichung im vorigen Beweis hat man

$$\|v+w\|^2 = \|v\|^2 + \langle v,w\rangle + \langle w,v\rangle + \|w\|^2.$$

Addiert man die beiden Gleichungen, so erhält man die Parallelogrammgleichung

$$\|v+w\|^2 + \|v-w\|^2 = 2\cdot\|v\|^2 + 2\cdot\|w\|^2.$$

Sie sagt intuitiv, dass die Quadrate über den Diagonalen in einem Parallelogramm in der Summe den Quadraten über den vier Seiten entsprechen. Tatsächlich ist eine Norm genau dann durch ein Skalarprodukt induziert, wenn sie die Parallelogrammgleichung erfüllt (Satz von Jordan und von Neumann).

Bemerkung 5.3.7. Den Winkel zwischen Vektoren $v, w \neq 0$ in einem euklidischen Vektorraum definiert man als

$$\sphericalangle(v,w) = \theta \in [0,\pi] \quad \text{für} \quad \cos(\theta) = \frac{\langle v,w\rangle}{\|v\|\cdot\|w\|}.$$

Dies ist wohldefiniert, denn mit Cauchy-Schwarz liegt der angegebene Bruch im Intervall $[-1,1]$, welches durch den Cosinus bijektiv auf $[0,\pi]$ abgebildet wird. Eine inhaltliche Rechtfertigung geben wir an dieser Stelle nicht, weil wir die Theorie der trigonometrischen Funktionen aus der Analysis nicht voraussetzen wollen.

Wir verbinden mit einem Grundbegriff der linearen Algebra:

Proposition 5.3.8. *Für paarweise orthogonale Vektoren $v_i \neq 0$ (also $v_i \perp v_j$ für alle $i \neq j$) ist die Menge $\{v_1,\dots,v_n\}$ linear unabhängig.*

Beweis. Aus $0 = \sum_{i=1}^n \lambda_i \cdot v_i$ ergibt sich jeweils

$$0 = \left\langle \sum_{i=1}^n \lambda_i \cdot v_i, v_j \right\rangle = \sum_{i=1}^n \lambda_i \cdot \langle v_i, v_j\rangle = \lambda_j \cdot \|v_j\|^2$$

und also $\lambda_j = 0$. □

Die Proposition motiviert den folgenden Begriff.

Definition 5.3.9. Sei B eine Basis eines Innenproduktraums. Wenn alle $v, w \in B$ die Beziehung

$$\langle v,w\rangle = \begin{cases} 1 & \text{wenn } v = w, \\ 0 & \text{sonst} \end{cases}$$

erfüllen, dann heißt B Orthonormalbasis.

Die Basisdarstellung bezüglich einer Orthonormalbasis ist leicht zu bestimmen:

Lemma 5.3.10. *Ist B eine Orthonormalbasis von V, so erfüllt jedes $v \in V$ die Beziehung*

$$v = \sum_{w\in B} \langle v,w\rangle \cdot w,$$

wobei $\langle v,w\rangle \neq 0$ nur für endlich viele $w \in B$ gilt.

Beweis. Wir schreiben $v = \sum_{i=1}^n \lambda_i \cdot v_i$ mit $v_i \in B$ und rechnen

$$\langle v, v_j\rangle = \sum_{i=1}^n \lambda_i \cdot \langle v_i, v_j\rangle = \lambda_j \cdot \|v_j\|^2 = \lambda_j.$$

Ähnlich ergibt sich $\langle v,w\rangle = 0$ für $w \in B\backslash\{v_1,\dots,v_n\}$. □

Der Beweis des folgenden Satzes ist als Gram-Schmidt-Verfahren bekannt.[100]

Proposition 5.3.11. *In jedem endlich-dimensionalen Innenproduktraum hat man eine Orthonormalbasis.*

Beweis. Wir beginnen mit einer beliebigen Basis $\{v_1, \dots, v_n\}$. Per Induktion konstruieren wir orthonormale Vektoren w_k mit

$$\operatorname{Span}(v_1, \dots, v_k) = \operatorname{Span}(w_1, \dots, w_k).$$

Im Induktionsanfang setzt man schlicht $w_1 = v_1/\|v_1\|$. Für den Induktionsschritt definieren wir

$$w_{k+1} = \frac{w}{\|w\|} \quad \text{mit} \quad w = v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, w_i \rangle \cdot w_i,$$

wobei man $w \neq 0$ wegen $v_{k+1} \notin \operatorname{Span}(w_1, \dots, w_k)$ beachte. Man möge sich überzeugen, dass der Spann von $v_1, \dots, v_{k+1}$ mit dem von $w_1, \dots, w_{k+1}$ übereinstimmt. Für $j \leq k$ hat man

$$\langle w_{k+1}, w_j \rangle = \frac{1}{\|w\|} \cdot \left(\langle v_{k+1}, w_j \rangle - \sum_{i=1}^{k} \langle v_{k+1}, w_i \rangle \cdot \langle w_i, w_j \rangle \right) = 0,$$

sodass die Vektoren orthonormal bleiben. □

Während jeder Vektorraum eine Basis hat, gibt es Innenprodukträume überabzählbarer Dimension, die keine Orthonormalbasis besitzen. Die Konstruktion eines Beispiels erfordert einigen Aufwand. Wir verweisen deshalb schlicht auf die Vorlesung zur Funktionalanalysis (beziehungsweise auf entsprechende Lehrbücher[101]). Dort betrachtet man deshalb auch allgemeinere Begriffe von Basis, die etwa unendliche Linearkombinationen zulassen.

Als Anwendung von Orthogonalität untersuchen wir nun, wie sich der Abstand zu einem Untervektorraums bestimmen lässt (Lot als kürzeste Verbindung).

Definition 5.3.12. Für eine Teilmenge $M \subseteq V$ eines Innenproduktraums V heißt

$$M^\perp = \{v \in V : v \perp w \text{ für alle } w \in M\}$$

das orthogonale Komplement von M.

Aus $M \subseteq N$ erhält man $M^\perp \supseteq N^\perp$. Weiter prüft man $M^\perp = \operatorname{Span}(M)^\perp$, weswegen man sich auf die orthogalen Komplemente von Untervektorräumen konzentrieren kann. Der folgende Beweis sei zur Übung überlassen.

Lemma 5.3.13. *Das orthogonale Komplement $M^\perp$ ist stets ein Untervektorraum.*

Wegen des folgenden Ergebnisses ist die Rede vom Komplement gerechtfertigt.

Proposition 5.3.14. *Für jeden endlich-dimensionalen Untervektorraum $U \subseteq V$ eines Innerproduktraums V ist $V = U \oplus U^\perp$ eine direkte Summe.*

---

[100] Dieses ist benannt nach Jørgen Pedersen Gram und Erhard Schmidt.

[101] Wer Mathematik auf Französisch lesen kann, beachte als klassiche Quelle auch Jacques Dixmier, *Sur les bases orthonormales dans les espaces préhilbertiens*, Acta Scientiarum Mathematicarum (Szeged) 15 (1953) 29–30.

Beweis. Zunächst sieht man, dass aus $v \in U \cap U^\perp$ schon $\langle v, v\rangle = 0$ und also $v = 0$ folgt. Sei nun $b_1, \ldots, b_n$ eine Orthonormalbasis von U. Zu gegebenem $v \in V$ definieren wir $w = \sum_{i=1}^n \langle v, b_i\rangle \cdot b_i$. Man erhält jeweils

$$\langle v - w, b_j\rangle = \langle v, b_j\rangle - \sum_{i=1}^n \langle v, b_i\rangle \cdot \langle b_i, b_j\rangle = 0$$

und also $v - w \in \{b_1, \ldots, b_n\}^\perp = U^\perp$. Mit $w \in U$ folgt $v = U \oplus U^\perp$. □

Als Folgerung erhalten wir ein weiteres Beispiel von Dualität:

Korollar 5.3.15. *Für endlich-dimensionales U gilt $(U^\perp)^\perp = U$.*

Beweis. Ist $u \in U$, so hat man $w \perp u$ für jedes $w \in U^\perp$ und also $u \in (U^\perp)^\perp$. Ist nicht nur U sondern auch der umgebende Raum endlich-dimensional, so erhält man die Behauptung aus Dimensionsgründen. Allgemeiner kann man jedes $v \in (U^\perp)^\perp$ wegen des vorigen Ergebnisses als $v = u + w$ mit $u \in U$ und $w \in U^\perp$ schreiben. Mit der bereits gezeigten Inklusion ergibt sich $w = v - u \in (U^\perp)^\perp$ und also $w = 0$, womit wir auf $v \in U$ schließen können. □

Der Begriff der direkten Summe erlaubt uns die folgende Konstruktion.

Definition 5.3.16. Sei $U \subseteq V$ ein endlich-dimensionaler Untervektorraum eines Innenproduktraums V. Die linearen Abbildungen $\rho_U : V \to U$ und $\rho_U^\perp : V \to U^\perp$ sind dadurch charakterisiert, dass $v = \rho_U(v) + \rho_U^\perp(v)$ für alle $v \in V$ gilt.

Aus der Definition ergibt sich unmittelbar $\rho_U \circ \rho_U = \rho_U$ (oder auch $\rho_U(u) = u$ für $u \in U$). Weiter gilt jeweils $v - \rho_U(v) \in U^\perp$, sodass die Verbindung $v - \rho_U(v)$ zu jedem $u \in U$ orthogonal ist. Man nennt ρ_U deshalb auch die orthogonale Projektion auf U. Wir weisen nach, dass diese den kürzesten Abstand realisiert.

Proposition 5.3.17. *Man hat $\|v - \rho_U(v)\| < \|v - u\|$ für jedes $u \in U \backslash \{\rho_U(v)\}$.*

Beweis. Wegen $v - \rho_U(v) \perp \rho_U(v) - u \in U$ erhält man

$$\begin{aligned}\|v - \rho_U(v)\|^2 < \|v - \rho_U(v)\|^2 + \|\rho_U(v) - u\|^2 \\ = \|v - \rho_U(v) + \rho_U(v) - u\|^2 = \|v - u\|^2,\end{aligned}$$

wobei die erste Gleichheit aus dem Satz des Pythagoras folgt. □

Wir bestimmen nun Abstände in einem konkreten Fall.

Beispiel 5.3.18. Mit Proposition 5.3.14 definiert jeder Vektor $w \in \mathbb{R}^3 \backslash \{0\}$ eine Ebene $w^\perp = \{w\}^\perp$ durch den Ursprung. Die parallele Ebene durch $v \in \mathbb{R}^3$ (mit Normalenvektor w) ist gegeben durch

$$v + w^\perp = \{v + u \in \mathbb{R}^3 : \langle u, w\rangle = 0\} = \{u' \in \mathbb{R}^3 : \langle u', w\rangle = \langle v, w\rangle\}.$$

Ist dabei w normiert, so wird die Darstellung als Hessesche Normalform bezeichnet. Der Abstand eines beliebigen Vektors $u \in \mathbb{R}^3$ von der Ebene $w^\perp$ berechnet sich mit der vorigen Proposition als

$$\|u - \rho_{\{w\}}^\perp(u)\| = \|u - (u - \rho_{\{w\}}(u))\| = \|\rho_{\{w\}}(u)\| = \|\langle u, w\rangle \cdot w\|.$$

Ist w normiert (also $\|w\| = 1$), so ergibt sich der Abstand als $|\langle u, w\rangle|$. Der Abstand von u zur Ebene $v + w^\perp$ ist gleich dem Abstand von $u - v$ zu $w^\perp$ und also im normierten Fall gleich $|\langle u, w\rangle - \langle v, w\rangle|$.

Als nächstes betrachten wir die folgenden längentreuen Abbildung.

Definition 5.3.19. Eine Abbildung $\varphi : V \to V$ auf einem Innenproduktraum[102] heißt Isometrie, wenn $\|\varphi(v) - \varphi(w)\| = \|v - w\|$ für alle $v, w \in V$ gilt.

Es folgt unmittelbar, dass jede Isometrie injektiv ist. Die Bedingung $\varphi(0) = 0$ im folgenden Ergebnis lässt sich durch eine Translation leicht erfüllen (betrachte statt φ die verschobene Abbildung $\hat{\varphi}$ mit $\hat{\varphi}(v) = \varphi(v) - \varphi(0)$).

Proposition 5.3.20. *Ist V euklidisch, so ist eine Isometrie $\varphi : V \to V$ mit der Eigenschaft $\varphi(0) = 0$ stets linear .*

Beweis. Man erhält zunächst

$$\|\varphi(v)\| = \|\varphi(v) - \varphi(0)\| = \|v - 0\| = \|v\|.$$

Aus dem Beweis von Satz 5.3.6 hat man

$$\langle v, w\rangle = \frac{\|v\|^2 + \|w\|^2 - \|v - w\|^2}{2}.$$

Man kann hier nun jeweils v und w durch $\varphi(v)$ beziehungsweise $\varphi(w)$ ersetzen. Dadurch ergibt sich $\langle\varphi(v), \varphi(w)\rangle = \langle v, w\rangle$. Man erhält nun

$$\begin{aligned}\|\varphi(\lambda \cdot v) - \lambda \cdot \varphi(v)\|^2 &= \|\varphi(\lambda \cdot v)\|^2 - 2\lambda \cdot \langle\varphi(\lambda \cdot v), \varphi(v)\rangle + \lambda^2 \cdot \|\varphi(v)\|^2 \\ &= \|\lambda \cdot v\|^2 - 2\lambda \cdot \langle\lambda \cdot v, v\rangle + \lambda^2 \cdot \|v\|^2 = 0\end{aligned}$$

und also $\varphi(\lambda \cdot v) = \lambda \cdot \varphi(v)$. Ähnlich zeigt man $\|\varphi(v + w) - \varphi(v) - \varphi(w)\|^2 = 0$. □

Die Proposition lässt sich unter geeigneten Bedingungen auf normierte Vektorräume ohne Skalarprodukt verallgemeinern, wo sie als Satz von Banach-Mazur bekannt ist. Im unitären Fall gilt sie so nicht, da etwa $\mathbb{C} \ni z \mapsto \overline{z} \in \mathbb{C}$ abstandserhaltend aber nicht linear ist (wegen $\overline{i \cdot i} = \overline{-1} = -1$ und $i \cdot \overline{i} = i \cdot (-i) = +1$). Wenn wir im Folgenden allgemeine Innenprodukträume betrachten, fordern wir deshalb explizit Linearität, auch wenn diese im euklidischen Fall automatisch ist. Die folgende Beobachtung aus dem vorigen Beweis gilt auch im unitären Fall. Intuitiv zeigt das Korollar, dass abstandstreue Abbildungen auch winkeltreu sind.

Korollar 5.3.21. *Ist $\varphi : V \to V$ eine lineare Isometrie, gilt $\langle\varphi(v), \varphi(w)\rangle = \langle v, w\rangle$ für alle $v, w \in V$.*

Beweis. Der euklidische Fall wurde im vorigen Beweis behandelt. Im unitären Fall schließt man analog mit der sogenannten Polarisationsidentität

$$\langle v, w\rangle = \frac{1}{4} \cdot \sum_{n=0}^{3} i^n \cdot \|v + i^n \cdot w\|^2,$$

die man durch explizite Berechnung der rechten Seite verifiziert. Man beachte, dass dabei noch Linearität in der Form $\varphi(i \cdot v) = i \cdot \varphi(v)$ eingeht. □

Das Folgende ist konvers zu einer Beobachtung im Beweis von Proposition 5.3.20.

Lemma 5.3.22. *Sei $\varphi : V \to V$ eine lineare Abbildung auf einem Innenproduktraum. Gilt $\|\varphi(v)\| = \|v\|$ für alle $v \in V$, so ist φ eine Isometrie.*

Beweis. Man erhält $\|\varphi(v) - \varphi(w)\| = \|\varphi(v - w)\| = \|v - w\|$. □

[102] Allgemeiner kann man alle normierten und modulo Notation alle metrischen Räume zulassen.

Wir beschäftigen uns nun mit den darstellenden Matrizen von Isometrien, wobei wir auf dem K^n stets das Standardskalarprodukt betrachten. Dies ist keine wesentliche Einschränkung: Hat man ein anderes Skalarprodukt gegeben, so betrachte man eine zugehörige Orthonormalbasis. Durch einen Basiswechsel auf die Einheitsbasis kann man auf das Standardskalarprodukt reduzieren.

Definition 5.3.23. Es sei $K = \mathbb{R}$ bzw. $K = \mathbb{C}$. Eine Matrix $M \in K^{n\times n}$ heißt orthogonal bzw. unitär, wenn $\mathrm{Abb}(M) : K^n \to K^n$ eine Isometrie ist.

Man sieht unmittelbar, dass die Komposition von Isometrien wieder eine Isometrie ist. Weiter sind lineare Isometrien auf endlich-dimensionalen Innenprodukträumen bijektiv, wobei die Umkehrabbildung wieder eine Isometrie ist. Hieraus ergibt sich das folgende Ergebnis.

Korollar 5.3.24. *Sind $M, N \in K^{n\times n}$ beide orthogonal beziehungsweise unitär, so sind es auch $M \cdot N$ und M^{-1}.*

Mit dem Korollar sind die folgenden Begriffe gerechtfertig.

Definition 5.3.25. Für $K = \mathbb{R}$ beziehungsweise $K = \mathbb{C}$ schreiben wir $O(n)$ beziehungsweise $U(n)$ für die Gruppe der orthogonalen beziehungsweise unitären Matrizen in $K^{n\times n}$ mit der Matrixmultiplikation als Verknüpfung.

In Anbetracht des folgenden Ergebnisses ist es etwas verwirrend aber üblich, von orthogonalen statt orthonormalen Matrizen zu sprechen. In vielen Texten werden (b) oder (c) zur Definition von orthogonalen und unitären Matrizen verwendet, wobei dann die Verbindung mit Isometrien zur Proposition wird.

Proposition 5.3.26. *Für $M \in K^{n\times n}$ mit $K = \mathbb{R}$ bzw. $K = \mathbb{C}$ sind äquivalent:*

(i) Die Matrix M ist orthogonal bzw. unitär.
(ii) Die Spalten $M \cdot e_1, \ldots, M \cdot e_n$ von M bilden eine Orthonormalbasis.
(iii) Man hat $M^\top \cdot M = \mathbb{1}$ bzw. $\overline{M}^\top \cdot M = \mathbb{1}$ (mit $\overline{M} = (\overline{\mu_{ij}})$ für $M = (\mu_{ij})$).

Beweis. Die Implikation von (i) nach (ii) gilt gemäß Korollar 5.3.21, da $e_1, \ldots, e_n$ eine Orthonormalbasis ist. Für die Rückrichtung schreiben wir $v = \sum_{i=1}^n \lambda_i \cdot e_i$ und rechnen

$$\begin{aligned}\|M \cdot v\|^2 &= \left\langle \sum_{i=1}^n \lambda_i \cdot M \cdot e_i, \sum_{j=1}^n \lambda_j \cdot M \cdot e_j \right\rangle \\ &= \sum_{i=1}^n \lambda_i \cdot \sum_{j=1}^n \overline{\lambda}_j \cdot \langle M \cdot e_i, M \cdot e_j\rangle = \sum_{i=1}^n \lambda_i \cdot \overline{\lambda}_i = \|v\|^2.\end{aligned}$$

Gemäß Lemma 5.3.22 ist also φ eine Isometrie, wie für (i) benötigt.

Die Matrix $M^\top \cdot \overline{M}$ hat an der Kreuzung der i-ten Zeile mit der j-ten Spalte den Eintrag $\langle M \cdot e_i, M \cdot e_j\rangle$. Daher ist (ii) äquivalent zu $M^\top \cdot \overline{M} = \mathbb{1}$. Indem man beide Seiten transponiert, erhält man die Äquivalenz mit (iii). □

Das vorige Ergebnis zeigt die Bedeutung des folgenden Begriffs.

Definition 5.3.27. Es sei $M \in \mathbb{C}^{n\times n}$. Man schreibt auch M^* für $\overline{M}^\top$ und nennt diese Matrix die adjungierte (oder hermitesch transponierte oder konjugiert transponierte) Matrix von M.

Bedingung (iii) aus Proposition 5.3.26 lässt sich auch als $M^{-1} = M^*$ oder als $M \cdot M^* = \mathbb{1}$ schreiben. Wegen $(M^*)^* = M$ erhält man

$$M \in U(n) \quad \Leftrightarrow \quad M^* \in U(n).$$

Gleiches gilt mit $O(n)$ anstelle von $U(n)$, wobei dann M^* gleich $M^\top$ ist. Weiter folgt nun, dass M genau dann orthogonal bzw. unitär ist, wenn die Zeilen von M eine Orthonormalbasis bilden.

Im Folgenden sind wir mit dem Standard-Skalarprodukt befasst. Man kann adjunigerte Abbildungen auch allgemeiner definieren, wobei diese im unendlichdimensionalen Fall nicht immer existieren. In der Kategorientheorie hat man einen abstrakten Begriff von Adjunktion, der durch dieselbe Struktureigenschaft charakterisiert ist.

Lemma 5.3.28. *Für alle $M \in \mathbb{C}^{n\times n}$ und $v, w \in \mathbb{C}^n$ gilt*

$$\langle M \cdot v, w\rangle = \langle v, M^* \cdot w\rangle.$$

Beweis. Man hat einerseits $\langle M \cdot v, w\rangle = (M \cdot v)^\top \cdot \overline{w} = v^\top \cdot M^\top \cdot \overline{w}$ und andererseits auch $\langle v, M^* \cdot w\rangle = v^\top \cdot \overline{M^* \cdot w} = v^\top \cdot M^\top \cdot \overline{w}$. □

Der folgende Begriff wird durch Beispiel 5.3.18 erhellt.

Definition 5.3.29. Für $w \in \mathbb{R}^n$ mit $\|w\| = 1$ heißt

$$\sigma_w : \mathbb{R}^n \to \mathbb{R}^n \quad \text{mit} \quad \sigma_w(v) = v - 2 \cdot \langle v, w\rangle \cdot w$$

die Spiegelung an der zu w orthogonalen Hyperebene.

Man kann direkt nachrechnen, dass Spiegelungen linear sind. Tatsächlich erfüllen sie die folgende stärkere Bedingung.

Lemma 5.3.30. *Jede Spiegelung σ_w ist eine Isometrie.*

Beweis. In Erweiterung des Satzes von Pythagoras schließen wir aus $u \perp \hat{u}$ zunächst auf $u \perp -\hat{u}$ und dann auf

$$\|u + \hat{u}\|^2 = \|u - (-\hat{u})\|^2 = \|u\|^2 + \|-\hat{u}\|^2 = \|u\|^2 + \|\hat{u}\|^2.$$

Im vorliegenden Fall erhält man

$$\begin{aligned}\|\sigma_w(v)\|^2 &= \|(v - \langle v, w\rangle \cdot w) - \langle v, w\rangle \cdot w\|^2 \\ &= \|v - \langle v, w\rangle \cdot w\|^2 + \|\langle v, w\rangle \cdot w\|^2 = \|(v - \langle v, w\rangle \cdot w) + \langle v, w\rangle \cdot w\|^2 = \|v\|^2\end{aligned}$$

wegen $v - \langle v, w\rangle \cdot w \perp w$. □

Wir zeigen nun, dass $O(n)$ durch Spiegelungen erzeugt ist.

Proposition 5.3.31. *Jede Isometrie $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ ist eine Komposition von höchstens n Spiegelungen (mit $n = 0$ für die Identität).*

Beweis. Ist φ nicht die Identität, so gehen wir modulo Symmetrie von $\varphi(e_1) \neq e_1$ aus. Wegen $\|\varphi(e_1)\| = \|e_1\| = 1$ gilt

$$\langle \varphi(e_1), \varphi(e_1) - e_1\rangle = 1 - \langle \varphi(e_1), e_1\rangle = \langle e_1 - \varphi(e_1), e_1\rangle = \langle -e_1, \varphi(e_1) - e_1\rangle$$

und also $2 \cdot \langle \varphi(e_1), \varphi(e_1) - e_1\rangle = \|\varphi(e_1) - e_1\|^2$. Für $w = (\varphi(e_1) - e_1)/\|\varphi(e_1) - e_1\|$ erhält man

$$\sigma_w \circ \varphi(e_1) = \varphi(e_1) - 2 \cdot \langle \varphi(e_1), w\rangle \cdot w$$

$$= \varphi(e_1) - 2 \cdot \frac{\langle \varphi(e_1), \varphi(e_1) - e_1 \rangle}{\|\varphi(e_1) - e_1\|^2} \cdot (\varphi(e_1) - e_1) = e_1.$$

Für $\sigma_w \circ \varphi = \mathrm{Abb}(M)$ mit $M = (\lambda_{ij})$ hat man also $\lambda_{11} = 1$ und $\lambda_{i1} = 0$ für $i > 1$. Wie im Abschnitt nach Definition 5.3.27 bemerkt, gilt dann auch $\lambda_{1i} = 0$ für $i > 1$, sodass der Untervektorraum $U = \mathrm{Span}(e_2, \ldots, e_n)$ unter $\sigma_w \circ \varphi$ abgeschlossen ist. Induktiv ist die Einschränkung von $\sigma_w \circ \varphi$ auf U eine Komposition von höchstens $n - 1$ Spiegelungen. Wir interpretieren diese als Spiegelungen des $\mathbb{R}^n$, welche e_1 unberührt lassen. Wegen $\sigma_w^2 = \mathrm{id}$ ist φ Summe von höchstens n Spiegelungen. □

Das folgende Ergebnis wird uns helfen, die Struktur von Isometrien noch besser zu verstehen.

Lemma 5.3.32. *Gilt $M \in O(n)$ oder $M \in U(n)$, so hat man $|\det(M)| = 1$.*

Beweis. Es ist

$$1 = \det(\mathbb{1}) = \det(M^* \cdot M) = \overline{\det(M)} \cdot \det(M) = |\det(M)|^2,$$

wie gewünscht. □

Gilt $\sigma_w = \mathrm{Abb}(M)$ mit $w \in \mathbb{R}^n$, so nennen wir auch $M \in O(n)$ eine Spiegelung.

Lemma 5.3.33. *Ist $M \in O(n)$ eine Spiegelung, so gilt $\det(M) = -1$.*

Beweis. Schreibe $\sigma_w = \mathrm{Abb}(M)$ und betrachte eine Orthonormalbasis $w_1, \ldots, w_n$ mit $w = w_1$. Ist T die Matrix mit Spalten w_i, so ist $T^{-1} \cdot M \cdot T$ die Spiegelung bezüglich e_1, denn man hat

$$T^{-1} \cdot M \cdot T \cdot e_1 = T^{-1} \cdot \sigma_w(w) = T^{-1} \cdot (-w) = -e_1$$

sowie $T^{-1} \cdot M \cdot T \cdot e_i = e_i$ für $i > 1$. Es folgt $\det(M) = \det(T^{-1} \cdot M \cdot T) = -1$. □

Die Komposition von zwei Spiegelungen ist eine Drehung. Wir verdeutlichen dies im Folgenden mit Hilfe der trigonometrischen Funktionen.[103]

Beispiel 5.3.34. Wir betrachten die Spiegelungen an $v = e_1$ und einem beliebigen weiteren Vektor w im $\mathbb{R}^2$. Angesichts der Identität $\sin(x)^2 + \cos(x)^2 = 1$ finden wir genau ein $\theta \in [0, 2 \cdot \pi)$ mit $w = (\cos(\theta), \sin(\theta))^\top$. Es ergibt sich

$$\sigma_w \circ \sigma_v(e_1) = -e_1 - 2 \cdot \langle -e_1, w \rangle \cdot w = \begin{pmatrix} -1 + 2 \cdot \cos(\theta)^2 \\ 2 \cdot \sin(\theta) \cdot \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos(2 \cdot \theta) \\ \sin(2 \cdot \theta) \end{pmatrix},$$

$$\sigma_w \circ \sigma_v(e_2) = e_2 - 2 \cdot \langle e_2, w \rangle \cdot w = \begin{pmatrix} -2 \cdot \sin(\theta) \cdot \cos(\theta) \\ 1 - 2 \cdot \sin(\theta)^2 \end{pmatrix} = \begin{pmatrix} -\sin(2 \cdot \theta) \\ \cos(2 \cdot \theta) \end{pmatrix}.$$

Damit ist $\sigma_w \circ \sigma_v$ eine Drehung um den Winkel $2 \cdot \theta$. Liegen v und w im $\mathbb{R}^3$ statt $\mathbb{R}^2$, so ist $\sigma_w \circ \sigma_v$ eine Drehung des Unterraums $\mathrm{Span}(v, w) \cong \mathbb{R}^2$, die dazu orthogonale Vektoren unverändert lässt.

In beliebiger Dimension führen wir den folgenden Begriff ein.

Definition 5.3.35. Die spezielle orthogonale beziehungsweise unitäre Gruppe ist gegeben durch

$$SO(n) = \{M \in O(n) : \det(M) = 1\} \quad \text{und} \quad SU(n) = \{M \in U(n) : \det(M) = 1\}.$$

Man bezeichnet die Elemente von $SO(n)$ auch als Drehungen.

[103] Da hier Resultate aus der Analysis eingehen, stellt der Autor in seinen Prüfungen zur Analysis keine Fragen zu dem folgenden Beispiel.

Speziell für $n = 3$ folgt aus Proposition 5.3.31, dass jedes Element von $SO(3)$ die Verknüpfung von null oder zwei Spiegelungen und also eine der in Beispiel 5.3.34 betrachteten Drehungen ist.

Bemerkung 5.3.36. Der Satz vom Fußball besagt, dass sich beim Anstoß der ersten und zweiten Halbzeit zwei Punkte der Fußballoberfläche an derselben Stelle befinden. Man kann hier argumentieren, dass sich jede Bewegung des Balls als diskrete Folge von Drehungen modellieren lässt. Der Satz folgt dann, weil $SO(3)$ unter Multiplikation abgeschlossen ist (und weil eine Drehnug die Punkte auf der Drehachse fixiert). Alternativ mag man argumentieren, dass die Lage des Fußballs zum Zeitpunkt t durch eine Isometrie und also eine Matrix $M_t \in O(3)$ zu beschreiben ist, wobei $M_0 = \mathbb{1}$ gilt. Aus Gründen der Stetigkeit wird man davon ausgehen, dass stets $\det(M_t) = +1$ gilt, womit M_t für jedes t eine Drehung beschreibt.

5.4. **Normalformen in Innenprodukträumen.** Um Normalformen über $\mathbb{R}$ und über $\mathbb{C}$ zu untersuchen, ist der Fundamentalsatz der Algebra unerlässlich. Wie bereits erwähnt ist durch diesen garantiert, dass jede lineare Abbildung auf dem $\mathbb{C}^n$ einen Eigenwert besitzt. Es gilt auch die umgekehrte Implikation:

Lemma 5.4.1. *Gilt* $\operatorname{Spec}(M) \neq \emptyset$ *für jede Matrix* $M \in K^{n\times n}$ *mit* $n > 0$*, so ist* K *algebraisch abgeschlossen.*

Beweis. Man hat

$$\begin{vmatrix} x & -a_0 \\ 1 & x + a_1 \end{vmatrix} = x \cdot (x + a_1) + a_0 = x^2 + a_1 \cdot x + a_0.$$

Durch Entwicklung nach der ersten Zeile erhält man

$$\begin{vmatrix} x & 0 & a_0 \\ 1 & x & -a_1 \\ 0 & 1 & x + a_2 \end{vmatrix} = x \cdot (x^2 + a_2 \cdot x + a_1) + a_0 = x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0.$$

Induktiv findet man für jedes Polynom $p(x) = \sum_{i=0}^{n} (-1)^n \cdot a_i \cdot x^i \in K[x]$ mit $a_n = 1$ eine Matrix $M \in K^{n\times n}$ mit $p(x) = \chi_M(x)$ und also $p(\lambda) = 0$ für $\lambda \in \operatorname{Spec}(M)$. □

Äquivalent zum Fundamentalsatz der Algebra können wir also auch direkt die Existenz von Eigenwerten nachweisen.[104]

Theorem 5.4.2. *Es gilt* $\operatorname{Spec}(M) \neq \emptyset$ *für alle* $M \in \mathbb{C}^{n\times n}$ *mit* $n > 0$.

Beweis. Wir behandeln zunächst den Fall, dass n ungerade ist. Hier betrachten wir

$$V = \{N \in \mathbb{C}^{n\times n} : N^* = N\}$$

als $\mathbb{R}$-Vektorraum (vgl. Definition 5.3.27). Dieser hat die ungerade Dimension n^2, denn wir können auf der Diagonalen n reelle Einträge und oberhalb der Diagonalen $\sum_{i<n} i$ komplexe Einträge wählen, wobei letztere $2 \cdot \sum_{i<n} i = (n-1) \cdot n$ reellen Parametern entsprechen.

Mit Proposition 4.3.13 hat jede lineare Abbildung $\varphi : V \to V$ einen Eigenvektor. Wir betrachten die linearen Abbildungen $\varphi_i : V \to V$ mit

$$\varphi_1(N) = \frac{M \cdot N + N \cdot M^*}{2} \quad \text{und} \quad \varphi_2(N) = \frac{M \cdot N - N \cdot M^*}{2i}.$$

[104]Der Beweis orientiert sich an Ausführungen von Harm Derksen, *The Fundamental Theorem of Algebra and Linear Algebra*, The American Mathematical Monthly 110:7 (2003) 620–623.

Unser Ziel ist es, einen gemeinsamen Eigenvektor $N \in V$ mit $\varphi_1(N) = \lambda \cdot N$ sowie $\varphi_2(N) = \mu \cdot N$ für geeignete $\lambda, \mu \in \mathbb{R}$ zu finden. Man erhält dann nämlich

$$(\lambda + i \cdot \mu) \cdot N = (\varphi_1 + i \cdot \varphi_2)(N) = M \cdot N,$$

sodass jede Spalte von N ein Eigenvektor von M ist.

Durch Rechnung prüft man $\varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1$, sodass unsere Abbildungen kommutieren. Durch Induktion über ungerades $d \in \mathbb{N}$ zeigen wir, dass zwei lineare Abbildungen $\psi_i : \mathbb{R}^d \to \mathbb{R}^d$ stets einen gemeinsamen Eigenvektor haben. Sei dazu λ ein beliebiger Eigenwert von ψ_1 (vgl. Proposition 4.3.13). Wir setzen

$$U = \ker(\psi_1 - \lambda \cdot \mathrm{id}) \qquad \text{und} \qquad W = \mathrm{img}(\psi_1 - \lambda \cdot \mathrm{id}).$$

Wegen $\psi_2(\psi_1(v) - \lambda \cdot v) = \psi_1(\psi_2(v)) - \lambda \cdot \psi_2(v)$ sind U und W unter ψ_2 sowie unter ψ_1 abgeschlossen. Mit dem Rangsatz hat U oder W ungerade Dimension. Weiter gilt $U \neq \{0\}$ und also $\dim(W) < d$. Wir können induktiv schließen, wenn nicht $U = \mathbb{R}^d$ gilt. Ist letzteres der Fall, so ist jeder Eigenvektor von ψ_2 ein gemeinsamer Eigenvektor.

Im allgemeinen Fall schreiben wir $n = 2^m \cdot k$ für ungerades k. Wir zeigen die Behauptung per Induktion über $m \in \mathbb{N}$. Den Basisfall mit $m = 0$ haben wir bereits behandelt. Wir nehmen nun $m > 0$ an und betrachten

$$V' = \{N \in \mathbb{C}^{n \times n} : N^\top = -N\}$$

als $\mathbb{C}$-Vektorraum. Da die Diagonaleinträge von $N \in \mathbb{C}^{n \times n}$ gleich Null sind, gilt

$$\dim(V') = n \cdot (n-1)/2 = 2^{m-1} \cdot k \cdot (n-1),$$

wobei $k \cdot (n-1)$ ungerade ist. Hat man zwei lineare Abbildungen $\psi : V' \to V'$, so hat also nach Induktionsvoraussetzung jede einen Eigenwert. Ähnlich wie zuvor folgert man, dass es sogar einen gemeinsamen Eigenvektor gibt. Genauer findet man gemeinsame Eigenwerte von kommutierenden Homomorphismen $\psi_i : \mathbb{R}^d \to \mathbb{R}^d$ mit $d = 2^j \cdot l$ per Induktion über $j < m$ und Hilfsinduktion über ungerades l.

Konkret betrachten wir die kommutierenden Homomorphismen $\varphi_i' : V' \to V'$ mit

$$\varphi_1'(N) = M \cdot N - N \cdot M^\top \qquad \text{und} \qquad \varphi_2'(N) = M \cdot N \cdot M^\top.$$

Wir finden also $N \in V' \backslash \{0\}$ sowie $\lambda, \mu \in \mathbb{C}$ mit $\varphi_1'(N) = \lambda \cdot N$ und $\varphi_2'(N) = \mu \cdot N$. Es folgt $N \cdot M^\top = M \cdot N - \lambda \cdot N$ und dann $\mu \cdot N = M \cdot (M \cdot N - \lambda \cdot N)$, also

$$(M^2 - \lambda \cdot M - \mu \cdot \mathbb{1}) \cdot v = 0$$

für eine beliebige Spalte v von N, wobei wir $v \neq 0$ annehmen können. Angesichts der Lösungsformel für quadratische Gleichungen betrachten wir eine komplexe Zahl η mit $\eta^2 = \lambda^2 + 4 \cdot \mu$ (welche nach Lemma 4.3.27 existiert). Es folgt

$$\left(M - \frac{\lambda + \mu}{2} \cdot \mathbb{1}\right) \cdot \left(M - \frac{\lambda - \mu}{2} \cdot \mathbb{1}\right) \cdot v = 0.$$

Also ist $(\lambda - \mu)/2$ oder $(\lambda + \mu)/2$ ein Eigenvektor von M. □

Mit Lemma 5.4.1 folgt der sogenannte Fundamentalsatz der Algebra:

Korollar 5.4.3. *Der Körper $\mathbb{C}$ ist algebraisch abgeschlossen.*

Jede Matrix über $\mathbb{C}$ lässt sich also in Jordan-Normalform bringen (aufgrund von Korollar 5.2.16) und insbesondere trigonalisieren. Mit dem Gram-Schmidt-Verfahren folgern, dass sich die Transformationsmatrix unitär wählen lässt.

Proposition 5.4.4 (Satz von Schur[105])**.** *Für jede Matrix $M \in \mathbb{C}^{n\times n}$ gibt es eine Matrix $U \in U(n)$, für die $U^{-1} \cdot M \cdot U$ in oberer Dreiecksform ist.*

Der Beweis wird zeigen, dass man $U \in O(n)$ wählen kann, wenn $M \in \mathbb{R}^{n\times n}$ über $\mathbb{R}$ trigonalisierbar ist.

Beweis. Gemäß Korollar 5.2.16 (siehe auch Bemerkung 5.2.18) lässt sich M trigonalisieren. Man hat also eine Basis $b_1, \dots, b_n$ mit der Eigenschaft, dass $\operatorname{Span}(b_1, \dots, b_j)$ jeweils unter $\operatorname{Abb}(M)$ abgeschlossen ist. Mit dem Beweis von Proposition 5.3.11 können wir annehmen, dass diese Basis orthonormal ist. Es gilt nun also jeweils

$$M \cdot b_j = \sum_{i=1}^{n} \lambda_{ij} \cdot b_i \quad \text{mit } \lambda_{ij} = 0 \text{ für } i > j.$$

Sei $U \in U(n)$ die Matrix mit Spalten $U \cdot e_i = b_i$ (vgl. Proposition 5.3.26). Gemäß Bemerkung 4.5.3 ist dann $U^{-1} \cdot M \cdot U = (\lambda_{ij})$. □

Als Variante des vorigen Beweises erhalten wir eine Zerlegung, die bereits in Bemerkung 5.2.18 angesprochen wurde.

Proposition 5.4.5 (QR-Zerlegung)**.** *Für jede invertierbare Matrix $M \in \mathbb{C}^{n\times n}$ gibt es eine Matrix $Q \in U(n)$ und eine obere Dreiecksmatrix $R \in \mathbb{C}^{n\times n}$ mit $M = Q \cdot R$.*

Der Beweis zeigt, dass dieselbe Aussage auch mit $\mathbb{R}$ und $O(n)$ anstelle von $\mathbb{C}$ beziehungsweise $U(n)$ gilt.

Beweis. Mit dem Gram-Schmidt-Verfahren (also dem Beweis von Proposition 5.3.11) erhalten wir eine Orthonormalbasis $b_1, \dots, b_n$ mit

$$\operatorname{Span}(M \cdot e_1, \dots, M \cdot e_j) = \operatorname{Span}(b_1, \dots, b_j)$$

für alle $i \leq n$. Man erhält also jeweils

$$M \cdot e_j = \sum_{i=1}^{n} \mu_{ij} \cdot b_i \quad \text{mit } \mu_{ij} = 0 \text{ für } i > j.$$

Es sei Q die Matrix mit Spalten $b_1, \dots, b_n$. Da die Einheitsvektoren e_j die Spalten der Matrix $\mathbb{1}$ ausmachen, liefert Bemerkung 4.5.3 die Gleichung $Q^{-1} \cdot M \cdot \mathbb{1} = (\lambda_{ij})$. Die Proposition gilt also mit $R = (\lambda_{ij})$. □

Zur konkreten Berechnung weisen wir darauf hin, dass im vorigen Beweis jeweils $\mu_{ij} = \langle M \cdot e_j, b_i \rangle$ gilt, wie sich aus Lemma 5.3.10 ergibt. Als nächstes beschäftigen wir uns mit Diagonalisierbarkeit über Innenproduktträumen, wobei sich der folgende Begriff als zentral herausstellt (vgl. auch den Beweis von Theorem 5.4.2).

Definition 5.4.6. Eine Matrix $M \in K^{n\times n}$ mit $K = \mathbb{R}$ bzw. $K = \mathbb{C}$ heißt symmetrisch bzw. selbstadjungiert, wenn $M = M^\top$ bzw. $M = M^*$ gilt.

Manchmal werden selbstadjungierte Matrizen auch als hermitesch bezeichnet.

Proposition 5.4.7. *Alle Eigenwerte einer selbstadjungierten Matrix sind reell.*

Beweis. Wir betrachten $M \in \mathbb{C}^{n\times n}$ mit $M = M^*$ sowie $M \cdot v = \lambda \cdot v$ für $v \neq 0$. Mit Lemma 5.3.28 erhält man

$$\lambda \cdot \|v\|^2 = \langle \lambda \cdot v, v \rangle = \langle M \cdot v, v \rangle = \langle v, M^* \cdot v \rangle = \langle v, M \cdot v \rangle = \langle v, \lambda \cdot v \rangle = \overline{\lambda} \cdot \|v\|^2.$$

Es gilt also $\lambda = \overline{\lambda}$ und somit $\lambda \in \mathbb{R}$. □

[105]Dieses Ergebnis ist benannt nach Issai Schur.

In diesem speziellen Fall garantiert der Fundamentalsatz der Algebra auch über den reellen Zahlen die Existenz von Nullstellen:

Korollar 5.4.8. *Für jede symmetrische Matrix $M \in \mathbb{R}^{n\times n}$ zerfällt das charakteristische Polynom χ_M über $\mathbb{R}$ in Linearfaktoren.*

Beweis. Wir können M auch als selbstadjungierte Matrix in $\mathbb{C}^{n\times n}$ auffasen und erhalten mit dem Fundamentalsatz der Algebra eine Faktorisierung

$$\chi_M(x) = (\lambda_1 - x)^{\alpha(1)} \cdot \ldots \cdot (\lambda_k - x)^{\alpha(k)}.$$

Hierbei gilt a priori $\lambda_i \in \mathbb{C}$ aber mit dem vorigen Ergebnis tatsächlich $\lambda_i \in \mathbb{R}$. □

Das folgende Ergebnis wird eine orthonormale Diagonalisierung ermöglichen.

Proposition 5.4.9. *Ist M symmetrisch oder selbstadjungiert, so sind Eigenvektoren zu verschiedenen Eigenwerten von M stets orthogonal.*

Beweis. Es gelte $M \cdot v = \lambda \cdot v$ und $M \cdot w = \mu \cdot w$. Nach Proposition 5.4.7 ist dabei $\mu \in \mathbb{R}$ und also $\overline{\mu} = \mu$. Mit Lemma 5.3.28 folgt

$$\begin{aligned}\lambda \cdot \langle v, w\rangle = \langle \lambda \cdot v, w\rangle = \langle M \cdot v, w\rangle = \langle v, M^* \cdot w\rangle \\ = \langle v, M \cdot w\rangle = \langle v, \mu \cdot w\rangle = \overline{\mu} \cdot \langle v, w\rangle = \mu \cdot \langle v, w\rangle\end{aligned}$$

Gilt $\lambda \neq \mu$, so folgt daraus $\langle v, w\rangle = 0$. □

Die folgende Äquivalenz bleibt gültig, wenn man in Aussage (ii) Diagonalmatrizen $D \in \mathbb{C}^{n\times n}$ zulässt (mit $\mathbb{C}$ statt $\mathbb{R}$) und dafür in (i) nur fordert, dass M normal ist, also $M \cdot M^* = M^* \cdot M$ erfüllt. Auch dieses Ergebnis ist als Spektralsatz bekannt.

Theorem 5.4.10 (Spektralsatz)**.** *Für $M \in \mathbb{C}^{n\times n}$ sind äquivalent:*

(i) Die Matrix M ist selbstadjungiert.
(ii) Es gibt $U \in U(n)$ und eine Diagonalmatrix $D \in \mathbb{R}^{n\times n}$ mit $M = U \cdot D \cdot U^{-1}$.

Beweis. Wir nehmen zunächst (i) an und folgern (ii). Betrachte $\lambda \in \mathrm{Spec}(M) \subseteq \mathbb{R}$ und schreibe $N = M - \lambda \cdot \mathbb{1}$. Aus $M^* = M$ folgt $N^* = M^* - \overline{\lambda} \cdot \mathbb{1} = N$ und damit

$$\|N \cdot v\|^2 = \langle N \cdot v, N \cdot v\rangle = \langle v, N^* \cdot N \cdot v\rangle = \langle v, N^2 \cdot v\rangle.$$

Gilt also $N^2 \cdot v = 0$, so folgt bereits $N \cdot v = 0$. Damit ist der Hauptraum $E^*_M(\lambda)$ gleich dem Eigenraum $E_M(\lambda)$ (vgl. den Beweis von Lemma 5.2.6). Gemäß Theorem 5.2.15 (und mit dem Fundamentalsatz der Algebra) erhält man

$$\mathbb{C}^n = \bigoplus \{E_M(\lambda) : \lambda \in \mathrm{Spec}(M)\}.$$

Mit dem Gram-Schmidt-Verfahren wählen wir jeweils eine Orthonormalbasis B_λ von $E_M(\lambda)$. Mit der vorigen Proposition ist dann $B = \bigcup\{B_\lambda : \lambda \in \mathrm{Spec}(M)\}$ eine Orthonormalbasis des $\mathbb{C}^n$. Ist U die Matrix mit den Vektoren aus B als Spalten, so ist also $U^{-1} \cdot M \cdot U$ die gewünschte Diagonalmatrix.

Gilt umgekehrt $M = U \cdot D \cdot U^{-1} = U \cdot D \cdot U^*$ wie in (ii), dann erhält man

$$M^* = (U^*)^* \cdot D^* \cdot U^* = U \cdot D \cdot U^* = M,$$

sodass (i) erfüllt ist. □

Für konkrete Berechnungen ist die folgende Beobachtung nützlich.

Bemerkung 5.4.11. Wir betrachten das vorige Theorem mit Eigenwerten λ_i und Basisvektoren b_i, wobei also jeweils $D \cdot e_i = \lambda_i \cdot e_i$ und $b_i = U \cdot e_i$ gilt. Es ergibt sich

$$U \cdot D \cdot U^{-1} \cdot b_i = U \cdot D \cdot e_i = U \cdot \lambda_i \cdot e_i = \lambda_i \cdot U \cdot e_i = \lambda_i \cdot b_i.$$

Nach Lemma 5.3.10 hat man stets $v = \sum_{i=1}^{n} \langle v, b_i \rangle \cdot b_i$ und damit

$$M \cdot v = \sum_{i=1}^{n} \langle v, b_i \rangle \cdot U \cdot D \cdot U^{-1} \cdot b_i = \sum_{i=1}^{n} \lambda_i \cdot \langle v, b_i \rangle \cdot b_i.$$

Diese Darstellung ist als Spektralzerlegung von M bekannt.

Im reellen Fall haben wir das folgende Ergebnis.

Korollar 5.4.12. *Für jede symmetrische Matrix M gibt es $O \in O(n)$ und eine Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ mit $M = O \cdot D \cdot O^{-1}$.*

Beweis. Man argumentiert wie im vorigen Beweis, wobei Korollar 5.4.8 die Rolle des Fundamentalsatzes der Algebra übernimmt. □

Symmetrische Matrizen treten unter anderem in der Analysis auf, wo man mit ihnen Maxima und Minima von Funktionen in mehreren Variablen bestimmen kann.

Beispiel 5.4.13. Wir betrachten die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x, y) = x^4 + y^4 - 4 \cdot x \cdot y.$$

Die partiellen Ableitungen (bei denen wir nach einer Variablen ableiten und die andere als Konstante behandeln) sind gegeben durch

$$\frac{\partial f}{\partial x}(x, y) = 4 \cdot x^3 - 4 \cdot y \quad \text{und} \quad \frac{\partial f}{\partial y}(x, y) = 4 \cdot y^3 - 4 \cdot x.$$

Als zweite partielle Ableitungen erhält man

$$\frac{\partial^2 f}{\partial^2 x}(x, y) = 12 \cdot x^2, \quad \frac{\partial^2 f}{\partial x \partial y}(x, y) = -4 = \frac{\partial^2 f}{\partial y \partial x}(x, y), \quad \frac{\partial^2 f}{\partial^2 y}(x, y) = 12 \cdot y^2.$$

Die sogenannte Hesse-Matrix ist damit gegeben durch

$$H_f(x, y) = \begin{pmatrix} 12 \cdot x^2 & -4 \\ -4 & 12 \cdot y^2 \end{pmatrix}.$$

Da die partiellen Ableitungen unter geeigneten Bedingungen vertauschbar sind, ist die Hesse-Matrix auch allgemein symmetrisch.

Im Zusammenhang mit der eben erwähnten Anwendung in der Analysis sind die folgenden Begriffe von Bedeutung.

Definition 5.4.14. Eine symmetrische bzw. selbstadjungierte Matrix $M \in K^{n \times n}$ heißt positiv definit, wenn $\langle v, M \cdot v \rangle > 0$ für alle $v \in \mathbb{K}^n \backslash \{0\}$ gilt.

Wegen $M = M^*$ ist $\langle v, M \cdot v \rangle = \langle M \cdot v, v \rangle$. In der Definition wird $\langle v, M \cdot v \rangle \in \mathbb{R}$ vorausgesetzt, was sich durch den folgenden Beweis erklärt.

Lemma 5.4.15. *Ist M symmetrisch bzw. selbstadjungiert, so ist M genau dann positiv definit, wenn alle Eigenwerte von M positiv sind.*

Beweis. Gilt $M \cdot v = \lambda \cdot v$ mit $v \neq 0$ (gemäß Proposition 5.4.7 mit $\lambda \in \mathbb{R}$), so hat man $\langle v, M \cdot v \rangle = \lambda \cdot \|v\|^2$. Ist M positiv definit, so muss also $\lambda > 0$ gelten. Wir nehmen nun umgekehrt an, dass alle Eigenwerte positiv sind. Selbiges gilt dann für die Diagonaleinträge der Matrix $D = U^{-1} \cdot M \cdot U$ aus dem Sprektralsatz. Für beliebiges $v \neq 0$ setzen wir $w = U^* \cdot v \neq 0$ (also $v = U \cdot U^* \cdot v = U \cdot w$). Es folgt

$$\langle v, M \cdot v \rangle = \langle U \cdot w, U \cdot D \cdot w \rangle = \langle w, D \cdot w \rangle > 0,$$

wobei wir Korollar 5.3.21 verwendet haben. □

Wir können nun das folgende Ergebnis ableiten.

Proposition 5.4.16 (Kriterium von Sylvester[106])**.** *Ist $M \in K^{n \times n}$ symmetrisch bzw. selbstadjungiert, so ist M genau dann positiv definit, wenn $\det(A_k) > 0$ für alle $k = \{1, \ldots, n\}$ gilt. Hierbei geht A_k aus A durch Streichen der letzten $k-1$ Zeilen und Spalten hervor (wobei man $\det(A_k)$ auch als k-ten Hauptminor bezeichnet).*

Beweis. Wir argumentieren per Induktion über $n > 0$. Im Basisfall von $n = 1$ schreiben wir $M = (\lambda)$ mit $\lambda \in \mathbb{R}$. Für $v = (z)$ ist dann $\langle v, M \cdot v \rangle = \lambda \cdot |z|^2$. Dies ist genau dann für alle $v \neq 0$ positiv, wenn $\det(A_1) = \det(A) = \lambda > 0$ gilt.

Im Induktionsschritt nehmen wir zunächst an, dass $M \in K^{n \times n}$ mit $n > 1$ positiv definit ist. Die Matrix M_{n-1} ist ebenfalls selbstadjungiert und positiv definit: Man hat nämlich $\langle v, M_{n-1} \cdot v \rangle = \langle w, M \cdot w \rangle$, wenn $w \in K^n$ den Vektor $v \in K^{n-1}$ um den letzten Eintrag Null verlängert. Induktiv ist also $\det(M_k) > 0$ für alle $k \leq n-1$. Weiter ist $\det(M_n) = \det(M)$ positiv: Sonst hätte die diagonalisierbare Matrix M (vgl. den Spektralsatz) einen Eigenwert $\lambda \leq 0$. Für einen zugehörigen Eigenvektor v würde dann aber $\langle v, M \cdot v \rangle = \lambda \cdot \|v\|^2 \leq 0$ gelten.

Für die Rückrichtung im Induktionsschritt nehmen wir an, dass $\det(M_k) > 0$ für alle $k \leq n$ gilt. Induktiv folgt daraus, dass M_{n-1} positiv definit ist. Der Spektralsatz liefert eine Orthonormalbasis $v_1, \ldots, v_n$ aus Eigenvektoren von M, wobei jeweils $M \cdot v_i = \lambda_i \cdot v_i$ gelte. Wegen $0 < \det(M) = \lambda_1 \cdot \ldots \cdot \lambda_n$ kann es nicht genau ein $i \in \{1, \ldots, n\}$ mit $\lambda_i \leq 0$ geben. Für einen Widerspruch nehmen wir nun an, es gibt $i \neq j$ mit $\lambda_i, \lambda_j \leq 0$. Sind z_i und z_j die letzten Einträge von v_i und v_j, so ist der letzte Eintrag von $w = z_j \cdot v_i - z_i \cdot v_j$ gleich Null. Da M_{n-1} positiv definit ist, muss dann $\langle w, A \cdot w \rangle > 0$ gelten. Andererseits erhält man

$$\langle w, A \cdot w \rangle = \langle z_j \cdot v_i - z_i \cdot v_j, z_j \cdot \lambda_i \cdot v_i - z_i \cdot \lambda_j \cdot v_j \rangle = |z_j|^2 \cdot \lambda_i + |z_i|^2 \cdot \lambda_j \leq 0,$$

was einen Widerspruch ergibt. Somit sind alle Eigenwerte von M positiv. Wir schließen mit dem vorigen Lemma. □

Beispiel 5.4.17. In der Analysis lernt man, Extrema der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ aus dem vorigen Beispiel wie folgt zu bestimmen. Zunächst sieht man, dass die Gleichungen

$$\frac{\partial f}{\partial x} = 4 \cdot x^3 - 4 \cdot y = 0 = 4 \cdot y^3 - 4 \cdot x = \frac{\partial f}{\partial x}$$

für (x, y) die Lösungen $(-1, -1)$ und $(0, 0)$ und $(1, 1)$ zulassen. Die Hessematrix

$$H_f(\pm 1, \pm 1) = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}$$

[106]Dieses ist benannt nach James Joseph Sylvester. Den hier gegebenen Beweis kenne ich aus einem Artikel von Giorgio Giorgi, *Various Proofs of the Sylvester Criterion for Quadratic Forms*, Journal of Mathematics Research 9:6 (2017) 55-66.

ist nach dem Kriterium von Sylvester positiv definit. Die Punkte $(-1,-1)$ und $(1,1)$ sind damit lokale Minima der Funktion f.

Zum Abschluss dieses Kapitels zeigen wir, wie die Jordan-Normalform über dem algebraisch abgeschlossenen Körper $\mathbb{C}$ auch eine Normalform für Matrizen über $\mathbb{R}$ induziert. Wir beschränken uns dabei auf einen konkreten Fall.

Beispiel 5.4.18. Wir betrachten die Matrix

$$M = \begin{pmatrix} 2 & 1 & 1 & -3 \\ 0 & 5 & 3 & 0 \\ 0 & -6 & -1 & 0 \\ 3 & 1 & 0 & 2 \end{pmatrix}.$$

Mit den Methoden aus Abschnitt 5.2 berechnet man die Jordan-Normalform

$$J = \begin{pmatrix} 2-3i & 1 & 0 & 0 \\ 0 & 2-3i & 0 & 0 \\ 0 & 0 & 2+3i & 1 \\ 0 & 0 & 0 & 2+3i \end{pmatrix}.$$

Es ist es kein Zufall, dass die Eigenwerte hier in konjugierten Paaren auftreten: Da die Koeffizienten von χ_M reell sind, folgt aus $\chi_M(z) = 0$ auch $\chi_M(\overline{z}) = \overline{\chi_M(z)} = 0$. Die Eigenvektoren ergeben sich als

$$M \cdot v_1 = \begin{pmatrix} -3-2i \\ 0 \\ 0 \\ 2-3i \end{pmatrix} = (2-3i) \cdot v_1 \quad \text{mit} \quad v_1 = \begin{pmatrix} -i \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

sowie als

$$M \cdot \overline{v_1} = \overline{M \cdot v_1} = \overline{(2-3i) \cdot v_1} = (2+3i) \cdot \overline{v_1}.$$

Mit Blick auf die Jordan-Ketten rechnet man

$$\big(M - (2-3i) \cdot \mathbb{1}\big) \cdot v_2 = v_1 \quad \text{für} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ -1-i \\ 0 \end{pmatrix}.$$

Es ergibt sich damit auch

$$\big(M - (2+3i) \cdot \mathbb{1}\big) \cdot \overline{v_2} = \overline{\big(M - (2-3i) \cdot \mathbb{1}\big) \cdot v_2} = \overline{v_1}.$$

Indem wir v_1 und v_2 in Real- und Imaginärteil aufspalten, erhalten wir die Vektoren

$$w_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad w_4 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}.$$

Wegen $w_1 = (v_1 + \overline{v_1})/2$ ergibt sich

$$M \cdot w_1 = \frac{M \cdot v_1 + M \cdot \overline{v}_1}{2} = \frac{(2-3i) \cdot v_1 + \overline{(2-3i) \cdot v_1}}{2} = \begin{pmatrix} -3 \\ 0 \\ 0 \\ 2 \end{pmatrix} = 2 \cdot w_1 + 3 \cdot w_2,$$

was dem Realteil von $M \cdot v_1$ entspricht. Mit $w_2 = (v_1 - \overline{v_1})/2$ und analogen Beschreibungen von w_3 und w_4 erhalten wir

$$M \cdot w_2 = -3 \cdot w_1 + 2 \cdot w_2, \quad M \cdot w_3 = w_1 + 2 \cdot w_3 + 3 \cdot w_4, \quad M \cdot w_4 = w_2 - 3 \cdot w_3 + 2 \cdot w_4.$$

Ist S die invertierbare Matrix mit Spalten $w_1, \ldots, w_4$, so hat man also

$$S^{-1} \cdot M \cdot S = \begin{pmatrix} 2 & -3 & 1 & 0 \\ 3 & 2 & 0 & 1 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 3 & 2 \end{pmatrix}.$$

In dieser 'reellen Jordan-Normalform' wurden die Eigenwerte $2 \pm 3i$ und die Einträge Eins oberhalb der Diagonalen ersetzt durch die 2×2-Matrizen

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} \quad \text{beziehungsweise} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$