Lineare Algebra und Geometrie

3–semestriger Zyklus ab WS 2014/15

Andreas Čap

Fakultät für Mathematik, Universität Wien, Oskar–Morgenstern–Platz 1, 1090 Wien
E-mail address: Andreas.Cap@univie.ac.at

Inhaltsverzeichnis

Vorwort

Das vorliegende Skriptum stellt eine Ergänzung zu dem dreisemestrigen Vorlesungszyklus über Lineare Algebra und Geometrie dar, den ich ab Wintersemester 2014/15 gehalten habe. Es baut auf mehrere Vorversionen für zweisemestrige Zyklen auf, deren letzter bereits aufbauend auf eine "Einführung in das mathematische Arbeiten" konzipiert war. Nachdem ich inzwischen die "Einführung in das mathematische Arbeiten" auch selbst gehalten habe, hoffe ich, dass die Anknüpfung an diese Vorlesung und das zugehörige Buch von Schichl und Steinbauer, das im weiteren als [EMA] zitiert wird, weiter verbessert wurde.

Die wesentlichen Inhalte für einen Vorlesungszyklus über lineare Algebra und Geometrie sind natürlich weitgehend vorgegeben, in der Wahl von Aufteilung und Reihenfolge gibt es doch einiges an Spielraum. Für den ersten Teil der Vorlesung ("Einführung in die lineare Algebra und Geometrie") liegt die Herausforderung in der Gestaltung am Anfang wohl zum Großteil in der Verarbeitung des Abstraktionsschocks. Danach habe ich mich bemüht die Inhalte in einem breiten Rahmen zu behandeln, um der Rolle der linearen Algebra als grundlegendes Werkzeug für große Teile der Mathematik gerecht zu werden. In den späteren Teilen wird auch versucht an weiterführende Themen heranzuführen. So enthält die Diskussion von Normen und inneren Produkten einige Ausblicke auf die Funktionalanalysis, und im Abschnitt über Tensorprodukte werden durchwegs Konstruktionen und Argumentationsweisen verwendet, die sich breit verallgemeinern lassen. Ich habe mich auch bemüht explizit auf geometrische Themen einzugehen, auch wenn das aus Zeitmangel oft nicht in dem von mir gewünschten Ausmaß möglich war.

Im ersten Teil der Zyklus ("Einführung in die lineare Algebra und Geometrie") wurden die Kapitel 1 bis 4 sowie die Anwendungen dieses Materials auf linear Codes und affine Geometrie in der ersten Hälfte von Kapitel 5 behandelt. Das zweite Semester ("Lineare Algebra und Geometrie 1") umfasste das Material über Quotienten und Dualräume aus Kapitel 5, sowie die Kapitel 6 bis 9. In der zweistündigen Vorlesung "Lineare Algebra und Geometrie 2", die den letzten Teil des Zyklus bildet, wurden die Kapitel 10 und 11 besprochen. Dazu ist zu sagen, dass in den Teilen am Ende der jeweiligen Semester (affine Geometrie in Kapitel 5, Hamilton'sche Quaternionen in Kapitel 9 und die zweite Hälfte von Kapitel 11) die Inhalte des Skriptum nur in verkürzter Form oder nur teilweise in der Vorlesung besprochen wurden. Es handelt sich hier jeweils um Material, das weiterführenden Charakter hat und in späteren Teilen jeweils nicht direkt benötigt wird.

Zum Inhalt des Skriptums: Durch die vorhergehende "Einführung in das mathematische Arbeiten" ist es nicht notwendig, Hintergrund über Mengen und ähnliche Themen zu behandeln, sondern man kann direkt in die Hauptinhalte der linearen Algebra einsteigen.

Um den Einstieg halbwegs sanft zu gestalten, habe ich ein Kapitel über lineare Gleichungssystem in zwei und drei Variablen über $\mathbb{R}$ an den Anfang der Vorlesung gestellt. Es geht dabei hauptsächlich darum, einige typische Phänomene der linearen

Algebra zu “entdecken” bzw. zu beobachten, deren Erklärung erst im Lauf der Vorlesung gelingen wird. Daneben wird auch die geometrisch–anschauliche Interpretation solcher Gleichungssysteme besprochen.

Im zweiten Kapitel werden die Grundkonzepte der linearen Algebra, Vektorräume und lineare Abbildungen, eingeführt und mit vielen Beispielen illustriert, und Kern und Bild einer linearen Abbildung besprochen. Kapitel 3 unterbricht die Entwicklung der abstrakten Theorie und wendet sich der konkreten Beschreibung von linearen Abbildungen zwischen Vektorräumen der Form $\mathbb{K}^n$ durch Matrizen zu. Das liefert auch eine Verbindung zu linearen Gleichungssystemen, deren allgemeine Lösung mit Hilfe des Gauß'schen Eliminationsverfahrens besprochen wird. Damit können Bild und Kern von linearen Abbildungen zwischen Räumen der Form $\mathbb{K}^n$ explizit und algorithmisch bestimmt werden. Den Abschluss des Kapitels bildet der Algorithmus zur Feststellung von Invertierbarkeit und zur Bestimmung der inversen Matrix.

In Kapitel 4 wird die Entwicklung der allgemeinen Theorie wieder aufgenommen und die zentralen Begriffe der linearen Algebra (Erzeugnis, lineare (un)abhängigkeit, Basen) werden entwickelt. Der Austauschsatz von Steinitz führt schnell zum Dimensionsbegriff und zu den fundamentalen Resultaten über Basen, die in den meisten Beweisen verwendet werden. Als nächstes werden die Dimensionssätze für Teilräume, Summen und für lineare Abbildungen und das Konzept des Ranges von linearen Abbildungen und Matrizen besprochen. Zum Abschluss des Kapitels wird der Matrizenkalkül auf allgemeine (endlichdimensionale) Vektorräume ausgedehnt, so das auch in diesem allgemeinen Setting alle Konzepte explizit gemacht werden können. Die grundlegenden Resultate über Ähnlichkeit von Matrizen bilden den Abschluss des Kapitels.

In Kapitel 5 wird zunächst eine Anwendungen der linearen Algebra in der Codierungstheorie besprochen, die auch die Nützlichkeit endlicher Körper illustriert. Dann wenden wir uns der Anwendung der linearen Algebra auf geometrische Probleme zu. Die Art von linearer Algebra, die wir im Verlauf der Vorlesung besprochen haben, führt nicht zur (zum Teil schon aus der Schule bekannten) Euklidischen Geometrie, sondern zur “schwächeren” sogenannten affinen Geometrie. Für die Euklidische Geometrie werden innere Produkte benötigt, die wir, wie auch einiges aus der Euklidischen Geometrie, erst im zweiten Semester der Vorlesung besprechen werden. Das Ende von Kapitel 5 bilden zwei wichtige allgemeine Konstruktionen der linearen Algebra, nämlich Quotientenräume und ihre universelle Eigenschaft einerseits und Dualräume, duale Abbildungen und Annihilatoren andererseits.

In Kapitel 6 wird die allgemeine Theorie der Determinanten besprochen. Um später das charakteristische Polynom einer Matrix tatsächlich als Polynom (und nicht nur als Polynomfunktion auf $\mathbb{K}$) definieren zu können, wird die Theorie über kommutativen Ringen mit Einselement entwickelt, was kein zusätzlichen Schwierigkeiten mit sich bringt.

Kapitel 7 ist dem Begriff der Diagonalisierbarkeit gewidmet. Konzeptuell gesehen wird in diesem Kapitel vor allem der Zusammenhang zwischen Polynomen und linearen Abbildungen hergestellt, der im Studium von Normalformen eine fundamentale Rolle spielt. Aus didaktischen Gründen schließt wird dieses Thema aber erst in Kapitel 10 im letzten Semester des Zyklus behandelt.

In Kapitel 8 und 9 werden Normen und innere Produkte behandelt, somit kommen als Grundkörper in diesen Kapiteln nur $\mathbb{R}$ und $\mathbb{C}$ in Frage. Wie schon gesagt finden sich hier einerseits mehrere Bezüge zur Analysis (etwa in der Besprechung des Matrizenexponentials) und zur Funktionalanalysis. Ich habe mich bemüht jeweils die Bezüge zu den allgemeinen Konzepten herauszustreichen, etwa in der Analogie zwischen orthogonalem Komplement und Annihilator im Dualraum. Nach einer Besprechung spezieller Abbildungen und der Singulärwertzerlegung endet das Kapitel mit Resultaten über Quaternionen und ihre Rolle in der Beschreibung der speziellen orthogonalen Gruppe in Dimension 3 und 4.

In Kapitel 10 kehren wir zur Situation allgemeiner Körper zurück und besprechen Normalformen. Im ersten Teil des Kapitels steht die Rolle der Polynome im Mittelpunkt des Interesses. Insbesondere wird die allgemeine Primärzerlegung parallel zur Primfaktorzerlegung des Minimalpolynoms (oder des charakteristischen Polynoms) bewiesen. Im zweiten Teil des Kapitels wird angenommen, dass das charakteristische Polynom in ein Produkt von Polynomen ersten Grades zerfällt, also alle Eigenwerte im betrachteten Grundkörper liegen. Für solche Abbildungen wird die Existenz der Jordan'schen Normalform vollständig bewiesen. Die allgemeine Normalform für lineare Abbildungen über $\mathbb{R}$ wird kurz diskutiert.

Den Abschluss der Vorlesung bildet eine Diskussion der Grundlagen der multilinearen Algebra in Kapitel 11. Hier gilt es wieder die konzeptuelle Hürde einer neuen Denkweise zu überwinden. Ich habe mich bemüht, das durch eine detaillierte Diskussion der Konzepts von Natürlichkeit zu erleichtern. Die zweite Hälfte des Kapitels geht wieder über den Inhalt der Vorlesung hinaus und bietet Einblicke in den Bezug zu Algebren bis hin zu einer Konstruktion von Clifford–Algebren.

Zum Abschluss noch eine Bemerkung zur Literatur. An sich ist alles, was an Fakten für die Vorlesungen benötigt wird, in diesem Skriptum enthalten. Trotzdem wird es für viele Studierende hilfreich sein, manche der Inhalte auch in Lehrbüchern anzusehen. Es gibt eine Unmenge solcher Lehrbücher mit verschiedenen Ansprüchen und verschiedener Gewichtung des Materials. Nachdem ich bis jetzt noch kein Lehrbuch gefunden habe, dass mir wirklich gefällt, möchte ich an dieser Stelle keine konkrete Literaturempfehlung abgeben. Die Wahl des "passenden" Lehrbuches scheint mir weitgehend Geschmackssache zu sein. Inhaltlich sollte das meiste Material in fast jedem Lehrbuch, das abstrakte Vektorräume nicht ganz weglässt, zu finden sein.

Schließlich möchte ich mich noch bei vielen Studierenden bedanken, die in Verschiedenen Versionen der Vorlesungen durch Fragen oder Korrekturen zur Verbesserung der Inhalte beigetragen haben, besonders bei Herrn Lukas Prader.

KAPITEL 1

Einleitung

Der Einstieg in die Vorlesung baut unmittelbar auf den Schulstoff und die Inhalte der Vorlesung "Einführung in das mathematische Arbeiten" (im folgenden kurz EMA genannt) auf. Zu der EMA gibt es das Buch von Schichl und Steinbauer (Springer-Verlag) das im weiteren als [EMA] zitiert wird. Wir werden uns kurz mit mit linearen Gleichungssystemen in bis zu drei Variablen beschäftigen, die ja schon aus der Schule bekannt sind. Dabei werden wir aber einerseits strukturell Aspekte in den Vordergrund stellen (was direkt auf die algebraischen Kapitel der EMA aufbaut). Andererseits können wir die Systeme auch geometrisch interpretieren, was an die Vektorrechnung aus der Schule anschließt (und wir werden dabei keinen Anspruch auf vollständige Exaktheit stellen). Dabei werden wir einige typische Phänomene bemerken, deren genaue Formulierung und allgemeine Erklärung zu den Hauptzielen der Vorlesung gehört. Ich hoffe, dass diese Beispiele, in denen man sich vieles anschaulich vorstellen kann, auch einen erste Illustration für den Zusammenhang zwischen algebraischen und geometrischen Konzepten liefern. Schließlich sollen auch noch die Rolle und die Vorteile von Abstraktion illustriert werden.

Zum Verständnis des Kapitels wird es hilfreich sein, sich die grundlegenden Fakten über Vektorrechnung aus Schule und EMA in Erinnerung zu rufen.

Einfache lineare Gleichungssysteme

1.1. Gleichungen in einer Variablen. Das erste Beispiel, das wir behandeln, sieht auf den ersten Blick völlig trivial aus, es lässt sich aber eine ganze Menge daraus lernen. Gegeben zwei reelle Zahlen $a, b \in \mathbb{R}$ suchen wir die reellen Lösungen x der Gleichung

$$ax = b \tag{1.1}$$

Die offensichtliche Idee *"durch a dividieren"* zur Lösung dieser Gleichung wird nach erfolgreicher Absolvierung der EMA hoffentlich durch *"erst überlegen ob $a \neq 0$ ist"* eingebremst.

Wir müssen also die Fälle $a \neq 0$ und $a = 0$ unterscheiden. Für $a \neq 0$ können wir tatsächlich beide Seiten der Gleichung durch a dividieren bzw. (wie unter Mathematikern populärer) mit a^{-1} multiplizieren, um $x = a^{-1}b$ als eindeutige Lösung zu erhalten.

Im Fall $a = 0$ ist $0x = 0$, also bekommt die Gleichung die Form $0 = b$ und die Variable x "kommt eigentlich gar nicht mehr vor". Damit gibt es für $b \neq 0$ keine Lösung der Gleichung, während für $b = 0$ jedes $x \in \mathbb{R}$ eine Lösung ist.

Soweit so simpel, was gibt es also zu lernen? Erstens können wir das Ergebnis genauer betrachten: Wir haben einerseits den "regulären" Fall $a \neq 0$, in dem es für jedes $b \in \mathbb{R}$ eine eindeutige Lösung gibt. Andererseits gibt es den "bösen" Fall $a = 0$ in dem entweder zu viele (nämlich unendlich viele für $b = 0$) oder zu wenige (nämlich keine für $b \neq 0$) Lösungen existieren. Wir sehen auch warum das passiert, nämlich weil für $a = 0$ "in Wirklichkeit" eine Gleichung in Null Variablen vorliegt.

Zweitens können wir uns überlegen, was wir im Fall $a \neq 0$ tatsächlich getan haben, und was wir dafür gebraucht haben um das zu tun. Zunächst haben wir benutzt, dass es zu jedem $a \in \mathbb{R} \setminus \{0\}$ ein multiplikativ inverses Element a^{-1} gibt, und dass $a^{-1}(ax) = (a^{-1}a)x = 1x = x$ gilt. Eigentlich haben wir auch nur gezeigt, wie eine Lösung aussehen muss, wenn es eine gibt und die Verifikation, dass $a(a^{-1}b) = (aa^{-1})b = 1b = b$ gilt stillschweigend weggelassen.

Mehr als die Assoziativität der Multiplikation und die Existenz von multiplikativ inversen Elementen haben wir aber sicher nicht benutzt. Die Rechnung funktioniert also in jedem Körper $\mathbb{K}$, siehe Kapitel 4 von [EMA]. (Im Fall $a \neq 0$ rechnet man genau wie oben, $0x = 0$ folgt in jedem Körper aus $0x = (0 + 0)x = 0x + 0x$.) Auch wenn man momentan vielleicht nur an $\mathbb{R}$ interessiert ist, ist die Tatsache, dass man nicht mehr als die Körperstruktur braucht sehr nützlich. Für große Teile der linearen Algebra benehmen sich etwa (ganz im Gegensatz zur Analysis) die rationalen Zahlen $\mathbb{Q}$ genau so gut wie die reellen.

Die Gleichung $ax = b$ macht natürlich viel allgemeiner Sinn als nur über Körpern, dann wird aber das Lösungsverhalten wesentlich komplizierter. Das sieht man schon, wenn man die Gleichung über $\mathbb{Z}$ betrachtet (immerhin ein kommutativer Ring mit Einselement, in dem es keine Nullteiler gibt), also $a, b \in \mathbb{Z}$ annimmt und eine Lösung $x \in \mathbb{Z}$ sucht. So hat die Gleichung $2x = b$ für jedes gerade b eine eindeutige Lösung in $\mathbb{Z}$ aber für ungerades b gibt es keine ganzzahlige Lösung.

Betrachten wir nun den Fall von zwei Gleichungen in einer Variablen. Wir suchen also eine gemeinsame Lösung des Systems

$$\begin{cases} a_1 x = b_1 \\ a_2 x = b_2 \end{cases} \tag{1.2}$$

von Gleichungen. Falls $a_1 = 0$ und $b_1 \neq 0$ oder $a_2 = 0$ und $b_2 \neq 0$ ist, dann hat schon eine der beiden Gleichungen keine Lösung, also gibt es auch keine Lösung des Systems (1.2). Anderseits kann man eine Gleichung, in der sowohl a als auch b gleich Null ist, einfach weglassen. Damit können wir uns auf den Fall beschränken, dass a_1 und a_2 beide ungleich Null sind. Dann hat aber die erste Gleichung die eindeutige Lösung $x = a_1^{-1}b_1$ und die zweite Gleichung die eindeutige Lösung $x = a_2^{-1}b_2$. Also hat das System (1.2) genau dann eine Lösung, wenn diese beiden Werte übereinstimmen.

Nun bedeutet aber $\frac{b_1}{a_1} = \frac{b_2}{a_2}$ natürlich $b_2 = \frac{a_2}{a_1}b_1$ und offensichtlich gilt $a_2 = \frac{a_2}{a_1}a_1$. Damit bedeutet die Bedingung für die Lösbarkeit genau, dass man die zweite Gleichung durch Multiplikation der ersten Gleichung mit $\frac{a_2}{a_1}$ erhält. Etwas großzügig gesagt, hat das System (1.2) also genau dann eine Lösung, wenn es "in Wirklichkeit" nur aus einer Gleichung besteht.

1.2. Gleichungen in zwei Variablen. Betrachten wir zunächst eine Gleichung in zwei Variablen. Für $a_1, a_2, b \in \mathbb{R}$ suchen wir alle Lösungen (x_1, x_2) der Gleichung

$$a_1x_1 + a_2x_2 = b \tag{1.3}$$

Ist $a_1 = a_2 = 0$, dann gibt es analog wie in 1.1 entweder keine Lösung (wenn $b \neq 0$) oder jedes Paar (x_1, x_2) ist eine Lösung (wenn $b = 0$). Ist mindestens eines der a's ungleich Null, dann läßt sich die Gleichung algebraisch ganz leicht lösen. Ist $a_1 \neq 0$, dann subtrahieren wir von beiden Seiten von (1.3) a_2x_2 und erhalten $a_1x_1 = b - a_2x_2$. Wie im Fall einer Variablen können wir jetzt durch a_1 dividieren, um $x_1 = a_1^{-1}(b - a_2x_2)$ zu erhalten. Wählen wir nun für x_2 ein beliebiges Element $t \in \mathbb{R}$, dann erhalten wir $x_1 = a_1^{-1}(b - a_2t)$, also muss jede Lösung der Gleichung von der Form $(a_1^{-1}(b - a_2t), t)$ für eine Zahl $t \in \mathbb{R}$ sein. Umgekehrt sieht man sofort, dass jedes solche Paar die Gleichung löst, also haben wir alle Lösungen gefunden. Auch diese Rechnung funktioniert ohne Änderungen über einem beliebigen Körper.

Es gibt also wieder zwei Fälle. Im "bösen" Fall $a_1 = a_2 = 0$ ist "in Wirklichkeit" wieder eine Gleichung in Null Variablen geben, die entweder zu wenige oder zu viele Lösungen hat. Im "regulären" Fall tritt in der Beschreibung der Lösungsmenge einer Gleichung in zwei Variablen ein freier Parameter auf.

Dieses Verhalten ist aus der geometrischen Deutung so einer Gleichung, die schon aus der Schule bekannt ist, vollkommen einsichtig. Betrachtet man (x_1, x_2) als die Koordinaten eines Punktes in der Ebene $\mathbb{R}^2$, dann ist für $a_1 \neq 0$ oder $a_2 \neq 0$ die Lösungsmenge von (1.3) eine Gerade. Dabei kann man (a_1, a_2) als den Normalvektor der Geraden deuten, während b (grob gesagt) angibt, wie weit die Gerade vom Nullpunkt weg verschoben ist. Die oben beschriebene Lösungsmethode findet dann genau die Parameterdarstellung der Geraden. Ist nämlich $a_1 \neq 0$, dann ist offensichtlich $(a_1^{-1}b, 0)$ ein Punkt der auf der Geraden liegt, während man als Richtungsvektor zum Beispiel $(-a_1^{-1}a_2, 1)$ verwenden kann und daraus ergibt sich sofort die Lösung von oben als $(a_1^{-1}b, 0) + t(-a_1^{-1}a_2, 1)$.

Betrachten wir nun ein System von zwei Gleichungen in zwei Variablen. Für gegebene a_{11}, a_{12}, a_{21}, a_{22}, b_1 und $b_2 \in \mathbb{R}$ suchen wir alle Lösungen (x_1, x_2) des Systems

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \tag{1.4}$$

Wie zuvor können wir annehmen, dass in jeder Zeile mindesten ein a ungleich Null ist (sonst gibt es entweder keine Lösung dieser einen Gleichung, oder man kann die Gleichung einfach weglassen). Man kann dieses System algebraisch lösen, indem man es auf ein System von einer Gleichung in einer Variablen reduziert. Ist etwa $a_{11} \neq 0$, dann kann man die erste Gleichung benutzen um wie oben $x_1 = a_{11}^{-1}(b_1 - a_{12}x_2)$ zu erhalten. Setzt man das in die zweite Gleichung ein, dann erhält man nach kurzer Umformung

$$(a_{22} - a_{11}^{-1}a_{21}a_{12})x_2 = b_2 - a_{11}^{-1}a_{21}b_1. \tag{1.5}$$

Damit haben wir eine Gleichung erhalten, in der nur noch x_2 vorkommt, und die wir nach 1.1 vollständig verstehen:

(i) Ist $a_{22} - a_{11}^{-1}a_{21}a_{12} \neq 0$, dann liefert (1.5) eine eindeutige Lösung für x_2.
(ii) Ist $a_{22} - a_{11}^{-1}a_{21}a_{12} = 0$ aber $b_2 - a_{11}^{-1}a_{21}b_1 \neq 0$, dann hat (1.5) keine Lösung, also hat auch (1.4) keine Lösung.
(iii) Ist $a_{22} - a_{11}^{-1}a_{21}a_{12} = 0$ und $b_2 - a_{11}^{-1}a_{21}b_1 = 0$, dann löst jedes x_2 (1.5).

Setzt man das in $x_1 = a_{11}^{-1}(b_1 - a_{12}x_2)$ ein, dann sieht man, dass das System (1.4) im Fall (i) eine eindeutige Lösung und im Fall (ii) keine Lösung besitzt, während im Fall (iii) die Lösungsmenge wieder durch einen freien Parameter beschrieben wird.

Diese Lösungsverhalten läßt sich wieder am besten mit der geometrischen Deutung von oben verstehen. Jede der beiden Gleichungen in (1.4) beschreibt eine Gerade in der Ebene $\mathbb{R}^2$. Die wesentliche Unterscheidung ist nun, ob die beiden Geraden parallel (also ihre Normalvektoren vielfache von einander) sind oder nicht. Im Fall $a_{11} \neq 0$ reduziert sich das aber gerade auf die Bedingung, die wir oben gefunden haben. Ist nämlich $(a_{21}, a_{22}) = \lambda(a_{11}, a_{12})$ ist, dann folgt aus der Betrachtung der ersten Koordinaten sofot $\lambda = \frac{a_{21}}{a_{11}}$, also sind die beiden Vektoren genau dann Vielfache von einander, wenn $a_{22} = \frac{a_{21}}{a_{11}}a_{12}$ gilt, was genau die Bedingung für die Fälle (ii) und (iii) von oben liefert.

Der Fall (i) von oben tritt also genau dann ein, wenn die beiden Geraden nicht parallel sind, und dann gibt es einen eindeutigen Schnittpunkt, dessen Kooridnaten die eindeutige Lösung von oben bilden.

Sind die beiden Normalvektoren Vielfache voneinander, dann können die beiden Geraden entweder zusammenfallen (und dann ist die Lösungsmenge genau diese Gerade) oder verschiedene parallele Gerade sein, die natürlich keinen Schnittpunkt haben. Die Bedingung für das Zusammenfallen ist (im Fall $a_{11} \neq 0$) natürlich gerade, dass auch noch $b_2 = \lambda b_1$ gilt, und das unterscheidet die Fälle (ii) und (iii) von oben. Im Fall (iii) (in dem die Geraden zusammenfallen) ist die zweite Gleichung in (1.4) einfach das λ–fache der ersten Gleichung, also hat man "in Wirklichkeit" nur eine Gleichung gegeben.

Für diejenigen, die das in der Schule gehört haben, lohnt es sich zu bemerken, dass man die Bedingung $a_{22} - a_{11}^{-1}a_{21}a_{12} \neq 0$ von oben umformen kann. Weil $a_{11} \neq 0$ gilt, ist sie äquivalent zu $a_{11}a_{22} - a_{21}a_{12} \neq 0$. Man erhält also genau die Determinante der Koeffizientenmatrix.

Der Fall mehr als zwei Gleichungen in zwei Variablen können wir auch gut verstehen. Betrachten wir etwa drei Gleichungen in zwei Variablen, dann können wir die erste Gleichung benutzen, um eine der Variablen auszudrücken. Setzt man in die anderen beiden Gleichungen ein, dann erhält man zwei Gleichungen in einer Variablen und damit einen Fall, den wir bereits in 1.1 untersucht haben. Es gibt also nur dann eine Lösung, wenn die beiden verbleibenden Gleichungen Vielfache voneinander sind. Mann kann sich direkt davon ueberzeugen, dass das genau dann der Fall ist, wenn man eine der drei ursprünglichen Gleichung als Summe von Vielfachen der beiden anderen Gleichungen schreiben kann.

1.3. Eine alternative Interpretation. Es gibt noch eine ganz andere Interpretation eines Systems von zwei Gleichungen in zwei Variablen. Betrachten wir nochmals das System

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases} \tag{1.4}$$

In 1.2 haben wir die Variablen x_1 und x_2 als Koordinaten eines Vektors in der Ebene $\mathbb{R}^2$ interpretiert. Jede der beiden Gleichungen haben wir dann als Beschreibung einer Gerade in $\mathbb{R}^2$ interpretiert, wobei jeweils (a_{i1}, a_{i2}) ein Normalvektor der Geraden war und b_i die Lage bestimmt. Man kann das ganze System aber auch direkt in Vektorform lesen, nämlich als

$$x_1\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{1.6}$$

Man fasst also die Koeffizienten a_{ij} in anderer Art als zuvor als Vektoren auf (in Matrixsprache betrachten wir die Spalten statt der Zeilen) und auch die b_i werden in offensichtlicher Weise als Vektor betrachtet. Das System kann man dann so lesen, dass man den Vektor der b's als Summe eines Vielfachen des ersten Spaltenvektors und eines Vielfachen des zweiten Spaltenevektors schreiben möchte wobei die beiden Faktoren gesucht sind.

Auch hier kann man das Lösungsverhalten gut geometrisch verstehen. Die entscheidende Frage ist wieder ob die beiden Vektoren $v = \binom{a_{11}}{a_{21}}$ und $w = \binom{a_{12}}{a_{22}}$ Vielfache voneinander sind oder nicht. Die Vielfachen eines Vektors (ungleich dem Nullvektor) bilden eine Gerade durch den Ursprung und die Frage ist gerade, ob v und w die gleiche Gerade liefern oder nicht. Sind die Geraden verschieden, dann kann man jeden Vektor

$\binom{b_1}{b_2}$ eindeutig als Summe eines Vielfachen von v und eines Vielfachen von w schrieben. (Dazu verschiebt man einfach eine der Geraden durch den Punkt $\binom{b_1}{b_2}$ und schneidet sie mit der anderen Geraden.) Das entspricht also genau dem "regulären" Fall (i) aus 1.2.

Nehmen wir andererseits an, dass $w = \lambda v$ gilt. Dann ist natürlich $x_1 v + x_2 w = (x_1 + \lambda x_2)v$ und das liegt immer auf der Geraden durch den Nullpunkt, die v erzeugt. Liegt $\binom{b_1}{b_2}$ nicht auf dieser Geraden, dann ist das System nicht lösbar (Fall (ii) aus 1.2). Liegt $\binom{b_1}{b_2}$ aber auf der Geraden, dann kann man diesen Vektor als μv für eine fixe Zahl $\mu \in \mathbb{R}$ schreiben. Dann wird unser System aber gerade zu $x_1 + \lambda x_2 = \mu$ und das liefert für beliebig gewähltes x_2 eine eindeutige Lösung für x_1. Man erhält also eine Lösungsmenge, die von einem freien Parameter abhängt, wie es dem Fall (iii) von 1.2 entspricht.

Wie nach den Ergebnissen von 1.2 zu erwarten ist, kann man direkt verifizieren, dass die Vektoren $\binom{a_{11}}{a_{21}}$ und $\binom{a_{12}}{a_{22}}$ genau dann Vielfache von einander sind, wenn $a_{11}a_{22} - a_{12}a_{21} = 0$ gilt. Man sieht also, dass man aus den Zeilen und aus den Spalten der Matrix jeweils auf die gleiche Bedingung für die Regularität des Geleichungssystems kommt. Phänomene wie dieses konzeptuell und allgemein zu verstehen ist ein wichtiges Ziel dieser Vorlesung.

1.4. Verallgemeinerungen. Die Lösungsmethode und die "zeilenweise" Interpretation eines Gleichungssystems aus 1.2 und auch die "spaltenweise" Interpretation aus 1.3 lassen sich auf Fälle mit mehreren Gleichungen bzw. mehreren Unbekannten verallgemeinern. Betrachten wir eine Gleichung mit drei Variablen, also

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1.$$

Nehmen wir wieder $a_{11} \neq 0$ an, und können dann die Gleichung zu

$$x_1 = a_{11}^{-1}(b_1 - a_{12}x_2 - a_{13}x_3)$$

umformen. Ist das die einzige Gleichung, dann können wir für x_2 und x_3 beliebige Werte $s, t \in \mathbb{R}$ einsetzen, und erhalten dann einen eindeutigen Wert für x_1. Die allgemeine Lösung hängt dann also von zwei Parametern ab. Kommen weitere Gleichungen dazu, dann kann man in diesen für x_1 einsetzen und dann wie zuvor weiter vorgehen.

Geometrisch beschreibt eine Gleichung in 3 Variable eine Ebene im Raum $\mathbb{R}^3$, was die beiden freien Parameter in der Lösung erklärt. Man kann (a_{11}, a_{12}, a_{13}) wieder als den Normalvektor der Ebene interpretieren und b_1 als Beschriebung der Lage der Ebene. Kommen weitere Gleichungen hinzu, dann kann man das Lösungsverhalten wieder durch das Schnittverhalten von Ebenen im $\mathbb{R}^3$ verstehen. Insbesondere erhält man für zwei Gleichugen in 3 Variablen im "regulären Fall" (nicht parallele Ebenen) eine Gerade als Lösungsmenge, also eine Beschriebung der Lösungsmenge mit einem freien Parameter. Für parallele Ebenen erhält man wieder entweder gar keine Lösung oder eine "zu große" Lösungsmenge (mit zwei freien Parametern in der Beschriebung). Bei drei Gleichungen in drei Variablen erhält man im regulären Fall eine eindeutige Lösung.

Für ein System von zwei Gleichungen in drei Variablen

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \end{cases}$$

funktioniert auch wieder die "spaltenweise" Interpretation aus 1.3. Man versucht hier den Vektor $\binom{b_1}{b_2}$ in der Ebene $\mathbb{R}^2$ als Summe von Vielfachen von drei gegebenen Vektoren (nämlich $v_i = \binom{a_{i1}}{a_{i2}}$ für $i = 1, 2, 3$) zu schreiben, wobei der Lösungen der Gleichung gerade die benötigten Faktoren sind. Der reguläre Fall ist hier, dass die Vektoren nicht alle drei parallel sind. Nehmen wir etwa an, dass v_1 und v_2 nicht parallel sind, dann wissen wir aus 1.3, dass man v_3 eindeutig als $\lambda_1 v_1 + \lambda_2 v_2$ und $\binom{b_1}{b_2}$ eindeutig als $\mu_1 v_1 + \mu_2 v_2$ schreiben kann. Setzt man den ersten Teil ein, dann erhält man

$$x_1 v_2 + x_2 v_2 + x_3 v_3 = (x_1 + \lambda_1 x_3) v_1 + (x_2 + \lambda_2 x_3) v_2$$

und damit ist unser System äquivalent zu

$$x_1 + \lambda_1 x_3 = \mu_1 \qquad x_2 + \lambda_2 x_3 = \mu_2.$$

Hier kann man offensichtlich x_3 beliebig wählen und erhält dann eindeutige Lösungen für x_1 und x_2.

Analog kann man ein System von drei Gleichungen in drei Unbekannten so interpretieren, dass man einen Vektor im Raum $\mathbb{R}^3$ als Summe von Vielfachen von drei gegebenen Vektoren schreiben kann, wobie die notwendigen Faktoren gesucht sind. Liegen die drei Vektoren nicht in einer Ebene durch den Nullpunkt, dann geht das immer eindeutig, ansonsten gibt es wieder entweder zu viele oder zu wenige Lösungen.

Natürlich gibt es keinen Grund, die Zahl der Gleichungen und/oder der Variablen auf zwei oder drei zu beschränken. Tatsächlich spielen lineare Gleichungssytems mit hunderten oder tausenden Gleichungen und Variablen eine zentrale Rolle in einer Fülle von Anwendungen der Mathematik. Man hat in diesem Fall zwar keine anschauliche Interpretation mehr, kann aber so wie in $\mathbb{R}^2$ und $\mathbb{R}^3$ auch in $\mathbb{R}^{1000}$ oder (dann eigentlich schon wieder leichter) in $\mathbb{R}^n$ mit beliebigem n rechnen. Auch die rellen Zahlen spielen hier keine besondere Rolle, es funktioniert vieles genau so gut über beliebigen Körpern. Im Endeffekt kann man dann die geometrische Anschauung (bis zu einem gewissen Grad) auf allgemeine Dimensionen und allgemeine Körper übertragen.

KAPITEL 2

Vektorräume und Lineare Abbildungen

Nachdem wir nun mit etwas anschaulicher Intuition gewappnet sind, werden wir in diesem Kapitel die zentralen Konzepte der linearen Algebra in allgemeiner (abstrakter) Form einführen, wobei die grundlagende Art des Zuganges schon aus der EMA bekannt ist. Ich möchte gleich vorweg bemerken, dass Vektorräume in der Hierarchie der algebraischen Strukturen eher am komplizierten Ende stehen. Trotzdem ist die Theorie der Vektorräume einer der einfachsten Teile der Algebra. Neben ihrer breiten Anwendbarkeit ist das einer der Gründe, warum die lineare Algebra in so einer frühen Phase des Studiums behandelt wird.

Vektorräume

2.1. Körper. Am Anfang der Vektorrechnung stehen zwei ganz verschiedene Operationen, die auch in unsere Besprechung von Gleichungssystemen in Kapitel 1 die Hauptrolle gespielt haven. Wir haben nämlich einerseits Gleichungen bzw. Vektoren zueinander addiert und andererseits Vielfache von Gleichungen bzw. Vektoren gebildet, d.h. sie mit Skalaren (rellen Zahlen) multipliziert. (Kurz ist über den Begriff des Normalvektors auch noch das innere Produkt aufgetaucht, das wir aber erst viel spaeter allgemein besprechen werden.) Jedenfalls müssen wir aber zwei verschiedene Arten von Objekten, nämlich Skalare und Vektoren, gemeinsam behandeln.

Wir werden uns hauptsächlich für den Fall interessieren, dass die Skalare reelle oder komplexe Zahlen und die Vektoren Elemente von $\mathbb{R}^n$ bzw. von $\mathbb{C}^n$ sind und man sollte diese Beispiele immer vor Augen haben. Wie in vielen Teilgebieten der Mathematik üblich und schon aus den passenden Teilen der EMA bekannt, werden wir die Theorie aber in einem allgemeineren Rahmen entwickeln. Neben dem offensichtlich Vorteil, dass die allgemeineren Resultate später nützlich sein werden, ist diese Vorgehensweise auch aus anderen Gründen vorzuziehen. Erstens ist es oft sehr hilfreich, klar zu sehen, welche Eigenschaften wirklich benötigt werden um Resultate zu beweisen. Zweitens zwingt der allgemeinere Rahmen automatisch zu einer konzeptuelleren Vorgangsweise, was oft zu Vereinfachungen führt, sogar für einfache Beispiele. Schließlich sind die anschaulichen Beispiele manchmal auch irreführend, weil es schwer ist, in diesen Situationen vorhandene zusätzliche Strukturen (etwa das innere Produkt) zu "vergessen".

Schließlich sei noch bemerkt, dass das Finden der "richtigen" Begriffe und des richtigen Abstraktionsniveaus ein langwieriger Prozess war. Eine gute Wahl von Begriffen ist oft schon ein großer Schritt zu einer guten mathematischen Theorie. Daher sollte man sich als AnfängerIn nicht darüber wundern, dass die genaue Wahl der Begriffe oft nicht unmittelbar einsichtig ist. Die Frage "Könnte man das nicht auch ein bisschen anders machen?" ist natürlich berechtigt, aber üblicherweise wurde das schon versucht und hat sich als weniger nützlich erwiesen.

Im Fall der linearen Algebra ist es günstig, als Skalare die Elemente eines beliebigen Körpers $\mathbb{K}$ zuzulassen. Dann definiert man den Begriff eines "Vektorraumes über einem Körper", und für gegebenes $\mathbb{K}$ sind dann die Vektorräume über $\mathbb{K}$ der richtige Rahmen um lineare Algebra betreiben zu können. Kommen wir zunächst zur Definition eines Körpers, die schon aus der EMA bekannt ist (siehe Abschnitt 5.4.1 in [EMA]):

Definition 2.1. Ein *Körper* ist eine Menge $\mathbb{K}$ zusammen mit zwei Operationen $+ : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ und $\cdot : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ mit folgenden Eigenschaften:

(K1) $(r + s) + t = r + (s + t) \quad \forall r, s, t \in \mathbb{K}$.
(K2) $r + s = s + r \quad \forall r, s \in \mathbb{K}$.
(K3) Es gibt ein Element $0 \in \mathbb{K}$, sodass $r + 0 = r \quad \forall r \in \mathbb{K}$.
(K4) Für jedes $r \in \mathbb{K}$ gibt es ein Element $(-r) \in \mathbb{K}$, sodass $r + (-r) = 0$ gilt.
(K5) $(r \cdot s) \cdot t = r \cdot (s \cdot t) \quad \forall r, s, t \in \mathbb{K}$
(K6) $r \cdot s = s \cdot r \quad \forall r, s \in \mathbb{K}$
(K7) Es gibt ein Element $1 \in \mathbb{K} \setminus \{0\}$, sodass $1 \cdot r = r \quad \forall r \in \mathbb{K}$.
(K8) Für jedes $r \in \mathbb{K} \setminus \{0\}$ gibt es ein Element $r^{-1} \in \mathbb{K}$, sodass $r \cdot r^{-1} = 1$ gilt.
(K9) $r \cdot (s + t) = (r \cdot s) + (r \cdot t) \quad \forall r, s, t \in \mathbb{K}$.

Auf einem Körper hat man also zwei Operationen, die wir wie gewohnt als Addition und Multiplikation bezeichnen werden. Für diese Operationen verlangen wir Assoziativität, Kommutativität, Distributivität und die Existenz von neutralen Elementen (die wir wie gewohnt mit 0 bzw. 1 bezeichnen) und von inversen Elementen, wie man sie von den Zahlbereichen $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ kennt. Der wesentliche Punkt hier ist, dass man keine weiteren Eigenschaften verlangt.

Wir werden ab sofort auch die Multiplikation meist nicht mit einem eigenen Symbol bezeichnen, sondern für $r \cdot s$ einfach rs schreiben. Die Bedingung $1 \neq 0$, die in (K7) versteckt ist, dient nur dazu, zu vermeiden, dass die einelementige Menge $\{0\}$ mit den Operationen $0 + 0 = 0$ und $0 \cdot 0 = 0$ ein Körper ist. (Sonst müssten wir in Zukunft überall "Sei $\mathbb{K}$ ein Körper mit mindestens zwei Elementen" schreiben.)

Aus der EMA ist auch schon bekannt, dass (K1)–(K4) genau sagen, dass $(\mathbb{K}, +)$ eine kommutative Gruppe ist während nach (K5)–(K8) auch $(\mathbb{K} \setminus \{0\}, \cdot)$ eine kommutative Gruppe ist. Die Bedingungen (K1)–(K6) und (K9) sagen genau, dass $(\mathbb{K}, +, \cdot)$ ein kommutativer Ring ist, und so weiter. Solange man sich aber mit diesen Begriffen nicht vertraut fühlt und man keine Resultate über Gruppen oder Ringe verwenden will, sind das aber nur (relativ leere) Worte.

Aus der EMA ist ebenfalls bekannt, dass neutrale und inverse Elemente in einer Gruppe eindeutig bestimmt sind. Dieses Resultat kann man natürlich auf die beiden Operationen in einem Körper anwenden und erhält die folgenden Resultate, die für $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ natürlich schon aus der Schule bekannt sind. Einer Wiederholung der Beweise aus der EMA ist an dieser Stelle eine gute Übung.

Lemma 2.1. *Sei* $(\mathbb{K}, +, \cdot)$ *ein Körper. Dann gilt:*
(1) Ist $s \in \mathbb{K}$ *so, dass* $s + r = r$ *für ein Element* $r \in \mathbb{K}$ *gilt, dann ist* $s = 0$.
(2) Ist $s \in \mathbb{K} \setminus \{0\}$ *so, dass* $s \cdot r = r$ *für ein Element* $r \in \mathbb{K} \setminus \{0\}$ *gilt, dann ist* $s = 1$.
(3) Gilt für $r, \hat{r} \in \mathbb{K}$, *dass* $r + \hat{r} = 0$ *ist, dann ist* $\hat{r} = -r$.
(4) Gilt für $r, \hat{r} \in \mathbb{K} \setminus \{0\}$, *dass* $r \cdot \hat{r} = 1$ *ist, dann ist* $\hat{r} = r^{-1}$.

Beispiele 2.1. (1) Aus der EMA ist bekannt, dass die üblichen Operationen auf $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ die in (K1)–(K9) verlangten Eigenschaften haben, also sind diese Zahlbereiche Körper.

(2) Die üblichen Operationen auf $\mathbb{Z}$ erfüllen alle Bedingungen aus Definition 2.1, außer (K8). Da etwa $2 \in \mathbb{Z}$ auch tatsächlich kein multiplikativ inverses Element besitzt ist $(\mathbb{Z}, +, \cdot)$ *kein* Körper.

(3) Betrachte die Menge $\mathbb{Z}_2 := \{0,1\}$ mit den Operationen $+$ und $\cdot$, die bestimmt sind durch $0+r = r+0 = r$, $0\cdot r = r\cdot 0 = 0$, $1\cdot r = r\cdot 1 = r$ und $1+1 = 0$. Dann verifiziert man leicht direkt, dass $(\mathbb{Z}_2,+,\cdot)$ ein Körper ist, siehe Abschnitt 5.4.4 in [EMA]. Analog lernt man für eine Primzahl p in der EMA auch den Körper $\mathbb{Z}_p$ kennen, der p Elemente hat.

Es gibt noch viele andere (teilweise sehr viel exotischere) Beispiele von Körpern, aber für unsere Zwecke werden die obigen für's erste ausreichen.

2.2. Vektorräume. Die Definitions eines Vektorraumes über einem Körper folgt im wesentlichen dem aus der EMA bekannten Schema. Man betrachtet eine Menge mit Operationen, von denen bestimmte Grundeigenschaften gefordert werden. Wie bereits gesagt, wollen wir Vektoren addieren und mit Skalaren multiplizieren können, also werden wir vorgegebene Skalare und zwei Operationen benötigen.

Definition 2.2. Sei $\mathbb{K}$ ein Körper. Ein *Vektorraum über* $\mathbb{K}$ (oder ein $\mathbb{K}$–*Vektorraum*) ist eine Menge V zusammen mit zwei Operationen $\oplus : V\times V\to V$ und $\odot : \mathbb{K}\times V\to V$ mit folgenden Eigenschaften:

(V1) $(u\oplus v)\oplus w = u\oplus(v\oplus w) \quad \forall u,v,w\in V$.
(V2) $u\oplus v = v\oplus u \quad \forall u,v\in V$.
(V3) Es gibt ein Element $0_V\in V$, sodass $v\oplus 0_V = v \quad \forall v\in V$.
(V4) Zu jedem $v\in V$ gibt es ein Element $\ominus v\in V$, sodass $v\oplus(\ominus v) = 0_V$ gilt.
(V5) $(r+s)\odot v = (r\odot v)\oplus(s\odot v) \quad \forall r,s\in\mathbb{K}, \forall v\in V$.
(V6) $r\odot(v\oplus w) = (r\odot v)\oplus(r\odot w) \quad \forall r\in\mathbb{K}, \forall v,w\in V$.
(V7) $(rs)\odot v = r\odot(s\odot v) \quad \forall r,s\in\mathbb{K}, \forall v\in V$.
(V8) $1\odot v = v \quad \forall v\in V$.

Die Bedingungen (V1)–(V4) sagen, dass $(V,\oplus)$ eine kommutative Gruppe ist. (Hoffentlich wird langsam sichtbar, dass es nützlich ist, die Konzepte aufbauend zu entwickeln.) Genau wie in Lemma 2.1 folgt, dass das neutrale Element von $\oplus$ durch (V3) und die inversen Elemente durch (V4) eindeutig bestimmt sind, was die Bezeichnungen 0_V und $\ominus v$ rechtfertigt.

Die Bedingungen (V5)–(V8) beschreiben den Zusammenhang zwischen den Rechenoperationen auf $\mathbb{K}$ und der Multiplikation von Elementen von V mit Elementen von $\mathbb{K}$. Wenn man (wie wir das in Kürze tun werden) die Addition auf V ebenfalls mit $+$ statt mit $\oplus$ und die Skalarmultiplikation mit $\cdot$ (oder ganz ohne Symbol) statt mit $\odot$ bezeichnet, dann sehen (V5)–(V7) wie Assoziativ- bzw. Distributivgesetze aus, sie betreffen aber jeweils zwei *verschiedene* Operationen. Die Bedingung (V8) ist eine intuitiv völlig einsichtige Forderung, interessant ist am ehesten, dass sie nicht aus den anderen Bedingungen folgt, sondern gefordert werden muss.

Beispiel 2.2. (0) Für einen beliebigen Körper $\mathbb{K}$ ist $V = \{0\}$ mit den offensichtlichen Operationen $0\oplus 0 = 0$, $r\odot 0 = 0$ für $r\in\mathbb{K}$ ein $\mathbb{K}$–Vektorraum, der *triviale* $\mathbb{K}$–*Vektorraum*.

(1) Sei wiederum $\mathbb{K}$ beliebig, und setze $V := \mathbb{K}$ und definiere $r\oplus s := r+s$ und $r\odot s = rs$. Dann sind (V1)–(V4) und (V7) genau die Bedingungen (K1)–(K4) und (K5) aus Definition 2.1, (V5) und (V6) folgen direkt aus dem Distributivgesetz (K9) für $\mathbb{K}$ und (V8) gilt nach Definition des Einselements. Also ist der Körper $\mathbb{K}$ selbst auch ein $\mathbb{K}$–Vektorraum.

(2) Der Körper $\mathbb{C}$ der komplexen Zahlen enthält die reellen Zahlen $\mathbb{R}$ als Teilkörper. Insbesondere können wir die Multiplikation von $\mathbb{C}$ zu einer Abbildung $\odot : \mathbb{R}\times\mathbb{C}\to\mathbb{C}$

einschränken. Wie in Beispiel (1) definieren wir $\oplus : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ durch $z \oplus w := z + w$, und wie dort folgt, dass diese Definitionen $\mathbb{C}$ zu einem $\mathbb{R}$–Vektorraum machen.

Ganz analog enthalten die reellen Zahlen $\mathbb{R}$ den Körper $\mathbb{Q}$ der rationalen Zahlen als Teilkörper. Verwenden wir die übliche Addition auf $\mathbb{R}$ als $\oplus$ und die Einschränkung der üblichen Multiplikation als $\odot : \mathbb{Q} \times \mathbb{R} \to \mathbb{R}$, dann macht das $\mathbb{R}$ zu einem $\mathbb{Q}$–Vektorraum.

(3) Für einen beliebigen Körper $\mathbb{K}$ und $n \in \mathbb{N}$ betrachte $\mathbb{K}^n := \{(v_1, \dots, v_n) : v_i \in \mathbb{K}\}$, die Menge aller geordneten n–Tupel von Elementen von $\mathbb{K}$. Wir werden solche n–Tupel im weiteren oft als Spaltenvektoren schreiben, und definieren

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \oplus \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} := \begin{pmatrix} v_1+w_1 \\ \vdots \\ v_n+w_n \end{pmatrix} \qquad r \odot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} rv_1 \\ \vdots \\ rv_n \end{pmatrix}.$$

Das sind die offensichtlichen Verallgemeinerungen der für $\mathbb{R}^2$, $\mathbb{R}^3$ und $\mathbb{R}^n$ aus Schule und EMA bekannten Operationen. Nun folgen (V1) und (V2) direkt indem man (K1) und (K2) auf jede Koordinate anwendet. Aus (K3) folgt, dass (V3) für $0_V = (0, \dots, 0)$ gilt und aus (K4) folgt (V4) für $\ominus(v_1, \dots, v_n) = (-v_1, \dots, -v_n)$. Für $v = (v_1, \dots, v_n)$, $w = (w_1, \dots, w_n)$ und $r \in \mathbb{K}$ gilt dann

$$r \odot (v \oplus w) = \begin{pmatrix} r(v_1+w_1) \\ \vdots \\ r(v_n+w_n) \end{pmatrix} \qquad (r \odot v) \oplus (r \odot w) = \begin{pmatrix} (rv_1)+(rw_1) \\ \vdots \\ (rv_n)+(rw_n) \end{pmatrix}$$

und diese Elemente stimmen nach dem Distributivgesetz (K9) für $\mathbb{K}$ überein, also gilt (V6). Analog verifiziert man (V5) und (V7), währen (V8) offensichtlich gilt. Damit ist $\mathbb{K}^n$ ein $\mathbb{K}$–Vektorraum, und das wir für uns das wichtigste Beispiel eines Vektorraumes sein.

(4) Sei wiederum $\mathbb{K}$ ein beliebiger Körper, und X eine beliebige Menge. Sein $\mathcal{F}(X, \mathbb{K})$ die Menge aller Funktionen $f : X \to \mathbb{K}$. Darauf definieren wir nun Operationen durch

$$(f \oplus g)(x) := f(x) + g(x) \qquad (r \odot f)(x) := r(f(x))$$

für $f, g : X \to \mathbb{K}$, $x \in X$ und $r \in \mathbb{K}$. Man nennt diese Operationen die *punktweisen* Operationen auf $\mathcal{F}(X, \mathbb{K})$, weil man Funktionen einfach addiert bzw. mit einem Skalar multipliziert indem man ihre Werte in jedem Punkt addiert bzw. mit dem Skalar multipliziert. Um zu verifizieren, dass diese Operationen $\mathcal{F}(X, \mathbb{K})$ zu einem $\mathbb{K}$–Vektorraum machen, muss man sich nur in Erinnerung rufen, dass zwei Funktionen $X \to \mathbb{K}$ genau dann übereinstimmen, wenn ihre Werte in allen Punkten von X übereinstimmen. Um etwa (V1) zu verifizieren genügt es also zu zeigen, dass für beliebige Funktionen $f, g, h : X \to \mathbb{K}$ und jedes $x \in X$ die Gleichung $((f \oplus g) \oplus h)(x) = (f \oplus (g \oplus h))(x)$. Nach Definition ist aber die linke Seite dieser Gleichung gegeben durch

$$(f \oplus g)(x) + h(x) = (f(x) + g(x)) + h(x),$$

und ganz analog sieht man, dass die rechte Seite gleich $f(x) + (g(x) + h(x))$ ist. Diese beiden Ausdrücke sind aber nach (K1) gleich. Völlig analog folgt (V2) aus (K2), (V3) für die *Nullfunktion* $0_{\mathcal{F}}$, die definiert ist durch $0_{\mathcal{F}}(x) = 0$ für alle $x \in X$ aus (K3) und (V4) aus (K4), wobei $\ominus f$ definiert ist durch $\ominus f(x) = -f(x)$ für alle $x \in X$. (V8) ist aus der Definition offensichtlich, und die restlichen Eigenschaften (V5)–(V7) werden analog zu (V1)–(V4) verifiziert (siehe Übungen). Damit ist $\mathcal{F}(X, \mathbb{K})$ mit den punktweisen Operationen ein $\mathbb{K}$–Vektorraum.

Dieses Beispiel mag hier exotisch aussehen, tatsächlich spielen Funktionenräume in weiten Bereichen der Mathematik und der Physik eine zentrale Rolle.

(5) Analog zu Beispiel (4) kann man für einen beliebigen Körper $\mathbb{K}$, eine Menge X und einen $\mathbb{K}$–Vektorraum V die Menge $\mathcal{F}(X, V)$ aller Funktionen $f : X \to V$ zu

einem Vektorraum machen. Die Operationen werden wieder punktweise definiert, also $(f \oplus g)(x) := f(x) \oplus g(x)$ (wobei aber jetzt das erste $\oplus$ die Addition auf $\mathcal{F}(X, V)$ bezeichnet, das zweite aber die Addition in V) und $(r \odot f)(x) = r \odot (f(x))$ (wobei auch $\odot$ zwei Bedeutungen hat). Die Eigenschaften (V1)–(V8) für $\mathcal{F}(X, V)$ folgen hier analog wie in (3) aus den entsprechenden Eigenschaften von V.

2.3. Als nächstes wollen wir einige allgemeine Rechenregeln für Vektorräume beweisen, die in allen oben angeführten Beispielen völlig offensichtlich sind. Der wesentliche Punkt hier ist aber, dass man nicht mehr als die Eigenschaften (K1)–(K9) von Körpern und (V1)–(V8) von Vektorräumen braucht, um diese Regeln zu beweisen. Wenn man sich nämlich bewusst macht, was die Aussagen eigentlich bedeuten (etwa "Wenn man ein beliebiges Element von V mit dem additiv neutralen Element von $\mathbb{K}$ multipliziert, dann erhält man das additiv neutrale Element von V.") dann sind sie gar nicht mehr offensichtlich.

Proposition 2.3. *Sein $\mathbb{K}$ ein Körper und V ein Vektorraum über $\mathbb{K}$. Dann gilt:*
(1) $0 \odot v = 0_V$ für alle $v \in V$.
(2) $(-1) \odot v = \ominus v$ für alle $v \in V$.
(3) $r \odot 0_V = 0_V$ für alle $r \in \mathbb{K}$.
(4) Für gegebenes $a \in \mathbb{K} \setminus \{0\}$ und $w \in V$ hat die Gleichung $a \odot v = w$ eine eindeutige Lösung in V, nämlich $v = a^{-1} \odot w$.

Beweis. (1) Nach (V5) ist $(0 \odot v) \oplus (0 \odot v) = (0 + 0) \odot v = 0 \odot v$, und wir haben bereits in 2.2 bemerkt, dass dies $0 \odot v = 0_V$ impliziert. (Man addiert einfach zu beiden Seiten der Gleichung $\ominus(0 \odot v)$.)

(2) Nach (V8), (V5) und Teil (1) gilt

$$((-1) \odot v) \oplus v = ((-1) \odot v) \oplus (1 \odot v) = ((-1) + 1) \odot v = 0 \odot v = 0_V.$$

Wir haben wiederum in 2.2 bemerkt, dass das $(-1) \odot v = \ominus v$ impliziert. (Man addiert einfach zu beiden Seiten der Gleichung $\ominus v$.)

(3) Nach (V6) gilt $(r \odot 0_V) \oplus (r \odot 0_V) = r \odot (0_V \oplus 0_V) = r \odot 0_V$ und wie in (1) folgt die Behauptung.

(4) Nach (V7) und (V8) gilt $a \odot (a^{-1} \odot w) = (aa^{-1}) \odot w = 1 \odot w = w$, also ist $a^{-1} \odot w$ eine Lösung der Gleichung. Umgekehrt folgt aus $a \odot v = w$ natürlich $a^{-1} \odot (a \odot v) = a^{-1} \odot w$, und wie oben folgt $a^{-1} \odot (a \odot v) = v$. □

An dieser Stelle wird es Zeit für eine **Vereinfachung der Notation**: Wir werden ab sofort die Addition in *jedem* Vektorraum mit $+$ und die Skalarmultiplikation immer mit $\cdot$ oder durch einfaches Hintereinanderschreiben der Elemente bezeichnen. Außerdem werden wir den *Nullvektor* (also das neutrale Element für die Addition in V) immer einfach mit 0 (statt mit 0_V) und das additiv inverse Element zu $v \in V$ einfach mit $-v$ bezeichnen.

Die Definition eines Vektorraumes ist so gewählt, dass diese Vereinfachung ohne Gefahr von Widersprüchen möglich ist, und auch die Aussagen in Proposition 2.3 zeigen, dass die "neue" Notation Sinn macht. Sie birgt aber auch gewisse Gefahren in sich, weil sie vom Leser mehr Aufmerksamkeit bzw. Mitdenken verlangt. Es muss nämlich aus dem Zusammenhang geschlossen werden, was die verwendeten Symbole genau bedeuten.

Teil (1) der Proposition liest sich etwa in der neuen Notation als $0 \cdot v = 0$. Um das richtig zu interpretieren muss beobachten, dass weil $v \in V$ ist, der Punkt nur die Skalarmultiplikation bezeichnen kann. Damit muss aber die Null auf der linken Seite der Gleichung in $\mathbb{K}$ und die Null auf der rechten Seite der Gleichung in V liegen. Analog kann für Teil (2), nämlich $(-1)v = -v$ wegen $v \in V$ links mit -1 nur das additiv inverse Element zu 1 in $\mathbb{K}$ gemeint sein, während rechts nur das additiv inverse Element zu v in V stehen kann.

Diese Probleme ziehen sich durch fast alles, was wir bisher gemacht haben. Die Vektorraumeigenschaft (V7) schreibt sich etwa in der neuen Notation als $(rs)v = r(sv)$ für $r, s \in \mathbb{K}$ und $v \in V$, und es ist dem/der LeserIn überlassen, zu bedenken, dass hier einmal die Multiplikation in $\mathbb{K}$ und dreimal die Skalarmultiplikation vorkommt. Für die Funktionenräume aus den Beispielen (3) und (4) von oben bekommt die Definition der Addition die Form $(f + g)(x) := f(x) + g(x)$ (und zwar in beiden Fällen) und man muss bei der Interpretation aufpassen. Ich kann nur allen Studierenden dringend empfehlen, in der nächsten Zeit diesen Aspekt sorgfältig im Auge zu behalten und nicht "drüberzulesen". Mit der Zeit stellt sich eine gewisse Automatik ein und es ist weniger bewusste Aufmerksamkeit nötig. Es lohnt sich aber, immer wieder einmal zu hinterfragen, ob wirklich klar ist, "was da eigentlich steht".

2.4. Teilräume. Der Begriff des Teilraums ist analog zu den Begriffen "Untergruppe", "Teilring" und "Teilkörper", die in der EMA kurz angesprochen wurden (und vermutlich eher schwierig zu verdauen waren). Algebraisch sauberer wäre es, einen Teilraum eines Vektorraumes V als eine Teilmenge $W \subset V$ zu definieren, sodass sich die Vektorraumoperationen auf V zu Vektorraumoperationen auf W einschränken lassen. Für AnfängerInnen vermutlich leichter verständlich, ist die folgende operative Definition:

Definition 2.4. Sei V ein Vektorraum über einem Körper $\mathbb{K}$. Eine nichtleere Teilmenge $W \subset V$ heißt ein *Teilraum* oder ein *Unterraum* von V, wenn für alle $w_1, w_2 \in W$ auch $w_1 + w_2 \in W$ und für alle $w \in W$ und $r \in \mathbb{K}$ auch $rw \in W$ gilt.

Teil (1) der folgenden Proposition ist eine alternative Beschreibung der Teilraumeigenschaft, die oft handlich ist, Teil (2) beweist die angestrebte Einschränkbarkeit der Operationen.

Proposition 2.4. *Sei V ein Vektorraum über $\mathbb{K}$. Dann gilt:*

(1) Eine Teilmenge $W \subset V$ ist genau dann ein Teilraum, wenn $0 \in W$ und für alle $w_1, w_2 \in W$ und $r \in \mathbb{K}$ auch $rw_1 + w_2 \in W$ gilt.

(2) Ist $W \subset V$ ein Teilraum, dann kann man die Addition und die Skalarmultiplikation zu Abbildungen $+ : W \times W \to W$ und $\cdot : \mathbb{K} \times W \to W$ einschränken, die W zu einem $\mathbb{K}$–Vektorraum machen.

Beweis. (1) Ist $W \subset V$ ein Teilraum, dann ist $W \neq \emptyset$, also gibt es ein Element $w \in W$. Dann liegt aber auch $(-1) \cdot w$ in W, was nach Teil (2) von Proposition 2.3 mit $-w$ übereinstimmt. Damit gilt aber $w + (-w) = 0 \in W$. Sind nun $w_1, w_2 \in W$ und $r \in \mathbb{K}$ beliebig, dann gilt nach Definition $rw_1 \in W$ und daher auch $rw_1 + w_2 \in W$, also sind die Bedingungen in (1) erfüllt.

Sind umgekehrt die Bedingungen in (1) erfüllt, dann ist $0 \in W$, also $W \subset V$ nichtleer. Für $w_1, w_2 \in W$ ist dann $1 \cdot w_1 + w_2 = w_1 + w_2 \in W$. Andererseits ist für $w \in W$ und $r \in \mathbb{K}$ auch $r \cdot w + 0 = r \cdot w \in W$, also ist $W \subset V$ ein Teilraum.

(2) Die Definition des Teilraumes bedeutet ja gerade, dass man die Operationen zu $+ : W \times W \to W$ und $\cdot : \mathbb{K} \times W \to W$ einschränken kann. Die Eigenschaften (V1), (V2) und (V5)–(V8) sind dann für die eingeschränkten Operationen offensichtlich erfüllt. Aus dem Beweis von Teil (1) wissen wir, dass $0 \in W$ gilt und nach Teil (2) von Proposition 2.3 ist $-w = (-1) \cdot w$, also liegt für $w \in W$ auch $-w \in W$. Somit sind auch (V3) und (V4) für die Einschränkung der Addition erfüllt. $\square$

Korollar 2.4. *Sei $\{W_i : i \in I\}$ eine beliebige Menge von Teilräumen eines $\mathbb{K}$–Vektorraumes V. Dann ist auch der Durchschnitt $W := \bigcap_{i\in I} W_i \subset V$ ein Teilraum.*

Beweis. Wir benutzen die Charakterisierung aus Teil (1) der Proposition. Da jedes $W_i \subset V$ ein Teilraum ist, gilt $0 \in W_i$ für alle $i \in I$, also $0 \in W$. Sind $w_1, w_2 \in W$ und $r \in \mathbb{K}$, dann gilt nach Definition $w_1, w_2 \in W_i$ für alle $i \in I$. Da jedes W_i ein Teilraum ist, gilt auch $rw_1 + w_2 \in W_i$ für alle $i \in I$, also $rw_1 + w_2 \in W$. Nach Teil (1) der Proposition ist damit $W \subset V$ ein Teilraum. □

Beispiel 2.4. (1) Für jeden $\mathbb{K}$–Vektorraum V sind die Teilmengen $\{0\} \subset V$ und $V \subset V$ offensichtlich Teilräume. Im $\mathbb{K}$–Vektorraum $\mathbb{K}$ sind das die einzigen Teilräume. Ist nämlich $W \subset \mathbb{K}$ ein Teilraum, der ein Element $r \neq 0$ enthält, dann ist auch $1 = r^{-1}r \in W$ und damit $s = s1 \in W$ für alle $s \in \mathbb{K}$.

(2) Betrachte den $\mathbb{K}$–Vektorraum $\mathbb{K}^n$ und eine Zahl k mit $1 \leq k \leq n$ und definiere $W := \{(r_1, \dots, r_k, 0, \dots, 0) : r_1, \dots, r_k \in \mathbb{K}\}$, also die Teilmenge aller jener Punkte, für die alle Koordinaten nach der k–ten gleich Null sind. Dann erfüllt W offensichtlich die Bedingungen von Definition 2.4 und ist damit ein Teilraum von $\mathbb{K}^n$. Offensichtlich kann man W mit $\mathbb{K}^k$ identifizieren (formal werden wir uns das noch genauer ansehen).

Analog können wir für eine Menge X, eine Teilmenge $Y \subset X$ und einen Körper $\mathbb{K}$ den Vektorraum $\mathcal{F}(X, \mathbb{K})$ aller $\mathbb{K}$–wertigen Funktionen auf X aus Beispiel (4) von 2.2 und die Teilmenge $W := \{f : X \to \mathbb{K} : f(y) = 0 \quad \forall y \in Y\}$. Verschwindet eine Funktion in y, dann natürlich auch jedes Vielfache dieser Funktion und für zwei Funktionen, die in y verschwinden, verschwindet auch ihre Summe in y. Damit erfüllt W die Bedingungen von Definition 2.4 und ist somit ein Teilraum von $\mathcal{F}(X, \mathbb{K})$.

(3) Betrachten wir den $\mathbb{K}$–Vektorraum $\mathbb{K}^n$ und die Teilmenge

$$W := \{(r_1, \dots, r_n) : r_1 + \dots + r_n = 0\}.$$

Dann verifiziert man wieder sofort, dass die Eigenschaften aus Definition 2.4 erfüllt sind, also $W \subset \mathbb{K}^n$ ein Teilraum ist.

Seien allgemeiner $a_1, \dots, a_n \in \mathbb{K}$ beliebige Elemente und definiere

$$W := \{(r_1, \dots, r_n) : a_1 r_1 + \dots + a_n r_n = 0\}.$$

Dann ist offensichtlich $0 \in W$ und für Elemente $v = (v_1, \dots, v_n)$ und $w = (w_1, \dots, w_n)$ von W und $r \in \mathbb{K}$ gilt $rv + w = (rv_1 + w_1, \dots, rv_n + w_n)$ und

$$\begin{aligned} a_1(rv_1 + w_1) + \dots + a_n(rv_n + w_n) &= r(a_1 v_1) + a_1 w_1 + \dots + r(a_n v_n) + a_n w_n \\ &= r(a_1 v_1 + \dots + a_n v_n) + (a_1 w_1 + \dots + a_n w_n) = 0, \end{aligned}$$

also ist W nach Proposition 2.4 ein Teilraum. Ist $\mathbb{K} = \mathbb{R}$, $n = 3$ und sind die a_i nicht alle gleich Null, dann ist W geometrisch gesehen eine Ebene durch den Nullpunkt in $\mathbb{R}^3$, nämlich die Normalebene auf den Vektor (a_1, a_2, a_3).

(4) Betrachte wieder den $\mathbb{K}$–Vektorraum $\mathbb{K}^n$ und die Teilmenge

$$W := \{(r_1, \dots, r_n) : r_n = 1\}.$$

Dann ist W kein Teilraum. Die einfachste Begründung dafür ist, der Nullvektor nicht in W liegt (siehe Proposition 2.4), es gibt aber auch viele andere Argumente.

Betrachten wir $\mathbb{R}^2$ und die Teilmenge $W := \{(x, y) : y = x^2\}$, dann ist das ebenfalls kein Teilraum, weil etwa $(1, 1)$ in W liegt, $2 \cdot (1, 1) = (2, 2)$ aber nicht.

(5) Wir wollen alle Teilräume von $\mathbb{R}^2$ bestimmen. Ist $W \subset \mathbb{R}^2$ ein Teilraum und $W \neq \{0\}$, dann wählen wir einen Vektor $w \in W \setminus \{0\}$. Nach Definition muss dann auch $r \cdot w \in W$ für alle $r \in \mathbb{R}$ gelten, also enthält W die Gerade, die durch 0 und w geht. Nun

bildet die Teilmenge $\{rw : r \in \mathbb{R}\} \subset \mathbb{R}^2$ einen Teilraum, weil $rw + sw = (r + s)w$ und $s(rw) = (sr)w$ für alle $r, s \in \mathbb{R}$ gilt. Wenn W nicht mit diesem Teilraum übereinstimmt, dann finden wir einen Vektor $\tilde{w} \in W$, der kein Vielfaches von w ist. Schreiben wir $w = (w_1, w_2)$ und $\tilde{w} = (\tilde{w}_1, \tilde{w}_2)$. Dann muss entweder $w_1 \neq 0$ oder $\tilde{w}_1 \neq 0$ gelten, weil sonst $\tilde{w} = \frac{\tilde{w}_2}{w_2} w$ wäre. (Wegen $w, \tilde{w} \neq 0$ macht die Division keine Probleme.) Nehmen wir ohne Beschränkung der Allgemeinheit $w_1 \neq 0$ an, dann wissen wir, dass $\tilde{w} - \frac{\tilde{w}_1}{w_1} w = (0, w_2 - \frac{\tilde{w}_1}{w_1}\tilde{w}_2) \in W$ gilt. Die zweite Komponente dieses Vektors muss ungleich Null sein, weil sonst $\tilde{w} = \frac{\tilde{w}_1}{w_1} w$ wäre. Multipliziert man mit dem Inversen dieser Komponente, dann erhält man $(0, 1) \in W$. Damit ist aber dann auch $(w_1, 0) = w - w_2 \cdot (0, 1) \in W$ und wegen $w_1 \neq 0$ folgt $(1, 0) \in W$. Damit ist aber $(r_1, r_2) = r_1 \cdot (1, 0) + r_2 \cdot (0, 1) \in W$ für alle $r_1, r_2 \in \mathbb{R}$, also $W = \mathbb{R}^2$. Damit sehen wir, dass neben $\{0\}$ und $\mathbb{R}^2$ nur die Geraden durch Null Teilräume von $\mathbb{R}^2$ bilden.

Analog zeigt man (siehe Übungen), dass die Teilräume von $\mathbb{R}^3$ genau $\{0\}$, die Geraden durch Null, die Ebenen durch Null, und $\mathbb{R}^3$ selbst sind.

Lineare Abbildungen

Es ist ein wichtiges Prinzip der Mathematik, dass zu einer gewissen Struktur (d.h. zu Objekten) immer auch die Abbildungen zwischen diesen Objekten betrachtet werden, die mit dieser Struktur verträglich sind. Das hilft auch dabei bei Objekten die in vielen verschiedenen Rollen auftreten (zum Beispiel die reellen Zahlen $\mathbb{R}$) die Übersicht zu bewahren, welche der vielen vorhandenen Strukturen im Moment die relevante ist. (Auf $\mathbb{R}$ kann man natürlich Körperhomomorphismen, lineare Abbildungen, stetige, differenzierbare oder integrierbare Funktionen betrachten, und jede dieser Klassen von Funktionen spiegelt eine natürliche Struktur wieder, die die Menge $\mathbb{R}$ trägt.)

2.5. Grundlegende Definitionen. Im Fall der Vektorräume sind die strukturerhaltenden Abbildungen die linearen Abbildungen. Dabei ist wichtige, dass es im Allgemeinen keinen Sinn macht, Vektorräume über verschiedenen Körpern zu "vergleichen". Der Begriff der linearen Abbildung ist nur für Vektorräume über dem gleichen Körper definiert. Ab sofort werden wir daher fast immer mit Vektorräumen über einem fixen (aber beliebigen) Körper $\mathbb{K}$ arbeiten.

Definition 2.5. Seien V und W Vektorräume über $\mathbb{K}$.

(1) Eine *lineare Abbildung* von V nach W ist eine Funktion $f : V \to W$, sodass $f(v_1 + v_2) = f(v_1) + f(v_2)$ und $f(rv) = rf(v)$ für alle $v, v_1, v_2 \in V$ und alle $r \in \mathbb{K}$ gilt.

(2) Die Menge aller linearen Abbildungen von V nach W wird mit $L(V, W)$ oder (wenn der Körper $\mathbb{K}$ aus dem Zusammenhang nicht klar ist oder betont werden soll) mit $L_{\mathbb{K}}(V, W)$ bezeichnet.

(3) Ein *linearer Isomorphismus* von V nach W ist eine bijektive lineare Abbildung $f : V \to W$.

(4) Man sagt, V und W sind *isomorph* und schreibt $V \cong W$, wenn es einen linearen Isomorphismus von V nach W gibt.

Bemerkung 2.5. (1) Die Definition einer linearen Abbildung ist eigentlich sehr einsichtig. Man hat auf einem Vektorraum zwei Operationen und eine lineare Abbildung ist einfach eine Funktion, die mit beiden Operationen verträglich ist. Weil in der Definition die Vektorraumoperationen auf V und W mit den selben Symbolen bezeichnet werden, muss man wieder aufpassen, was die Dinge genau bedeuten.

(2) Isomorphe Vektorräume sind vom Standpunkt der linearen Algebra aus gesehen als "gleich" (oder vielleicht besser als "ununterscheidbar") zu betrachten. Ein linearer

Isomorphismus $f : V \to W$ kann benutzt werden, um Punkte von V mit entsprechenden Punkten von W zu identifizieren. Da f linear ist, sehen die Vektorraumoperationen in beiden Bildern gleich aus. Beim Übergang zu einem isomorphen Vektorraum gibt man sozusagen nur den Elementen des Raumes "andere Namen", lässt aber sonst alles unverändert.

Wir können nun leicht einige fundamentale Eigenschaften von linearen Abbildungen beweisen. Dazu erinnern wir uns an Abschnitt 4.3 der EMA: Eine Funktion $f : V \to W$ ist *injektiv*, wenn $f(v_1) = f(v_2)$ nur für $v_1 = v_2$ gilt und *surjektiv*, wenn es zu jedem $w \in W$ ein Element $v \in V$ gibt, für das $f(v) = w$ gilt. Schließlich heißt f *bijektiv*, wenn f injektiv und surjektiv ist. Das ist äquivalent dazu, dass es eine Funktion $g : W \to V$ gibt, sodass $g \circ f = \mathrm{id}_V$ und $f \circ g = \mathrm{id}_W$ gilt. Die Funktion g ist dann eindeutig bestimmt und wird als die inverse Funktion f^{-1} zu f bezeichnet.

Satz 2.5. *Seien U, V und W Vektorräume über $\mathbb{K}$.*

(1) Eine Funktion $f : V \to W$ ist genau dann eine lineare Abbildung, wenn

$$f(rv_1 + v_2) = rf(v_1) + f(v_2)$$

für alle $v_1, v_2 \in V$ und alle $r \in \mathbb{K}$ gilt.

(2) Ist $f : V \to W$ eine lineare Abbildung, dann gilt für $v_1, \dots, v_n \in V$ und $r_1, \dots, r_n \in \mathbb{K}$, dass $f(\sum_i r_i v_i) = \sum_i r_i f(v_i)$.

(3) Die Komposition linearer Abbildungen ist linear, d.h. sind $f : U \to V$ und $g : V \to W$ linear, dann ist $g \circ f : U \to W$ linear.

(4) Für einen linearen Isomorphismus $f : V \to W$ ist auch die inverse Funktion $f^{-1} : W \to V$ ein linearer Isomorphismus.

(5) Betrachtet man die Menge $\mathcal{F}(V, W)$ als $\mathbb{K}$–Vektorraum mit den Operationen aus Beispiel (5) von 2.2, dann bildet $L(V, W) \subset \mathcal{F}(V, W)$ einen Teilraum. Insbesondere ist also $L(V, W)$ ein Vektorraum unter den punktweisen Operationen.

Beweis. (1) Ist f linear, dann ist $f(rv_1 + v_2) = f(rv_1) + f(v_2) = rf(v_1) + f(v_2)$. Erfüllt umgekehrt f diese Bedingung, dann ist

$$f(v_1 + v_2) = f(1 \cdot v_1 + v_2) = 1 \cdot f(v_1) + f(v_2) = f(v_1) + f(v_2).$$

Damit ist aber $f(0) = f(0 + 0) = f(0) + f(0)$, also $f(0) = 0$. Benutzt man das, dann erhält man $f(rv) = f(rv + 0) = rf(v) + 0 = rf(v)$, also ist f linear.

(2) Nach Teil (1) gilt

$$f((r_1v_1 + \dots + r_{n-1}v_{n-1}) + r_nv_n) = f(r_1v_1 + \dots + r_{n-1}v_{n-1}) + r_nf(v_n)$$

und damit folgt das allgemeine Resultat sofort durch Induktion nach n.

(3) Nach Definition gilt $(g \circ f)(u) = g(f(u))$ für $u \in U$. Unter Benutzung von Teil (1) folgt $(g \circ f)(0) = g(0) = 0$ und

$$\begin{aligned}(g \circ f)(ru_1 + u_2) &= g(f(ru_1 + u_2)) = g(rf(u_1) + f(u_2)) = rg(f(u_1)) + g(f(u_2)) \\ &= r(g \circ f)(u_1) + (g \circ f)(u_2),\end{aligned}$$

also ist $g \circ f$ wieder linear.

(4) Sei f ein linearer Isomorphismus und $f^{-1} : W \to V$ die inverse Funktion zu f. Da $f(0) = 0$ gilt, folgt natürlich sofort $f^{-1}(0) = 0$. Andererseits ist für $w_1, w_2 \in W$ und $r \in \mathbb{K}$

$$f(rf^{-1}(w_1) + f^{-1}(w_2)) = rf(f^{-1}(w_1)) + f(f^{-1}(w_2)) = rw_1 + w_2.$$

Das bedeutet aber gerade, dass $f^{-1}(rw_1 + w_2) = rf^{-1}(w_1) + f^{-1}(w_2)$ gilt, also ist f^{-1} linear nach Teil (1).

(5) Das ist eine einfache Rechnung, bei der aber die Interpretation schwierig ist, weil nun schon drei verschiedene Operationen mit + bzw. mit $\cdot$ bezeichnet werden.

Um zu zeigen, dass $L(V,W) \subset \mathcal{F}(V,W)$ ein Teilraum ist, müssen wir nach Definition 2.4 zeigen, dass für $f, g \in L(V,W)$ und $r \in \mathbb{K}$ auch $rf : V \to W$ und $f+g : V \to W$ in $L(V,W)$ liegen, also linear sind. Nach Definition der Vektorraumstruktur auf $\mathcal{F}(V,W)$ in Beispiel (5) von 2.3 ist $(rf)(v) = r(f(v))$ und $(f+g)(v) = f(v)+g(v)$. Da $f(0) = 0$ und $g(0) = 0$ gilt, erhalten wir sofort $(rf)(0) = 0$ und $(f+g)(0) = 0$. Für $v_1, v_2 \in V$ und $s \in \mathbb{K}$ gilt weiters

$$\begin{aligned}(rf)(sv_1+v_2) &= r(f(sv_1+v_2)) = r(sf(v_1)+f(v_2)) = s((rf)(v_1)) + (rf)(v_2)\\ (f+g)(sv_1+v_2) &= f(sv_1+v_2)+g(sv_1+v_2) = sf(v_1)+f(v_2)+sg(v_1)+g(v_2)\\ &= s(f+g)(v_1)+(f+g)(v_2).\end{aligned}$$

□

Ein Ausdruck der Form $\sum_i r_i v_i$ wie in Teil (2) der Proposition heißt eine *Linearkombination* der Vektoren $v_1, \dots, v_n$.

Beispiel 2.5. (0) Für beliebige Vektorräume V und W ist die Nullabbildung $0 : V \to W$, $0(v) = 0$ für alle $v \in V$ klarerweise linear. (Auch hier tritt das Symbol “0” in zwei Bedeutungen auf, nämlich als Nullelement in den Vektorräumen V und $L(V,W)$.)

(1) Wir können relative einfach alle linearen Abbildungen $f : \mathbb{K}^n \to \mathbb{K}$ beschreiben: Für Elemente $a_1, \dots, a_n \in \mathbb{K}$ betrachte die Abbildung

$$f(r_1, \dots, r_n) := \textstyle\sum_i a_i r_i = a_1 r_1 + \dots + a_n r_n.$$

Zunächst können wir sofort nachrechnen, dass f linear ist. Offensichtlich gilt $f(0) = 0$ und für $v = (v_1, \dots, v_n)$ und $w = (w_1, \dots, w_n)$ in $\mathbb{K}^n$ und $r \in \mathbb{K}$ ist $rv + w$ gegeben durch $(rv_1 + w_1, \dots, rv_n + w_n)$ und damit

$$f(rv+w) = \textstyle\sum_i a_i(rv_i + w_i) = r\sum_i a_i v_i + \sum_i a_i w_i = rf(v) + f(w).$$

Umgekehrt ist aber auch jede lineare Abbildung von dieser Form: Für $i = 1, \dots, n$ definieren wir $e_i \in \mathbb{K}^n$ als den Vektor, dessen i–te Komponente gleich 1 ist, während alle anderen Komponenten Null sind.

Ist nun $f : \mathbb{K}^n \to \mathbb{K}$ eine beliebige lineare Abbildung dann setzen wir $a_i := f(e_i) \in \mathbb{K}$ für $i = 1, \dots, n$. Dann gilt aber $(r_1, \dots, r_n) = r_1 e_1 + \dots + r_n e_n$, also

$$f(r_1, \dots, r_n) = f(\textstyle\sum_i r_i e_i) = \sum_i r_i f(e_i) = \sum_i r_i a_i.$$

(2) Sei V ein beliebiger $\mathbb{K}$–Vektorraum und betrachte eine Funktion $f : V \to \mathbb{K}^n$ für beliebiges n. Dann ist für jedes $v \in V$ das Bild $f(v)$ ein n–Tupel von Elementen von $\mathbb{K}$ und wir bezeichnen mit $f_i(v) \in \mathbb{K}$ die i–te Komponente dieses Tupels. Dann definiert das Funktionen $f_i : V \to \mathbb{K}$ für $i = 1, \dots, n$, und $f(v) = (f_1(v), \dots, f_n(v))$. Nun ist für $v_1, v_2 \in V$ die i–te Komponente von $f(v_1 + v_2)$ natürlich gerade $f_i(v_1 + v_2)$, während die i–te Komponente von $f(v_1) + f(v_2)$ gleich $f_i(v_1) + f_i(v_2)$ ist. Daher gilt $f(v_1 + v_2) = f(v_1) + f(v_2)$ genau dann, wenn $f_i(v_1 + v_2) = f_i(v_1) + f_i(v_2)$ für alle $i = 1, \dots, n$ gilt. Analog sieht man sofort, dass für $v \in V$ und $r \in \mathbb{K}$ die Gleichung $f(rv) = rf(v)$ äquivalent zu $f_i(rv) = rf_i(v)$ für alle $i = 1, \dots, n$ ist. Somit sehen wir, dass eine Funktion $f : V \to \mathbb{K}^n$ genau dann linear ist, wenn alle Koeffizientenfunktionen $f_i : V \to \mathbb{K}$ linear sind.

Zusammen mit Beispiel (1) liefert das natürlich eine vollständige Beschreibung aller linearen Abbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ für beliebige $m, n \in \mathbb{N}$. Diese Beschreibung ist der Ausgangspunkt für den Matrizenkalkül, den wir in Kürze entwickeln werden.

(3) Für $k < n$ betrachte die Funktion $f : \mathbb{K}^k \to \mathbb{K}^n$, die gegeben ist durch

$$f(r_1, \dots, r_k) := (r_1, \dots, r_k, 0, \dots, 0).$$

(Man hängt also einfach an einen Vektor der Länge k noch $n - k$ Nullen an.) Dann ist f nach Definition linear und offensichtlich ist f injektiv. Natürlich hat f Werte in dem Teilraum W aus Beispiel (2) von 2.4, und als Abbildung $\mathbb{K}^k \to W$ ist f natürlich bijektiv, also ein linearer Isomorphismus.Das ist die präzise Version der Aussage aus 2.4, dass man W mit $\mathbb{K}^k$ identifizieren, also $\mathbb{K}^k$ als Teilraum von $\mathbb{K}^n$ betrachten kann.

(4) Betrachte die Menge $X := \{1, \dots, n\}$ und den Vektorraum $\mathcal{F}(X, \mathbb{K})$ aller Funktionen $f : X \to \mathbb{K}$. Definiere eine Funktion $\varphi : \mathcal{F}(X, \mathbb{K}) \to \mathbb{K}^n$ durch

$$\varphi(f) := (f(1), \dots, f(n)).$$

Für $f, g \in \mathcal{F}(X, \mathbb{K})$, $r \in \mathbb{K}$ und $i = 1, \dots, n$ gilt $(rf)(i) = r(f(i))$ und $(f + g)(i) = f(i) + g(i)$, also folgt sofort, dass $\varphi(rf) = r\varphi(f)$ und $\varphi(f + g) = \varphi(f) + \varphi(g)$ gilt. Somit ist φ linear. Da man aus $\varphi(f)$ die Werte von f in allen Punkten ablesen kann, ist φ injektiv. Ist $v = (v_1, \dots, v_n) \in \mathbb{K}^n$ beliebig, dann definiert $f(i) := v_i$ natürlich eine Funktion $f : X \to \mathbb{K}$, sodass $\varphi(f) = v$ gilt. Damit ist φ auch surjektiv, also ein linearer Isomorphismus.

2.6. Polynome. Polynome spielen eine wichtige Rolle in der linearen Algebra. Für's erste lernen wir sie als ein weiteres Beispiel für einen Vektorraum kennen.

Definition 2.6. Sei $\mathbb{K}$ ein beliebiger Körper, und $k \in \mathbb{N}$.

(1) Ein *Polynom vom Grad* $\leq k$ über $\mathbb{K}$ in der Variablen x ist ein Ausdruck der Form $p = \sum_{i=0}^{k} a_i x^i = a_0 + a_1 x + \dots + a_k x^k$ für Koeffizienten $a_i \in \mathbb{K}$.

(2) Die Menge aller Polynome vom Grad $\leq k$ über $\mathbb{K}$ in der Variablen x wird mit $\mathbb{K}_k[x]$ bezeichnet, die Menge aller solchen Polynome (für beliebiges k) mit $\mathbb{K}[x]$.

Bemerkung 2.6. (1) Für $n \geq k$ kann man ein Polynom $p \in \mathbb{K}_k[x]$ auch als Element von $\mathbb{K}_n[x]$ auffassen, indem man die Koeffizienten a_i für $i > k$ Null setzt. Das muss man tun, um die Menge $\mathbb{K}[x]$ aller Polynome bilden zu können.

(2) Wie haben Polynome hier eigentlich nur als formalen Ausdruck, also durch ihre Koeffizienten definiert. Insbesondere bedeutet das, dass zwei Polynome genau dann gleich sind, wenn ihre Koeffizienten gleich sind, wobei man bei Polynomen verschiedenen Grades auch noch (1) benutzt. Die x^i spielen spielen noch keine wirkliche Rolle, sie werden später für die Multiplikation von Polynomen nützlich sein. Dieser formale Zugang mag ungewohnt sein, wenn man aus der Schule reelle Polynome als spezielle Funktionen auf $\mathbb{R}$ kennt, er ist aber aus mehreren Gründen nützlich. Wie wir gleich sehen werden, definiert zwar jedes Polynom über $\mathbb{K}$ eine Funktion $\mathbb{K} \to \mathbb{K}$, aber für endliche Körper ist ein Polynom durch diese Funktion nicht eindeutig bestimmt. Außerdem wird es später wichtig sein, dass man in ein Polynom über $\mathbb{K}$ auch allgemeinere Objekte einsetzen kann, als Elemente von $\mathbb{K}$.

Proposition 2.6. *(1) Für $p, q \in \mathbb{K}_n[x]$ mit $p = \sum a_i x^i$ und $q = \sum b_i x^i$, sowie $r \in \mathbb{K}$ definiere $p + q := \sum (a_i + b_i) x^i$ und $rp := \sum r a_i x^i$. Das macht $\mathbb{K}_n[x]$ zu einem Vektorraum über $\mathbb{K}$, der isomorph zu $\mathbb{K}^{n+1}$ ist.*

(2) Für $p = \sum a_i x^i \in \mathbb{K}[x]$ definiere eine Funktion $\varphi_p : \mathbb{K} \to \mathbb{K}$ durch $\varphi_p(s) := \sum a_i s^i$ für $s \in \mathbb{K}$. Dann definiert $p \mapsto \varphi_p$ eine lineare Abbildung $\mathbb{K}[x] \to \mathcal{F}(\mathbb{K}, \mathbb{K})$. Diese Abbildung ist aber nicht notwendigerweise injektiv oder surjektiv.

Beweis. (1) Definiere $f : \mathbb{K}_n[x] \to \mathbb{K}^{n+1}$ durch $f(p) := (a_0, \dots, a_n)$, wobei $p = \sum a_i x^i$ und $g : \mathbb{K}^{n+1} \to \mathbb{K}_n[x]$ durch $g(r_0, \dots, r_n) := \sum r_i x^i$. Dann sind nach Konstruktion sowohl f als auch g mit der Addition und der Skalarmultiplikation auf den beiden Räumen verträglich, d.h. es gilt $f(p+q) = f(p)+f(q)$ und $g(v+w) = g(v)+g(w)$ und so weiter. Damit folgen die definierenden Eigenschaften eines Vektorraumes für $\mathbb{K}_n[x]$ sofort aus den analogen Eigenschaften für $\mathbb{K}^{n+1}$. Hat man diese Eigenschaften verifiziert, dann ist natürlich f ein linearer Isomorphismus zwischen den beiden Räumen.

(2) Für $p = \sum a_i x^i$ und $q = \sum b_i x^i$ folgt aus den Definitionen, dass $\varphi_{p+q}(s) = \sum (a_i+b_i)s^i = \sum (a_i s^i + b_i s^i)$ gilt. Da die Summen endlich ist, können wir sie problemlos umordnen und als $(\sum a_i s^i) + (\sum b_j s^j) = \varphi_p(s) + \varphi_q(s)$ schreiben. Daraus folgt aber $\varphi_{p+q} = \varphi_p + \varphi_q$ und analog zeigt man $\varphi_{rp} = r\varphi_p$.

Natürlich gibt es viele Funktionen $\mathbb{R} \to \mathbb{R}$, die nicht als Polynome geschrieben werden können, also ist $p \mapsto \varphi_p$ als Funktion $\mathbb{R}[x] \to \mathcal{F}(\mathbb{R}, \mathbb{R})$ nicht surjektiv. Betrachten wir andererseits $\mathbb{K} := \mathbb{Z}_2$, dann hat $\mathcal{F}(\mathbb{Z}_2, \mathbb{Z}_2)$ nur vier Elemente. (Jede Funktion $f : \mathbb{Z}_2 \to \mathbb{Z}_2$ ist durch Ihre Werte auf 0 und 1 festgelegt, für die es jeweils zwei Möglichkeiten gibt.) Andererseits sind aber nach Teil (1) schon die Polynome vom Grad ≤ 2 über $\mathbb{Z}_2$ isomorph zu $(\mathbb{Z}_2)^3$, also gibt es acht solche Polynome. Tatsächlich verifiziert man sofort, dass das Polynom $p = x + x^2 = 0 + 1x + 1x^2 \in \mathbb{Z}_2[x]$ die Nullfunktion auf $\mathbb{Z}_2$ definiert, also $\varphi_p = 0$ erfüllt. Somit ist $p \mapsto \varphi_p$ in diesem Fall nicht injektiv. □

2.7. Kern und Bild. Um den Zusammenhang zwischen linearen Abbildungen und Teilräumen zu studieren, erinnern wir uns kurz an die Begriff von Bild und Urbild unter Funktionen aus der EMA: Sei $f : X \to Y$ eine Funktion zwischen zwei Mengen. Dann kann man für jede Teilmenge $A \subset X$ das *Bild* $f(A) := \{y \in Y : \exists a \in A : f(a) = y\}$ und für jede Teilmenge $B \subset Y$ das Urbild $f^{-1}(B) := \{x \in X : f(x) \in B\}$ bilden. (Das Symbol f^{-1} für das Urbild ist von der inversen Funktion zu unterscheiden, die ja im Allgemeinen nicht existiert.)

Definition 2.7. Seien V und W Vektorräume über $\mathbb{K}$ und sein $f : V \to W$ eine lineare Abbildung.

(1) Der *Kern von f* ist die Teilmenge $\mathrm{Ker}(f) := f^{-1}(\{0\}) = \{v \in V : f(v) = 0\} \subset V$.

(2) Das *Bild von f* ist $\mathrm{Im}(f) := f(V) = \{w \in W : \exists v \in V : f(v) = w\} \subset W$.

Proposition 2.7. *Sei $f : V \to W$ eine lineare Abbildung zwischen zwei $\mathbb{K}$–Vektorräumen. Dann gilt:*

(1) Ist $\tilde{V} \subset V$ ein Teilraum, dann ist auch $f(\tilde{V}) \subset W$ ein Teilraum. Insbesondere ist $\mathrm{Im}(f) \subset W$ ein Teilraum.

(2) Ist $\tilde{W} \subset W$ ein Teilraum, dann ist auch $f^{-1}(\tilde{W}) \subset V$ ein Teilraum. Insbesondere ist $\mathrm{Ker}(f) \subset V$ ein Teilraum.

(3) f ist genau dann surjektiv, wenn $\mathrm{Im}(f) = W$ gilt.

(4) f ist genau dann injektiv, wenn $\mathrm{Ker}(f) = \{0\}$ gilt.

Beweis. (1) Da $\tilde{V} \subset V$ ein Teilraum ist, gilt $0 \in \tilde{V}$, also $0 = f(0) \in f(\tilde{V})$. Sind $w_1, w_2 \in f(\tilde{V})$ und ist $r \in \mathbb{K}$, dann finden wir nach Definition $v_1, v_2 \in \tilde{V}$ mit $f(v_i) = w_i$ für $i = 1, 2$. Da $\tilde{V} \subset V$ ein Teilraum ist, gilt $rv_1 + v_2 \in \tilde{V}$, also $f(rv_1 + v_2) = rf(v_1) + f(v_2) = rw_1 + w_2 \in f(\tilde{V})$. Damit ist $f(\tilde{V})$ ein Teilraum. Die Aussage über $\mathrm{Im}(f)$ folgt natürlich, weil $V \subset V$ ein Teilraum ist.

(2) Natürlich gilt $0 \in f^{-1}(\tilde{W})$, und für $v_1, v_2 \in f^{-1}(\tilde{W})$ und $r \in \mathbb{K}$ ist $f(v_1), f(v_2) \in \tilde{W}$. Da $\tilde{W} \subset W$ ein Teilraum ist, gilt $rf(v_1) + f(v_2) = f(rv_1 + v_2) \in \tilde{W}$, also $rv_1 + v_2 \in f^{-1}(\tilde{W})$. Die Aussage über $\mathrm{Ker}(f)$ folgt, weil $\{0\} \subset W$ ein Teilraum ist.

(3) Das ist die Definition von Surjektivität.

(4) Ist f injektiv und $v \in \operatorname{Ker}(f)$, dann gilt $f(v) = 0 = f(0)$, also $v = 0$ wegen der Injektivität von f. Umgekehrt folgt aus $f(v_1) = f(v_2)$ für $v_1, v_2 \in V$ natürlich $0 = f(v_1) - f(v_2) = f(v_1 - v_2)$, also $v_1 - v_2 \in \operatorname{Ker}(f)$. Ist $\operatorname{Ker}(f) = \{0\}$ dann impliziert das $v_1 = v_2$. $\square$

KAPITEL 3

Matrizen und lineare Gleichungssysteme

Nachdem wir die fundamentalsten Begriffe der linearen Algebra entwickelt haben, verlassen wir vorübergehen die abstrakten Konzepte und wenden uns ihrer konkreten Realisierung zu. Wir werden zeigen, wie man lineare Abbildungen zwischen den Vektorräumen $\mathbb{K}^n$ (für verschiedene n) explizit beschrieben und ihre Kerne und Bilder bestimmen kann, indem man lineare Gleichungssysteme löst. Wir werden später sehen, dass sich diese Methoden auf allgemeinere Vektorräume übertragen lassen. Wir fixieren weiterhin einen beliebigen Körper $\mathbb{K}$.

Matrizen

Als ersten Schritt beschreiben wir lineare Abbildungen zwischen Räumen der Form $\mathbb{K}^n$ durch Matrizen. Die Vektorraumoperationen auf Räumen linearer Abbildungen und die Komposition von linearen Abbildungen übersetzen sich in einfache Rechenoperationen mit Matrizen. Es zeigt sich aber, dass es oft einfacher ist, die Eigenschaften dieser Operationen aus den allgemeinen Konzepten abzuleiten als sie nachzurechnen.

3.1. Grundlegende Definitionen. Die Definition einer Matrix ist (zumindest in Spezialfällen) schon aus der Schule und aus der EMA bekannt:

Definition 3.1. Sei $\mathbb{K}$ ein Körper.

(1) Eine $m \times n$*–Matrix* über $\mathbb{K}$ ist ein rechteckiges Schema von Elementen von $\mathbb{K}$ mit m Zeilen und n Spalten. Wir schrieben so eine Matrix als $A = (a_{ij})$ mit $i = 1, \dots, m$ und $j = 1, \dots, n$ oder als $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$, wobei $a_{ij} \in \mathbb{K}$ für alle i, j.

(2) Für $i = 1, \dots m$ ist der i–te *Zeilenvektor* von A die $1 \times n$–Matrix $(a_{i1}, \dots, a_{in})$ und für $j = 1, \dots, n$ ist der j–te *Spaltenvektor* von A die $m \times 1$–Matrix $\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$.

(3) Die Menge aller $m \times n$–Matrizen über $\mathbb{K}$ wird mit $M_{m,n}(\mathbb{K})$ bezeichnet. Für $m = n$ schreibt man $M_n(\mathbb{K})$ statt $M_{n,n}(\mathbb{K})$.

(4) Die $n \times n$–Einheitsmatrix $\mathbb{I}_n \in M_n(\mathbb{K})$ ist definiert als $\mathbb{I}_n = (\delta_{ij})$ wobei $\delta_{ij} = 1$ für $i = j$ und $\delta_{ij} = 0$ für $i \neq j$.

Bemerkung 3.1. (1) Wir können die Zeilenvektoren einer $m \times n$–Matrix über $\mathbb{K}$ natürlich als Elemente von $\mathbb{K}^n$ betrachten. Im Rahmen des Matrizenkalküls ist es aber wesentlich günstiger, die Elemente des Vektorraumes $\mathbb{K}^n$ als Spaltenvektoren zu betrachten. (Natürlich macht es logisch gesehen keinen Unterschied, ob man die Elemente eines n–Tupels mit Beistrichen getrennt nebeneinander oder ohne Trennung untereinander schreibt.) In diesem Bild kann man sich die Formeln, die wir im weiteren entwickeln werden, viel leichter merken. Dementsprechend werden die Spaltenvektoren einer Matrix für uns (zumindest anfänglich) wichtiger sein als die Zeilenvektoren. Wir werden später aber auch eine gute Interpretation für die Zeilenvektoren finden.

(2) Nach Definition ist eine $m \times n$–Matrix einfach eine geordnete Ansammlung von mn Elementen eines Körpers. Wir können sie aber auch als Familie der Spaltenvektoren, d.h. als (geordnete) Kollektion von n Elementen von $\mathbb{K}^m$, oder als Familie der Zeilenvektoren betrachten. Im weiteren wird es oft nützlich sein, zwischen diesen verschiedenen Bildern hin– und herzuschalten

Wie wir bereits bemerkt haben, ist eine Matrix $A \in M_{m,n}(\mathbb{K})$ nur eine Ansammlung von $m \cdot n$ vielen Elementen von $\mathbb{K}$ (diesmal nur rechteckig statt nebeneinander oder untereinander angeordnet). Damit können wir die Mengen $M_{m,n}(\mathbb{K})$ der $m \times n$–Matrizen über $\mathbb{K}$ auf offensichtliche Weise zu einem $\mathbb{K}$–Vektorraum machen: Man addiert Matrizen einfach komponentenweise und multipliziert sie auch komponentenweise mit Skalaren. Für Matrizen $A = (a_{ij})$ und $B = (b_{ij})$ in $M_{m,n}(\mathbb{K})$ haben wir also $A + B = (c_{ij})$, wobei $c_{ij} = a_{ij} + b_{ij}$ und für $r \in \mathbb{K}$ ist $rA = (d_{ij})$ mit $d_{ij} = ra_{ij}$ für alle $i = 1, \dots, m$ und $j = 1, \dots, n$. Wir werden dafür auch kurz $A+B = (a_{ij}+b_{ij})$ und $rA = (ra_{ij})$ schreiben. Beachte, dass die Addition nur für Matrizen gleicher Größe definiert ist.

3.2. Zusammenhang mit linearen Abbildungen. Wir haben die wesentliche Grundlage für den Matrizenkalkül schon in Beispiel 2.5 kennen gelernt. Betrachten wir für beliebige $n, m \in \mathbb{N}$ eine lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$. Wie in Beispiel (3) von 2.5 betrachten wir die Koeffizientenfunktionen $f_i : \mathbb{K}^n \to \mathbb{K}$ für $i = 1, \dots, m$, d.h. $f(v) = \begin{pmatrix} f_1(v) \\ \vdots \\ f_m(v) \end{pmatrix}$ für $v \in \mathbb{K}^n$. Aus diesem Beispiel wissen wir, dass die Abbildung f genau dann linear ist, wenn jede der Abbildungen $f_i : \mathbb{K}^n \to \mathbb{K}$ linear ist. Aus Beispiel (2) von 2.5 wissen wir dann aber, dass es Elemente $a_{i1}, \dots, a_{in} \in \mathbb{K}$ gibt, sodass f_i den Vektor $(v_1, \dots, v_n) \in \mathbb{K}^n$ auf $\sum_j a_{ij}v_j = a_{i1}v_1 + \dots + a_{in}v_n$ abbildet.

Fassen wir nun die Koeffizienten a_{ij} zu einer $m \times n$–Matrix $A = (a_{ij})$ zusammen, dann können wir diese Abbildungsvorschrift als $f(v) = A \cdot v$ schreiben, wobei

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} a_{11}v_1 + \dots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + \dots + a_{mn}v_n \end{pmatrix}.$$

Man ordnet also der Matrix $A \in M_{m,n}(\mathbb{K})$ und dem Vektor $v \in \mathbb{K}^n$ ein Element $A \cdot v \in \mathbb{K}^m$ zu. Man erhält die i–te Komponente von $A \cdot v$ indem man die Elemente der i–ten Zeile von A mit den Komponenten von v multipliziert und dann aufsummiert. Wir werden im weiteren statt $A \cdot v$ einfach Av schreiben.

Wir wissen also aus Beispiel 2.5 bereits, dass $f(v) := Av$ eine lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ definiert. Wir können aus Beispiel 2.5 auch direkt ablesen, wie die Matrix A mit der linearen Abbildung f zusammenhängt. Aus diesem Beispiel kennen wir die Vektoren $e_j \in \mathbb{K}^n$ für $j = 1, \dots, n$, wobei e_j als j–te Komponente 1 und alle anderen Komponenten gleich 0 hat. Der Vektor e_j heißt der *j–te Einheitsvektor* in $\mathbb{K}^n$. In Beispiel 2.5 haben wir gesehen, dass für $f_i(v) = \sum_j a_{ij}v_j$ der Koeffizient a_{ij} das Bild von e_j unter f_i ist, d.h. $a_{ij} = f_i(e_j)$. Damit gilt aber für den j–ten Spaltenvektor der Matrix A

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} f_1(e_j) \\ \vdots \\ f_m(e_j) \end{pmatrix} = f(e_j).$$

Damit sehen wir, dass *die n Spaltenvektoren der Matrix A zu einer linearen Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ genau die Bilder der n Einheitsvektoren $e_1, \dots, e_n$ von $\mathbb{K}^n$ unter f*

sind. Betrachtet man die Matrix A als ein n–Tupel von Spaltenvektoren, dann gilt also $A = (f(e_1), \dots, f(e_n))$. Das beweist den Anfang von

Proposition 3.2. *Jede lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ kann eindeutig als $f(v) = Av$ für eine Matrix $A \in M_{m,n}(\mathbb{K})$ geschrieben werden. Dabei sind die Spaltenvektoren von A genau die Bilder der Einheitsvektoren $e_j \in \mathbb{K}^n$ unter f.*

Stattet man die Menge $L(\mathbb{K}^n, \mathbb{K}^m)$ mit der Vektorraumstruktur aus Teil (5) von Satz 2.5 und $M_{m,n}(\mathbb{K})$ mit der Vektorraumstruktur aus 3.1 aus, dann definiert diese Konstruktion einen linearen Isomorphismus $L(\mathbb{K}^n, \mathbb{K}^m) \cong M_{m,n}(\mathbb{K})$.

Beweis. Da wir bereits festgestellt haben, dass die Abbildung $f \mapsto (f_i(e_j)) \in M_{m,n}(\mathbb{K})$ bijektiv ist (die inverse Abbildung bildet A auf die Funktion $v \mapsto Av$ ab), bleibt nach Teil (4) von Satz 2.5 nur noch zu zeigen, dass diese Abbildung linear ist. Sei also $g \in L(\mathbb{K}^n, \mathbb{K}^m)$ eine weitere Funktion. Dann ist nach Definition $f + g \in L(\mathbb{K}^n, \mathbb{K}^m)$ gegeben durch $(f + g)(v) = f(v) + g(v)$ und weil die Addition in $\mathbb{K}^m$ komponentenweise definiert ist, ist die i–te Komponente von $f + g$ gegeben durch $(f + g)_i(v) = f_i(v) + g_i(v)$. Für $v = e_j$ zeigt das, dass die Matrix zu $f + g$ gerade die komponentenweise Summe der Matrizen zu f und zu g ist. Genauso verifiziert man sofort, dass $(rf)_i(e_j) = r(f_i(e_j))$ für $r \in \mathbb{K}$ gilt, und vervollständigt den Beweis der Linearität der Abbildung $f \mapsto (f_i(e_j))$. □

Beispiel 3.2. (1) Betrachten wir die Identitätsabbildung $\mathrm{id} : \mathbb{K}^n \to \mathbb{K}^n$, dann ist natürlich $\mathrm{id}(e_j) = e_j$ für alle $j = 1, \dots, n$. Daher entspricht die Identitätsabbildung auf $\mathbb{K}^n$ der $n \times n$–Einheitsmatrix $\mathbb{I}_n$ aus Teil (4) von Definition 3.1.

(2) Betrachten wir die Abbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$, die gegeben ist durch $f\binom{v_1}{v_2} := \binom{2v_1 - v_2}{-3v_1 + 4v_2}$. Dann gilt natürliche $f(e_1) = \binom{2}{-3}$ und $f(e_2) = \binom{-1}{4}$, also ist die Matrix zu f gegeben durch $A = \begin{pmatrix} 2 & -1 \\ -3 & 4 \end{pmatrix}$.

3.3. Die Matrizenmultiplikation. Nachdem wir nun lineare Abbildungen zwischen $\mathbb{K}$–Vektorräumen der Form $\mathbb{K}^n$ durch Matrizen beschreiben können und aus Satz 2.5 wissen, dass die Komposition von zwei linearen Abbildungen wieder linear ist, liegt es nahe zu überlegen, wie die Komposition im Bild der Matrizen aussieht. Natürlich können wir für zwei lineare Abbildungen f und g die Komposition $g \circ f$ nur dann bilden, wenn die Funktion g auf jenem Raum definiert ist, in dem f Werte hat. Beginnen wir also mit einer Funktion $f : \mathbb{K}^n \to \mathbb{K}^m$, dann muss g eine Abbildung $\mathbb{K}^m \to \mathbb{K}^p$ für ein $p \in \mathbb{N}$ sein, und dann ist $g \circ f$ eine Abbildung $\mathbb{K}^n \to \mathbb{K}^p$. Dann entspricht f eine $m \times n$–Matrix, g eine $p \times m$–Matrix, und der Komposition $g \circ f$ eine $p \times n$–Matrix.

Bezeichnen wir die Matrizen zu f mit $B = (b_{ij})$ und zu g mit $A = (a_{k\ell})$. Wollen wir die Matrix zu $g \circ f$ bestimmen, dann müssen wir die Bilder der Einheitsvektoren $e_1, \dots, e_n \in \mathbb{K}^n$ unter der Abbildung $g \circ f$ bestimmen. Nun ist aber nach Definition $(g \circ f)(e_j) = g(f(e_j))$, und $f(e_j)$ ist ja gerade der j–te Spaltenvektor von B. Das Bild dieses Vektors unter g erhält man aber gerade, indem man die Matrix A auf diesen Spaltenvektor anwendet. Das motiviert

Definition 3.3. Seien $A = (a_{ij}) \in M_{p,m}(\mathbb{K})$ und $B = (b_{ij}) \in M_{m,n}(\mathbb{K})$ Matrizen über einem Körper $\mathbb{K}$. Dann definiert man das Matrixprodukt $A \cdot B = AB = (c_{ij}) \in$

$M_{p,n}(\mathbb{K})$ durch $c_{ij} := \sum_{k=1}^{m} a_{ik}b_{kj}$ bzw. durch

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{p1} & \dots & a_{pm} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix}$$
$$= \begin{pmatrix} a_{11}b_{11} + \dots + a_{1m}b_{m1} & \dots & a_{11}b_{1n} + \dots + a_{1m}b_{mn} \\ \vdots & & \vdots \\ a_{p1}b_{11} + \dots + a_{pm}b_{m1} & \dots & a_{p1}b_{1n} + \dots + a_{pm}b_{mn} \end{pmatrix}$$

Das Element c_{ij} erhält man also, indem man die Elemente der i–ten Zeile von A mit den jeweiligen Elementen der j–ten Spalte von B multipliziert und dann aufsummiert.

Bemerkung 3.3. (1) **Achtung**: Das Produkt AB ist nur dann definiert, wenn die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B ist. (Ansonsten macht auch die obige Regel keinen Sinn.) Versuche, Matrizen zu multiplizieren, deren Größen nicht (in diesem Sinn) zusammen passen, zeugen von schlimmem Unverständnis.

(2) Man kann einen Spaltenvektor $v \in \mathbb{K}^n$ natürlich als eine $n \times 1$–Matrix auffassen. Damit ist für eine $m \times n$–Matrix A das Matrizenprodukt Av definiert und stimmt nach Definition mit der Vorschrift aus 3.2 überein.

Wir können nun relativ einfach die wesentlichsten Eigenschaften der Matrizenmultiplikation abklären.

Satz 3.3. *(1) Die Multiplikation von Matrizen entspricht der Komposition von linearen Abbildungen.*

(2) Die Matrizenmultiplikation ist assoziativ. Genauer: Kann man für drei Matrizen A, B, C eines der Produkte $(AB)C$ und $A(BC)$ bilden, dann auch das andere und es gilt $(AB)C = A(BC)$.

(3) Die Matrizenmultiplikation ist auf beiden Seiten distributiv bezüglich der Addition, d.h. für $A, A_1, A_2 \in M_{p,m}(\mathbb{K})$ und $B, B_1, B_2 \in M_{m,n}(\mathbb{K})$ gilt

$$(A_1 + A_2)B = A_1B + A_2B \qquad A(B_1 + B_2) = AB_1 + AB_2.$$

(4) Für $A \in M_{m,n}(\mathbb{K})$ und die Einheitsmatrizen $\mathbb{I}_\ell \in M_\ell(\mathbb{K})$ gilt $\mathbb{I}_m A = A\mathbb{I}_n = A$.

Beweis. (1) In der Definition der Matrizenmultiplikation haben wir bemerkt, dass man die Spalten von AB erhält, indem man die Matrix A auf die Spalten von B anwendet. Das bedeutet aber gerade, dass $(AB)e_j = A(Be_j)$ für alle j gilt, und die Behauptung folgt.

(2) Diese und die weiteren Eigenschaften kann man direkt nachrechnen, es ist aber einfacher, sie aus den Eigenschaften der Komposition linearer Abbildungen herzuleiten.

Sei $A \in M_{m,n}(\mathbb{K})$, $B \in M_{k,\ell}(\mathbb{K})$ und $C \in M_{r,s}(\mathbb{K})$. Dann existiert das Produkt AB genau dann, wenn $n = k$ ist, und dann ist $AB \in M_{m,\ell}(\mathbb{K})$. Damit existiert das Produkt $(AB)C$ genau dann, wenn $n = k$ und $\ell = r$. Analog existiert BC genau dann wenn $\ell = r$ gilt, und $A(BC)$ existiert genau dann, wenn zusätzlich $n = k$ gilt. Somit ist die Existenz beider Produkte äquivalent. Wenn die beiden Produkte existieren, dann repräsentieren sie aber die linearen Abbildungen $(f \circ g) \circ h$ bzw. $f \circ (g \circ h)$, wobei $f(u) = Au$, $g(v) = Bv$ und $h(w) = Cw$. Nun ist aber

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

also sind diese beiden Kompositionen gleich.

(3) Analog zu (2) genügt es hier $(f_1+f_2)\circ g = (f_1\circ g)+(f_2\circ g)$ und $f\circ(g_1+g_2) = (f\circ g_1)+(f\circ g_2)$ zu beweisen. Nun ist aber

$$\begin{aligned}((f_1+f_2)\circ g)(v) =&(f_1+f_2)(g(v)) = f_1(g(v))+f_2(g(v))\\ &=(f_1\circ g)(v)+(f_2\circ g)(v) = ((f_1\circ g)+(f_2\circ g))(v),\end{aligned}$$

also folgt die erste Gleichung. Die zweite Gleichung verifiziert man analog.

(4) Das folgt direkt aus $f\circ \mathrm{id} = \mathrm{id}\circ f = f$. □

Die Matrizenmultiplikation ist *nicht* kommutativ. Die Frage der Kommutativität macht nur für quadratische Matrizen, also $A, B \in M_n(\mathbb{K})$ Sinn, für die man beide Produkte AB und BA bilden kann. Als Beispiel für die Nichtkommutativität betrachten wir $\left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right) = \left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$ und $\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)\left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right) = \left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$, was über jedem Körper $\mathbb{K}$ Sinn macht. Es ist sehr empfehlenswert, dieses Beispiel im Bild der linearen Abbildungen durchzudenken.

Beispiel 3.3. Betrachten wir die Matrix $A = \left(\begin{smallmatrix}2&-1\\-3&4\end{smallmatrix}\right)$ aus Beispiel (2) von 3.2. Dann können wir $A^2 = AA$ berechnen, nämlich

$$\begin{pmatrix}2&-1\\-3&4\end{pmatrix}\begin{pmatrix}2&-1\\-3&4\end{pmatrix} = \begin{pmatrix}7&-6\\-18&19\end{pmatrix},$$

und man verifiziert leicht, dass dies tatsächlich die Matrix zu $f^2 = f\circ f : \mathbb{R}^2 \to \mathbb{R}^2$ für die Abbildung f Beispiel (2) von 3.2 repräsentiert.

3.4. Invertierbare Matrizen. Zum Abschluss können wir noch lineare Isomorphismen im Bild der Matrizen beschreiben. Im nächsten Kapitel werden wir sehen, dass es einen linearen Isomorphismus $\mathbb{K}^n \to \mathbb{K}^m$ nur für $m = n$ geben kann. Da wir schon wissen, dass die Identitätsabbildung auf $\mathbb{K}^n$ durch die Einheitsmatrix $\mathbb{I}_n$ repräsentiert wird, ist die folgende Definition nahe liegend.

Definition 3.4. Eine quadratische Matrix $A \in M_n(\mathbb{K})$ heißt *invertierbar*, wenn es eine Matrix $B \in M_n(\mathbb{K})$ gibt, sodass $AB = BA = \mathbb{I}_n$ gilt, wobei $\mathbb{I}_n \in M_n(\mathbb{K})$ die $n\times n$–Einheitsmatrix bezeichnet.

Satz 3.4. *(1) Eine $n\times n$–Matrix $A \in M_n(\mathbb{K})$ ist genau dann invertierbar, wenn die Abbildung $f(v) := Av$ ein linearer Isomorphismus $\mathbb{K}^n \to \mathbb{K}^n$ ist*

(2) Ist A invertierbar, dann ist die Matrix B in Definition 3.4 eindeutig bestimmt. Sie heißt die inverse Matrix zu A und wird mit A^{-1} bezeichnet.

(3) Sind $A, B \in M_n(\mathbb{K})$ beide invertierbar, dann ist auch das Produkt AB invertierbar und $(AB)^{-1} = B^{-1}A^{-1}$.

Beweis. (1) Ist A invertierbar, dann bedeutet $AB = BA = \mathbb{I}_n$ genau dass die Funktion $v \mapsto Bv$ invers zu f ist, also ist f ein linearer Isomorphismus.

Ist umgekehrt $f : \mathbb{K}^n \to \mathbb{K}^n$ ein linearer Isomorphismus, dann gibt es eine lineare Abbildung $g : \mathbb{K}^n \to \mathbb{K}^n$, sodass $g\circ f = f\circ g = \mathrm{id}_{\mathbb{K}^n}$. Sind A und B die Matrizen zu f und g, dann bedeutet das aber genau $AB = BA = \mathbb{I}_n$.

(2) Gilt $AC = \mathbb{I}_n$, dann erhält man durch multiplizieren mit B von links sofort $B(AC) = B$ und $B(AC) = (BA)C = \mathbb{I}_n C = C$. Analog folgt aus $CA = \mathbb{I}_n$, dass $C = B$ gelten muss.

(3) Wir nehmen an, das A und B invertierbar sind und rechnen

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A\mathbb{I}_nA^{-1} = AA^{-1} = \mathbb{I}_n$$

Analog folgt $(B^{-1}A^{-1})(AB) = \mathbb{I}_n$ und damit die Behauptung nach (2). □

Bemerkung 3.4. (1) Wir werden in Kürze klären, wie man tatsächlich feststellen kann, ob eine Matrix invertierbar ist und wie man die inverse Matrix ausrechnen kann.

(2) Betrachten wir die Menge $GL(n,\mathbb{K})$ aller invertierbaren $n \times n$–Matrizen über $\mathbb{K}$. Dann folgt aus Teil (3) von Satz 3.4, dass für $A, B \in GL(n,\mathbb{K})$ auch das Produkt $AB \in GL(n,\mathbb{K})$ liegt. Also können wir die Matrizenmultiplikation als Abbildung $GL(n,\mathbb{K}) \times GL(n,\mathbb{K}) \to GL(n,\mathbb{K})$ betrachten, sie definiert also eine assoziative Verknüpfung auf $GL(n,\mathbb{K})$. Außerdem ist für eine invertierbare Matrix A auch die inverse Matrix A^{-1} invertierbar (und $(A^{-1})^{-1} = A$). Schließlich ist die Einheitsmatrix $\mathbb{I}_n$ auch invertierbar (mit $\mathbb{I}_n^{-1} = \mathbb{I}_n$). Damit folgt aber, dass $GL(n,\mathbb{K})$ eine Gruppe unter der Matrizenmultiplikation mit neutralem Element $\mathbb{I}_n$ bildet, siehe Kapitel 5.2 von [EMA]. Diese Gruppe ist (für $n > 1$) nicht kommutativ. Sie stellt eines der wichtigsten Beispiele für eine nicht kommutative Gruppe dar.

Beispiel 3.4. (1) Für eine reelle Zahl $a \in \mathbb{R}$ betrachten wir die lineare Abbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$, die gegeben ist durch $f\binom{x}{y} = \binom{x+ay}{y}$. Dann sieht man sofort, dass f invertierbar ist und die Inverse Funktion $g : \mathbb{R}^2 \to \mathbb{R}^2$ gegeben ist durch $g\binom{u}{v} = \binom{u-av}{v}$. In Termen der zugehörigen Matrizen bedeutet das, dass die Matrix $A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ invertierbar ist und die inverse Matrix durch $A^{-1} = \begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix}$ gegeben ist, was man sofort durch nachrechnen überprüft.

(2) Betrachte eine allgemeine 2×2–Matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M_2(\mathbb{R})$. Dann rechnet man sofort nach, dass

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (a_{11}a_{22} - a_{12}a_{21})\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ist $a_{11}a_{22} - a_{12}a_{21} \neq 0$ (vergleiche mit 1.2!), dann folgt sofort, dass A invertierbar ist und $A^{-1} = (a_{11}a_{22} - a_{12}a_{21})^{-1}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$ gilt. Das liefert natürlich auch Beispiel (1) als Spezialfall.

Lineare Gleichungssysteme

Wir wenden uns nun der Lösung von linearen Gleichungssystemen zu. Das wird uns auch eine explizite Methode liefern um für Abbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ Kern und Bild zu berechnen. Weiters werden wir entschieden können, ob eine gegebene quadratische Matrix invertierbar ist und in diesem Fall die Inverse explizit berechnen können.

3.5. Grundlegende Begriffe und Eigenschaften. Wir haben Spezialfälle von linearen Gleichungssystemen schon in Kapitel 1 oberflächlich kennen gelernt. Jetzt werden wir die nötigen Begriff formal sauber definieren.

Definition 3.5. Sei $\mathbb{K}$ ein Körper

(1) Ein *lineares Gleichungssystem* mit m Gleichungen in n Unbekannten über $\mathbb{K}$ ist gegeben durch Angabe einer $m \times n$–Matrix $A = (a_{ij}) \in M_{m,n}(\mathbb{K})$, der *Koeffizientenmatrix* des Systems, sowie von m Elementen $b_1, \dots, b_m \in \mathbb{K}$. Eine *Lösung* des Gleichungssystems ist gegeben durch n Elemente $x_1, \dots, x_n \in \mathbb{K}$, sodass

$$\begin{gathered} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m. \end{gathered}$$

(2) Für das System $Ax = b$ aus (1) ist die *erweiterte Koeffizientenmatrix* die $m \times (n+1)$–Matrix $(A|b)$, die man erhält indem man den Vektor $b = (b_1, \dots, b_m)$ zu A als zusätzliche Spalte hinzufügt.

(3) Das Gleichungssystem heißt homogen, wenn $b_1 = \cdots = b_m = 0$ gilt. Sonst heißt das System inhomogen.

(4) Für ein inhomogenes lineares Gleichungssystem ist das *zugehörige homogene System* das System, das man erhält wen man die Matrix A unverändert lässt, aber alle b_i durch 0 ersetzt.

Unsere wesentliche Aufgabe hier wird sein, eine Methode zu finden, wie man explizit alle Lösungen eines gegebenen linearen Gleichungssystems beschreiben kann. Zunächst können wir die Frage nach Lösungen mit den Begriffen aus dem letzten Kapitel in Verbindung bringen, indem wir das Problem ein wenig uminterpretieren: Betrachten wir $b = (b_1, \dots, b_m)$ als Element von $\mathbb{K}^m$ dann können wir das lineare Gleichungssystem einfach als $Ax = b$ für $x = (x_1, \dots, x_n) \in \mathbb{K}^n$ schreiben. Das ist aber nun gerade die Gleichung $f(x) = b$, wobei $f : \mathbb{K}^n \to \mathbb{K}^m$ die lineare Abbildung $x \mapsto Ax$ ist. Damit erhalten wir aber leicht:

Proposition 3.5. *Sei $\mathbb{K}$ ein Körper, $A \in M_{m,n}(\mathbb{K})$, $b \in \mathbb{K}^m$ und $f : \mathbb{K}^n \to \mathbb{K}^m$ die lineare Abbildung $x \mapsto Ax$. Dann gilt:*

(1) Das Gleichungssystem $Ax = b$ hat genau dann eine Lösung, wenn b im Teilraum $\mathrm{Im}(f) \subset \mathbb{K}^m$ liegt.

(2) Die Menge aller Lösungen des entsprechenden homogenen Systems $Ax = 0$ ist genau der Teilraum $\mathrm{Ker}(f) \subset \mathbb{K}^n$.

(3) Falls das System $Ax = b$ eine Lösung x_0 besitzt, dann ist die Menge aller Lösungen genau die Menge $x_0 + \mathrm{Ker}(f) := \{x_0 + x : Ax = 0\}$. Man erhält also alle Lösungen des inhomogenen Systems indem man zu einer Lösung des inhomogenen Systems alle Lösungen des zugehörigen homogenen Systems addiert.

(4) Ist $m = n$ und A invertierbar, dann hat das inhomogene Gleichungssystem $Ax = b$ für jedes $b \in \mathbb{K}^n$ die eindeutige Lösung $x = A^{-1}b$.

Beweis. Nach Definition ist $\mathrm{Im}(f) = \{f(x) : x \in \mathbb{K}^n\}$ und $f(x) = Ax$, also folgt (1). Genauso folgt (2) sofort aus der Definition von $\mathrm{Ker}(f)$.

(3) Ist $f(x_0) = b$ und $f(x) = 0$, dann ist $f(x_0 + x) = f(x_0) + f(x) = b$, also ist $x_0 + \mathrm{Ker}(f)$ in der Menge aller Lösungen enthalten. Gilt umgekehrt $f(y) = b$, dann ist $f(y - x_0) = f(y) - f(x_0) = 0$, also $x := y - x_0 \in \mathrm{Ker}(f)$ und $y = x_0 + x$.

(4) Offensichtlich ist $AA^{-1}b = b$, also $x = A^{-1}b$ eine Lösung. Umgekehrt folgt aus $Ax = b$ für invertierbares A natürlich $A^{-1}(Ax) = A^{-1}b$ und $A^{-1}(Ax) = (A^{-1}A)x = \mathbb{I}_n x = x$. □

Bemerkung 3.5. Natürlich hilft uns dieses Resultat nicht dabei, lineare Gleichungssysteme tatsächlich zu lösen. Die Bedeutung liegt darin, dass wir sehen, dass wir durch Lösen linearer Gleichungssysteme auch die Kerne und Bilder von linearen Abbildungen $f : \mathbb{K}^n \to \mathbb{K}^m$ beschreiben können.

3.6. Äquivalenz von linearen Gleichungssystemen. Der erste Schritt zur Lösung linearer Gleichungssysteme ist, dass man solche Systeme äquivalent umformen kann, und bei solchen Umformungen bleibt die Lösungsmenge unverändert. Der benutzte Begriff von Äquivalenz lässt sich mit Hilfe des Matrizenkalküls sehr elegant formulieren. Dann spezifizieren wir eine spezielle Klasse von äquivalenten Umformungen, die sogenannten elementaren Zeilenoperationen.

Definition 3.6. (1) Zwei Systeme $Ax = b$ und $A'x = b'$ von m linearen Gleichungen in n Unbekannten über $\mathbb{K}$ heißen *äquivalent* wenn es eine invertierbare $m \times m$–Matrix $T \in M_m(\mathbb{K})$ gibt, sodass $A' = TA$ und $b' = Tb$.

(2) Für eine beliebige Matrix C ist eine *elementare Zeilenoperation* eine der folgenden drei Typen von Operationen:

(i) Vertausche zwei Zeilen von C.
(ii) Multipliziere eine der Zeilen von C mit einem Skalar $\lambda \in \mathbb{K}$ mit $\lambda \neq 0$.
(iii) Addiere zu einer der Zeilen von C ein beliebiges Vielfaches einer *anderen* Zeile.

Bemerkung 3.6. (1) Nach Definition der Matrizenmultiplikation sind zwei lineare Gleichungssysteme genau dann äquivalent, wenn die erweiterte Koeffizientenmatrix $(A'|b')$ des zweiten Systems aus der erweiterten Koeffizientenmatrix $(A|b)$ des ersten Systems durch Multiplikation von links mit einer invertierbaren Matrix hervorgeht.

(2) Wir können die Zeilenoperationen in Teil (2) der Definition noch expliziter beschreiben. Betrachten wir C und schreiben $c_i = (c_{i1}, \dots, c_{i\ell})$ für die i–te Zeile von C. Bezeichnen wir das Ergebnis der Zeilenoperation mit C' und die Spalten dieser Matrix mit c'_i. Für Typ (i), angewandt auf die i–te und die j–te Zeile verwenden wir die Notation $(i \leftrightarrow j)$. Die Operation vom Typ (ii), angewandt auf die i–te Zeile mit Faktor $\lambda \in \mathbb{K} \setminus \{0\}$ bezeichnen wir mit $(i \mapsto \lambda \cdot i)$, und bei Typ (iii) verwenden wir die Bezeichnung $(i \mapsto i + \lambda \cdot j)$, wenn wir zur i–ten Zeile das λ–fache der j–ten Zeile addieren (für beliebiges λ aber nur für $j \neq i$). Dann sehen die elementaren Zeilenoperationen tabellarisch wie folgt aus:

$$\begin{aligned} (i \leftrightarrow j): &\qquad c'_i = c_j, \quad c'_j = c_i, \quad c'_\ell = c_\ell \text{ für } \ell \neq i, j \\ (i \mapsto \lambda \cdot i): &\qquad c'_i = \lambda c_i, \quad c'_\ell = c_\ell \text{ für } \ell \neq i \\ (i \mapsto i + \lambda \cdot j): &\qquad c'_i = c_i + \lambda c_j, \quad c'_\ell = c_\ell \text{ für } \ell \neq i \end{aligned}$$

Für den Beweis des folgenden Satzes benötigen wir noch eine Beobachtung über die Matrizenmultiplikation. Betrachten wir eine $m \times n$–Matrix $C = (c_{ij}) \in M_{m,n}(\mathbb{K})$, dann können wir insbesondere C von links mit jeder $m \times m$–Matrix $T = (t_{ij}) \in M_m(\mathbb{K})$ multiplizieren. Nach Definition ist das Produkt $TC =: C' = (c'_{ij}) \in M_{m,n}(\mathbb{K})$ gegeben durch $c'_{ij} = \sum_k t_{ik}c_{kj}$. Bezeichnen wir wieder die Zeilenvektoren von C mit $c_i := (c_{i1}, \dots, c_{in})$ und die von C' mit c'_i für $i = 1, \dots, m$, dann bedeutet das genau, dass $c'_i = \sum_j t_{ij}c_j$ (das ist jetzt eine Gleichung von Vektoren!) gilt. Man kann also das Matrizenprodukt so interpretieren, dass die Zeilen von TC Linearkombinationen der Zeilen von C sind, wobei die Koeffizienten in diesen Linearkombinationen gerade durch die Eintragungen t_{ij} von T gegeben sind.

Satz 3.6. *(1) Äquivalenz von linearen Gleichungssystemen ist eine Äquivalenzrelation, also reflexiv, symmetrisch und transitiv (siehe Kapitel 4.2 von [EMA]).*

(2) Äquivalente lineare Gleichungssysteme haben die gleichen Lösungsmenge.

(3) Wendet man auf die erweiterte Koeffizientenmatrix eines linearen Gleichungssystems eine Folge von elementaren Zeilenoperationen an, dann erhält man die erweiterte Koeffizientenmatrix eines äquivalenten linearen Gleichungssystems.

Beweis. (1) Nach Teil (1) von Bemerkung 3.6 genügt es zu zeigen, dass $C \sim C'$ genau dann, wenn es eine invertierbare Matrix $T \in M_m(\mathbb{K})$ mit $C' = TC$ gibt, eine Äquivalenzrelation auf $M_{m,n}(\mathbb{K})$ definiert. Nun ist aber $\mathbb{I}_m$ invertierbar und $C = \mathbb{I}_m C$ (siehe Bemerkung 3.4 (2) und Teil (4) von Satz 3.3), also $C \sim C$. Für eine invertierbare Matrix T ist natürlich auch T^{-1} invertierbar und aus $C' = TC$ folgt $T^{-1}C' = C$, also folgt aus $C \sim C'$ schon $C' \sim C$. Schließlich ist für invertierbare Matrizen $T_1, T_2 \in M_m(\mathbb{K})$ auch das Produkt $T_1T_2 \in M_m(\mathbb{K})$ invertierbar, und aus $C' = T_1C$ und $C'' = T_2C'$ folgt $C'' = T_2(T_1C) = (T_1T_2)C$. Damit folgt aber aus $C \sim C'$ und $C' \sim C''$ auch $C \sim C''$, was den Beweis vervollständigt.

(2) Beschreiben (A, b) und (A', b') äquivalente Systeme, dann gibt es eine invertierbare Matrix $T \in M_m(\mathbb{K})$, sodass $A' = TA$ und $b' = Tb$ gilt. Ist $x \in \mathbb{K}^n$ eine Lösung des ersten Systems, dann gilt $Ax = b$ und daraus folgt $A'x = (TA)x = T(Ax) = Tb = b'$. Nun gilt aber auch $A = T^{-1}A'$ und $b = T^{-1}b'$ also zeigt man ganz analog, dass jede Lösung des zweiten Systems auch eine Lösung des ersten Systems ist und wir erhalten die gleichen Lösungsmengen.

(3) Wegen Teil (1) genügt es zu zeigen, dass eine einzelne elementare Zeilenoperation von der erweiterten Koeffizientenmatrix eines linearen Gleichungssystems zur erweiterten Koeffizientenmatrix eines äquivalenten Systems führt, d.h. dass man die elementaren Zeilenoperationen durch Multiplikation von links mit einer invertierbaren Matrix darstellen kann. Das müssen wir jetzt nur einfach für die drei Typen von Operationen überprüfen.

Betrachten wir zunächst eine Operation $(i \leftrightarrow j)$ vom Typ (i) und definieren wir $T = (t_{k\ell})$ durch $t_{ij} = t_{ji} = 1$, $t_{kk} = 1$ für $k \neq i, j$ und alle anderen Eintragungen gleich Null. Für eine Matrix $C \in M_{m,n}(\mathbb{K})$ mit Zeilen c_k sind die Zeilen c'_k von $C' = TC$ gegeben durch $c'_k = \sum_\ell t_{k\ell} c_\ell$. Für $k \neq i, j$ ist $t_{kk} = 1$ und $t_{k\ell} = 0$ für $\ell \neq k$, also $c'_k = c_k$. Für $k = i$ ist $t_{kj} = 1$ und $t_{k\ell} = 0$ für $\ell \neq j$, also $c'_i = c_j$ und analog ist $c'_j = c_i$. Um zu sehen, dass T invertierbar ist, kann man entweder einfach $T^2 = TT = \mathbb{I}_m$ nachrechnen.

Für die Operation $(i \mapsto \lambda \cdot i)$ vom Typ (ii) definiert man $T = (t_{k\ell})$ durch $t_{ii} = \lambda$, $t_{kk} = 1$ für $k \neq i$ und alle anderen Eintragungen gleich Null. Wie oben rechnet man sofort nach, dass TC aus C durch Anwenden der elementaren Zeilenoperation $(i \mapsto \lambda \cdot i)$ hervorgeht. Um zu sehen, dass T invertierbar ist, verifiziert man einfach, dass die Matrix $S = (s_{ij})$, die alle Eintragungen gleich Null hat, außer $s_{ii} = 1/\lambda$ und $s_{kk} = 1$ für $k \neq i$ invers zu T ist.

Für die Operation $(i \mapsto i + \lambda \cdot j)$ vom Typ (iii) definiert man $T = (t_{k\ell})$ durch $t_{kk} = 1$ für alle k, $t_{ij} = \lambda$ und alle anderen Eintragungen gleich Null. Man verifiziert wieder sofort, dass T die Zeilenoperation $(i \mapsto i + \lambda \cdot j)$ repräsentiert. Die lineare Abbildung $g : \mathbb{K}^m \to \mathbb{K}^m$ zu T bildet e_i auf $e_i + \lambda e_j$ und alle e_k für $k \neq i$ auf sich selbst ab. Damit folgt aber sofort, dass die analoge Abbildung mit $-\lambda$ statt λ invers ist (vergleiche mit Beispiel 3.4 (1)). □

Natürlich spielt es im Allgemeinen eine Rolle, in welcher Reihenfolge elementare Zeilenoperationen durchgeführt werden. Zum Beispiel erhält man offensichtlich ein anderes Ergebnis, wenn man erst $(i \leftrightarrow j)$ und dann $(i \mapsto \lambda \cdot i)$ anwendet, als wenn man erst $(i \mapsto \lambda \cdot i)$ und dann $(i \leftrightarrow j)$ anwendet. Bei Zeilenoperationen, die keine gemeinsamen Zeilen betreffen spielt die Reihenfolge aber keine Rolle, und wir werden diese Operationen oft in einem Schritt durchführen um nicht allzu viel aufschreiben zu müssen. So kann man etwa für fixes j und verschiedene $i_1, \dots, i_k$, die auch alle ungleich j sind, die Operationen $i_1 \mapsto i_1 + \lambda_1 \cdot j$ bis $i_k \mapsto i_k + \lambda_k \cdot j$ gleichzeitig durchführen.

3.7. Zeilenstufenform und reduzierte Zeilenstufenform. Der zweite Schritt zur Lösung linearer Gleichungssysteme ist die Beobachtung, dass es eine Form von erweiterten Matrizen gibt, für die man die allgemeine Lösung des Systems leicht bestimmen kann. Im letzten Schritt werden wir dann eine Methode (den "Gauß'schen Algorithmus") finden, wie man jede Matrix durch eine Folge elementarer Zeilenoperationen in diese Form bringen kann. Das dies durch eine explizite Prozedur geschieht, erhalten wir ein Rezept zur Lösung linearer Gleichungssysteme.

Definition 3.7. Sei $C = (c_{ij}) \in M_{m,n}(\mathbb{K})$ eine Matrix. Man sagt, C hat *Zeilenstufenform*, falls es eine natürliche Zahl $r \leq m$ und natürliche Zahlen $j_1, \dots, j_r$ mit

$1 \leq j_1 < j_2 < \cdots < j_r \leq n$ gibt, sodass einerseits $c_{ij_i} \neq 0$ für alle $i = 1, \dots, r$ gilt, und andererseits $c_{ij} = 0$ ist, falls $i > r$ oder $j < j_i$ gilt.

Gilt sogar $c_{ij_i} = 1$ für $i = 1, \dots, r$ und zusätzlich $c_{kj_i} = 0$ für $k < i$, dann sagt man, die Matrix A hat *reduzierte Zeilenstufenform.*

Die technische Definition der Zeilenstufenform ist etwas kompliziert, anschaulich ist aber leicht zu interpretieren, was sie bedeutet. Zunächst sind für eine Matrix C in Zeilenstufenform nach Definition ab der $(r+1)$–ten Zeile alle Zeilen Null. Für $i = 1, \dots, r$ beginnt die i–te Zeile mit Nullen, bis zum j_i–ten Eintrag, der $\neq 0$ ist, ab dann kommen beliebige Einträge. Die j_i wandern mit steigendem i immer weiter nach rechts, daher der Name Zeilenstufenform.

Für lineare Gleichungssysteme, deren erweiterte Koeffizientenmatrix reduzierte Zeilenstufenform hat, kann man die Menge aller Lösungen leicht bestimmen:

Proposition 3.7. *Sei $Ax = b$ ein lineares Gleichungssystem über einem Körper $\mathbb{K}$ mit m–Gleichungen in n Unbekannten, sodass die erweiterte Koeffizientenmatrix $(A|b)$ reduzierte Zeilenstufenform mit Indizes $r, j_1, \dots, j_r$ hat. Dann gilt:*

(1) Das System besitzt genau dann eine Lösung, wenn $j_r \leq n$ gilt.

(2) Ist das der Fall, dann erhält man die allgemeine Lösung wie folgt: Sei $J := \{1, \dots n\} \setminus \{j_1, \dots, j_r\}$, und für jedes $j \in J$ wähle ein beliebiges Element $x_j \in \mathbb{K}$. Dann setzte

$$x_{j_i} := b_j - \textstyle\sum_{k \in J, k > j_i} a_{ik} x_k$$

für $i = 1, \dots, r$, um die Lösung $(x_1, \dots, x_n)$ zu erhalten. Die allgemeine Lösung hängt also von $n - r$ freien Parametern ab.

Beweis. (1) Die Zeilen von A haben die Form $(a_{i1}, \dots, a_{in}, b_i)$. Ist $j_r = n + 1$ (das ist die einzige Möglichkeit, wie $j_r \leq n$ nicht gelten kann), dann hat die r–te Zeile die Form $(0, \dots, 0, 1)$. Diese Zeile liefert aber dann die Gleichung $0x_1 + \cdots + 0x_n = 1$, für die es natürlich keine Lösung gibt. Ist $j_r \leq n$, dann werden wir im Beweis von Teil (2) eine Lösung konstruieren.

(2) Da A in Zeilenstufenform ist, bestehen alle Zeilen ab der r–ten nur als Nullen. Diese Zeilen liefen alle die Gleichung $0x_1 + \cdots + 0x_n = 0$, sind also für das Gleichungssystem uninteressant. Ist $j_r \leq n$, dann ist die r–te Zeile der erweiterten Matrix $(0, \dots, 0, a_{rj_r}, \dots, a_{rn}, b_r)$ mit $a_{rj_r} = 1$, entspricht also der Gleichung $x_{j_r} + \sum_{k > j_r} a_{rk} x_k = b_r$. Die Zeilen darüber haben die Form $(0, \dots, 0, 1, a_{i,j_i+1}, \dots, a_{in}, b_i)$, wobei $a_{ij_\ell} = 0$ für $\ell = i + 1, \dots, r$ gilt. Diese Zeilen liefern also Gleichungen der Form

$$x_{j_i} + \textstyle\sum_{k=j_i+1}^{n} a_{ik} x_k = b_i,$$

wobei in der Summe nur Terme mit $k \in J$ eine Rolle spielen. Wählt man also beliebige Elemente x_k für $k \in J$, dann sind die Gleichungen offensichtlich äquivalent zu

$$x_{j_i} = b_i - \textstyle\sum_{k > j_i, k \in J} a_{ik} x_k$$

für alle $i = 1, \dots, r$. □

Wenn die erweiterte Koeffizientenmatrix nur Zeilenstufenform hat (und nicht reduzierte Zeilenstufenform), dann kann man die Lösungsmenge fast genauso leicht ablesen. Zunächst gilt Teil (1) der Proposition völlig unverändert, wenn $(A|b)$ nur Zeilenstufenform hat. Für den zweiten Teil, kann man wieder Elemente x_k für $k \in J$ beliebig wählen. Dann entspricht die $r$te Zeile einer Gleichung der Form $a_{rj_r} x_{j_r} + \sum_{k > j_r} a_{rk} x_k = b_r$ mit

$a_{rj_r} \neq 0$, ist also offensichtlich äquivalent zu

$$x_{j_r} = (a_{rj_r})^{-1}(b_r - \sum_{k=j_r+1}^{n} a_{rk}x_k). \tag{3.1}$$

Die Zeile darüber liefert dann eine Gleichung der Form

$$a_{r-1,j_{r-1}}x_{j_{r-1}} + \sum_{k=j_{r-1}}^{n} a_{r-1,k}x_k = b_{r-1}.$$

Da alle x_k für $k > j_{r-1}$ schon bekannt sind (entweder gewählt oder in (3.1) berechnet), ist diese Gleichung äquivalent zu

$$x_{j_{r-1}} = (a_{r-1,j_{r-1}})^{-1}(b_{r-1} - \sum_{k=j_{r-1}}^{n} a_{r-1,k}x_k),$$

wobei für x_{j_r} der in (3.1) bestimmte Wert eingesetzt wird. So kann man die Gleichungen schrittweise von unten nach oben auflösen, und abgesehen davon, dass man noch durch die Faktoren $a_{ij_i} \neq 0$ dividieren muss, ist das der einzige Unterschied zur reduzierten Zeilenstufenform.

3.8. Das Gauß'sche Eliminationsverfahren. Als letzten Schritt zur Lösung allgemeiner linearer Gleichungssysteme besprechen wir nun den sogenannten Gauß'schen Algorithmus, mit dem man jede Matrix durch elementare Zeilenoperationen auf reduzierte Zeilenstufenform bringen kann.

Satz 3.8. *Sei C eine $m \times n$–Matrix über einem Körper $\mathbb{K}$. Dann kann man C durch eine endliche Folge von elementaren Zeilenoperationen auf reduzierte Zeilenstufenform bringen.*

Beweis. Ist C die Nullmatrix, dann hat C Zeilenstufenform (mit $r = 0$). Sind nicht alle Eintragungen von C gleich Null, dann sei $j_1 \in \{1, \dots, n\}$ der kleinste Index, sodass es eine Eintragung $c_{ij_1} \neq 0$ in C gibt. Durch Anwenden der Zeilenoperationen $(i \mapsto (c_{ij_1})^{-1} \cdot i)$ und dann $(1 \leftrightarrow i)$ erhalten wir eine Matrix $C' = (c'_{ij})$ für die $c'_{1j_1} = 1$ und $c'_{ij} = 0$ für alle $j < j_1$ gilt. Wenden wir auf diese Matrix die Zeilenoperationen $(i \mapsto i - c'_{ij_1} \cdot 1)$ für $i = 2, \dots, n$ an, dann erhalten wir eine Matrix $C'' = (c''_{ij})$, für die $c''_{1j_1} = 1$ und $c''_{ij} = 0$ falls $j < j_1$ oder $j = j_1$ und $i > 1$ gilt.

Zur Vereinfachung der Notation nennen wir die resultierende Matrix jetzt wieder $C = (c_{ij})$. Dann sei $j_2 \in \{1, \dots, n\}$ der minimale Index, sodass es ein Element $c_{ij_2} \neq 0$ mit $i \geq 2$ gibt. Dann ist nach Konstruktion $j_2 > j_1$. Mittels $(i \mapsto (c_{ij_2})^{-1} \cdot i)$ und dann $(i \leftrightarrow 2)$ erreicht man eine Matrix $C' = (c'_{ij})$ deren erste Zeile mit der ersten Zeile von C übereinstimmt und $c'_{2j_2} = 1$ sowie $c'_{ij} = 0$ für $i \geq 2$ und $j < j_2$ erfüllt. Analog wie oben wendet man nun die Zeilenoperationen $(i \mapsto i - a_{ij_2} \cdot 2)$ für $i = 1, 3, \dots, m$ an, um den Rest der j_2–ten Spalte zu "vernichten".

Verfolgt man diese Prozedur weiter, dann können entweder nach "abarbeiten" der r–ten Zeile alle weiteren Zeilen nur noch aus Nullen bestehen. Dann hat die resultierende Matrix nach Konstruktion reduzierte Zeilenstufenform mit Indices $r, j_1, \dots, j_r$. Die andere Möglichkeit ist, dass man bis zur letzten Zeile Eintragungen ungleich Null hat. Dann hat aber die entstehende Matrix reduzierte Zeilenstufenform mit Indices $r = m, j_1, \dots, j_m$. □

Damit können wir nun lineare Gleichungssysteme vollständig lösen.

Beispiel 3.8. (1) Betrachten wir das folgende lineare Gleichungssystem über $\mathbb{R}$:

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= b_1 \\ 3x_1 + 2x_2 - 5x_3 &= b_2 \\ 5x_1 + x_2 - 2x_3 &= b_3 \end{aligned}$$

Dann erhalten wir als erweiterte Koeffizientenmatrix die Matrix

$$\begin{pmatrix} 2 & -1 & 3 & b_1 \\ 3 & 2 & -5 & b_2 \\ 5 & 1 & -2 & b_3 \end{pmatrix}$$

Der erste Schritt im Gauß'schen Algorithmus ist, eine Zeile, in der x_1 einen Koeffizienten $\neq 0$ hat ganz nach oben zu tauschen. Dazu brauchen wir also nichts zu tun. Als nächstes machen wir den Koeffizienten von x_1 in dieser Zeile zu 1, indem wir $(1 \mapsto \frac{1}{2} \cdot 1)$ anwenden und erhalten

$$\begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} & \frac{b_1}{2} \\ 3 & 2 & -5 & b_2 \\ 5 & 1 & -2 & b_3 \end{pmatrix}$$

Um die erste Spalte zu e_1 zu machen, wenden wir die Zeilenoperationen $(2 \mapsto 2 - 3 \cdot 1)$ und $(3 \mapsto 2 - 5 \cdot 1)$ an. Das liefert die Matrix

$$\begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} & \frac{b_1}{2} \\ 0 & \frac{7}{2} & -\frac{19}{2} & \frac{1}{2}(2b_2 - 3b_1) \\ 0 & \frac{7}{2} & -\frac{19}{2} & \frac{1}{2}(2b_3 - 5b_1) \end{pmatrix}.$$

An dieser Stelle ist natürlich ein "Abschneider" sichtbar, weil die beiden letzten Zeilen schon so ähnlich aussehen, aber wir folgen einfach dem Algorithmus. Die zweite Zeile hat bereits einen nichttrivialen Koeffizienten bei x_2, also ist keine Vertauschung nötig. Wir können also gleich die Operation $(2 \mapsto \frac{2}{7} \cdot 2)$ anwenden um diesen Koeffizienten zu 1 zu machen und erhalten

$$\begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} & \frac{b_1}{2} \\ 0 & 1 & -\frac{19}{7} & \frac{1}{7}(2b_2 - 3b_1) \\ 0 & \frac{7}{2} & -\frac{19}{2} & \frac{1}{2}(2b_3 - 5b_1) \end{pmatrix}.$$

Um als zweite Spalte e_2 zu erhalten, brauchen wir die Operationen $(1 \mapsto 1 + \frac{1}{2} \cdot 2)$ und $(3 \mapsto 3 - \frac{7}{2} \cdot 2)$. Das führt zu

$$\begin{pmatrix} 1 & 0 & \frac{1}{7} & \frac{1}{7}(2b_1 + b_2) \\ 0 & 1 & -\frac{19}{7} & \frac{1}{7}(2b_2 - 3b_1) \\ 0 & 0 & 0 & -b_1 - b_2 + b_3 \end{pmatrix}.$$

An dieser Stelle gibt es zwei Möglichkeiten: Entweder, es gilt $b_1 + b_2 - b_3 = 0$. Dann besteht die letzte Zeile aus lauter Nullen, und unsere Matrix ist bereits in reduzierter Zeilenstufenform mit Indices $r = 2$, $j_1 = 1$ und $j_2 = 2$. In diesem Fall erhalten wir die allgemeine Lösung unseres Gleichungssystems nach Teil (2) von Proposition 3.7. Betrachten wir etwa den Fall $b_1 = b_2 = 1$, $b_3 = b_1 + b_2 = 2$, dann ist die erweiterte Koeffizientenmatrix

$$\begin{pmatrix} 1 & 0 & \frac{1}{7} & \frac{3}{7} \\ 0 & 1 & -\frac{19}{7} & -\frac{1}{7} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Wir können dann für x_3 eine beliebige Zahl $t \in \mathbb{R}$ wählen und erhalten $x_1 = \frac{1}{7}(3-t)$ und $x_2 = \frac{1}{7}(-1+19t)$. Die Lösungsmenge ist also

$$\left\{\left(\tfrac{1}{7}(3-t), \tfrac{1}{7}(-1+19t), t\right) : t \in \mathbb{R}\right\} \subset \mathbb{R}^3.$$

Ist andererseits $-b_1 - b_2 + b_3 \neq 0$, dann ist es nicht notwendig, die erweiterte Koeffizientenmatrix tatsächlich in Zeilenstufenform zu bringen. Wir sehen nämlich schon, dass in diesem Fall die Zeilenstufenform die Indices $r = 3$, $j_1 = 1$, $j_2 = 2$ und $j_3 = 4$ haben wird, und es damit nach Teil (1) von Proposition 3.7 keine Lösung geben kann.

Das Beispiel erlaubt auch eine schöne Interpretation in Termen linearer Abbildungen. Ist $f : \mathbb{R}^3 \to \mathbb{R}^3$ gegeben durch

$$f(x_1, x_2, x_3) := (2x_1 - x_2 + 3x_3, 3x_1 + 2x_2 - 5x_3, 5x_1 + x_2 - 2x_3)$$

und $b = (b_1, b_2, b_3)$, dann beschreibt unser Gleichungssystem natürlich gerade die Gleichung $f(x) = b$. Also besteht das Bild $\mathrm{Im}(f)$ gerade aus jenen $b \in \mathbb{R}^3$ für die das System lösbar ist, d.h. $\mathrm{Im}(f) = \{(b_1, b_2, b_3) \in \mathbb{R}^3 : b_3 = b_1 + b_2\}$, und der Kern von f besteht aus allen Lösungen des zugehörigen homogenen Systems, also $\mathrm{Ker}(f) = \{t \cdot (-1, 19, 7) : t \in \mathbb{R}\}$.

3.9. Invertierbarkeit und Bestimmung der inversen Matrix. In diesem Abschnitt betrachten wir eine $n \times n$–Matrix A über einem Körper $\mathbb{K}$. Wir wollen ein explizites Verfahren entwickeln, das entscheidet, ob die Matrix A invertierbar ist, und falls ja, die inverse Matrix A^{-1} explizit bestimmen. Sei $T \in M_n(\mathbb{K})$ invertierbar. Dann ist für eine invertierbare Matrix $A \in M_n(\mathbb{K})$ nach Teil (3) von Satz 3.4 auch $TA \in M_n(\mathbb{K})$ invertierbar. Hat umgekehrt $A \in M_n(\mathbb{K})$ die Eigenschaft, dass TA invertierbar ist, dann ist $A = T^{-1}(TA)$ ebenfalls invertierbar.

Da elementare Zeilenoperationen durch Multiplikation mit invertierbaren Matrizen von links dargestellt werden kann sehen wir, dass wir für eine beliebige Matrix A den Gauß'schen Algorithmus anwenden können, um A in reduzierte Zeilenstufenform zu bringen und A genau dann invertierbar ist, wenn das Resultat dieser Operationen invertierbar ist. Das liefert nicht nur ein einfaches Kriterium für die Invertierbarkeit einer Matrix sondern auch einen Algorithmus zur Bestimmung der Inversen:

Satz 3.9. *(1) Eine $n \times n$–Matrix über einem Körper $\mathbb{K}$, die reduzierte Zeilenstufenform hat, ist genau dann invertierbar, wenn sie Indices $r = n$ (und damit $j_i = i$ für $i = 1, \dots, n$) hat. Das ist nur für die Einheitsmatrix $\mathbb{I}_n$ erfüllt.*

(2) Ist $A \in M_n(\mathbb{K})$ invertierbar, dann kann man A durch eine endliche Folge von elementaren Zeilenoperationen in die Einheitsmatrix $\mathbb{I}_n$ umwandeln. Man erhält die inverse Matrix A^{-1} indem man die dazu nötigen Zeilenoperationen in der selben Reihenfolge auf die Einheitsmatrix $\mathbb{I}_n$ anwendet.

Beweis. (1) Wenn $A \in M_n(\mathbb{K})$ reduzierte Zeilenstufenform hat, dann hat auch $(A|0) \in M_{n,n+1}(\mathbb{K})$ reduzierte Zeilenstufenform mit den gleichen Indices. Ist $r < n$, dann hat nach Proposition 3.7 das lineare Gleichungssystem $Ax = 0$ Lösungen $x \neq 0$. Damit ist aber die lineare Abbildung $f(x) := Ax$ nicht injektiv, also kein linearer Isomorphismus, also kann A nach Satz 3.4 nicht invertierbar sein.

Ist andererseits $r = n$, dann muss wegen $1 \leq j_1 < j_2 < \dots < j_r \leq n$ natürlich $j_i = i$ für alle $i = 1, \dots, n$ gelten. Nach Definition der reduzierten Zeilenstufenform müssen die Eintragungen a_{ij} von A dann $a_{ii} = 1$ für $i = 1, \dots, n$ und $a_{ij} = 0$ für $j \neq i$ erfüllen. Damit ist aber $A = \mathbb{I}_n$ und die Einheitsmatrix ist invertierbar.

(2) Wir können A nach Satz 3.8 durch eine endliche Folge von elementaren Zeilenoperationen in reduzierte Zeilenstufenform bringen. Wir haben bereits beobachtet, dass

die resultierende Matrix ebenfalls invertierbar, also nach Teil (1) gleich $\mathbb{I}_n$, sein muss. Diese Folge von elementaren Zeilenoperationen kann man als Linksmultiplikation mit einer invertierbaren Matrix T realisieren, also gilt $TA = \mathbb{I}_n$. Multipliziert man diese Gleichung von rechts mit A^{-1}, dann folgt $T = A^{-1}$. Dann ist aber $A^{-1} = T\mathbb{I}_n$, also die Matrix, die man erhält, indem man die gegebene Folge von elementaren Zeilenoperationen auf $\mathbb{I}_n$ anwendet. □

Bemerkung 3.9. (1) Der obige Satz liefert uns einen einfachen Algorithmus, der entscheidet, ob eine gegebene Matrix $A \in M_n(\mathbb{K})$ invertierbar ist und im Fall der Invertierbarkeit sofort die inverse Matrix berechnet. Man bildet einfach die $n \times 2n$–Matrix $(A|\mathbb{I}_n)$ die als erste n–Spalten A und als zweite n–Spalten $\mathbb{I}_n$ hat. Dann wendet man den Gauß'schen Algorithmus an um diese Matrix auf reduzierte Zeilenstufenform zu bringen. Führt das zu $r = n$ und $j_i = i$ für alle $i = 1, \dots, n$, also zu $\mathbb{I}_n$ in den ersten n Spalten, dann bilden die letzten n Spalten gerade die Inverse Matrix A^{-1}. Erhält man $j_i > i$ für ein i, dann kann A nicht invertierbar sein.

(2) In theoretischer Hinsicht sind sowohl der Gauß'sche Algorithmus als auch das hier beschriebene Verfahren zur Bestimmung der inversen Matrix sehr befriedigend. Beide Verfahren können natürlich auch leicht im Computer implementiert werden (sofern der Computer im Körper $\mathbb{K}$ rechnen kann, was aber meist nicht allzu schwierig ist). Hier sind die beiden Verfahren für allgemeine Matrizen sehr effizient. In vielen Anwendungsfällen hat man es aber mit sehr großen Matrizen zu tun (etwa in der Größenordnung 1000×1000 oder 10000×10000), und da werden diese Verfahren schon sehr aufwändig. In vielen dieser Fälle erhält man aber aus den Anwendungsproblemen zusätzliche Informationen über die Matrizen, etwa, dass sie "dünn besetzt" sind, also viele Eintragungen gleich Null sind. Für solche Systeme gibt es dann schnellere Verfahren zur Lösung.

Beispiel 3.9. Für die Matrix $A = \left(\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{smallmatrix}\right) \in M_3(\mathbb{R})$ sieht unser Algorithmus wie folgt aus:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 \\ 1 & 4 & 9 & 0 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 3 & 8 & -1 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & -1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 & -3 & 1 \end{pmatrix} \mapsto$$
$$\begin{pmatrix} 1 & 0 & -1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -\frac{3}{2} & \frac{1}{2} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 & 3 & -\frac{5}{2} & \frac{1}{2} \\ 0 & 1 & 0 & -3 & 4 & -1 \\ 0 & 0 & 1 & 1 & -\frac{3}{2} & \frac{1}{2} \end{pmatrix}$$

Damit ist A invertierbar und die "zweite Hälfte" der letzten Matrix ist die inverse Matrix A^{-1}.

KAPITEL 4

Basen und Dimension

Wir kommen nun zum theoretischen Kern der Vorlesung, nämlich zum Konzept der Dimension eines Vektorraumes. Intuitiver Hintergrund ist die Idee, dass ein Element von $\mathbb{K}^n$ durch n Elemente von $\mathbb{K}$ beschrieben werden kann, also sollte die Dimension von $\mathbb{K}^n$ gerade n sein. Wir werden sehen, dass sich dieses Konzept auf beliebige Vektorräume verallgemeinern lässt und dass ein Vektorraum mit Dimension n automatisch isomorph zu $\mathbb{K}^n$ ist.

Erzeugendensysteme und lineare Unabhängigkeit

In 2.5 haben wir die Einheitsvektoren $e_i \in \mathbb{K}^n$ kennen gelernt. Die Tatsache, dass man jedes Element von $\mathbb{K}^n$ als Linearkombination dieser Vektoren schreiben kann war für die Beschreibung der linearen Abbildungen auf $\mathbb{K}^n$ wichtig. Dieses Konzept können wir als erstes verallgemeinern.

4.1. Die erste zentrale Beobachtung ist, dass es zu jeder Teilmenge eines Vektorraumes einen kleinsten Teilraum gibt, der die Teilmenge enthält.

Satz 4.1. *Sei V ein $\mathbb{K}$–Vektorraum, und $A \subset V$ eine beliebige Teilmenge.*

(1) Es gibt einen eindeutigen Teilraum $\langle A\rangle \subset V$, sodass $A \subset \langle A\rangle$ und für jeden Teilraum $W \subset V$ mit $A \subset W$ gilt $\langle A\rangle \subset W$.

(2) Ein Element $v \in V$ liegt genau dann in dem Teilraum $\langle A\rangle$ aus (1), wenn es als Linearkombination von Elementen von A geschrieben werden kann, also wenn es endlich viele Elemente $a_i \in A$ und $r_i \in \mathbb{K}$ gibt, sodass $v = \sum_i r_i a_i$ gilt.

Beweis. (1) Für gegebenes A definiere $\langle A\rangle$ als den Durchschnitt über alle Teilräume $W \subset V$, die $A \subset W$ erfüllen. Dann ist $\langle A\rangle \subset V$ nach Korollar 2.4 ein Teilraum, der nach Konstruktion die gewünschte Eigenschaften hat. Ist $W \subset V$ ein weiterer Teilraum mit diesen Eigenschaften, dann gilt wegen $A \subset W$ und der Eigenschaft von $\langle A\rangle$ natürlich $\langle A\rangle \subset W$. Umgekehrt gilt wegen $A \subset \langle A\rangle$ und der Eigenschaft von W auch $W \subset \langle A\rangle$, also folgt $W = \langle A\rangle$.

(2) Ist $v = \sum_i r_i a_i$ für endlich viele Elemente $a_i \in A$ und $r_i \in \mathbb{K}$, dann liegen die a_i in $\langle A\rangle$ und weil $\langle A\rangle$ ein Teilraum ist, gilt $v \in \langle A\rangle$. Bezeichnen wir also mit W die Menge aller Elemente $v \in V$, die in der obigen Form geschrieben werden können, dann ist $W \subset \langle A\rangle$.

Andererseits ist für $v = \sum_i r_i a_i \in W$ natürlich $rv = \sum_i (rr_i)a_i \in W$. Analog liegt die Summe von zwei Elementen von W wieder in W, also ist $W \subset V$ ein Teilraum. Da offensichtlich $A \subset W$ gilt, folgt $\langle A\rangle \subset W$ nach (1). □

Definition 4.1. Sei V ein Vektorraum über $\mathbb{K}$ und $A \subset V$ eine Teilmenge.

(1) Der Teilraum $\langle A\rangle$ von V aus Satz 3.1 heißt das *lineare Erzeugnis* oder die *lineare Hülle* von A.

(2) A heißt ein *Erzeugendensystem* für V wenn $\langle A\rangle = V$ gilt. Man sagt dann auch, die Teilmenge A *erzeugt den Vektorraum V*.

(3) Der Vektorraum V heißt *endlich erzeugt* oder *endlichdimensional* wenn er ein endliches Erzeugendensystem besitzt.

Bemerkung 4.1. Der Ursprung des Begriffs "endlich erzeugt" ist unmittelbar einsichtig, während "endlichdimensional" an dieser Stelle eigentlich noch nicht viel Sinn macht. Wir werden erst später sehen, dass ein Vektorraum tatsächlich eine wohldefinierte Dimension hat, die genau dann endlich ist, wenn der Vektorraum endlich erzeugt ist. Der Begriff "endlichdimensional" ist aber für Vektorräume so viel üblicher als "endlich erzeugt", dass wir ihn trotzdem schon hier einführen.

Beispiel 4.1. (1) Ist V ein beliebiger $\mathbb{K}$–Vektorraum, dann ist das Erzeugnis $\langle\emptyset\rangle$ der leeren Menge der Teilraum $\{0\} \subset V$. Für einen Teilraum $W \subset V$ ist nach Definition $\langle W\rangle = W$. Insbesondere ist $V = \langle V\rangle$, also besitzt jeder Vektorraum ein Erzeugendensystem.

(2) Ist V ein $\mathbb{K}$–Vektorraum und sind $A, B \subset V$ Teilmengen mit $A \subset B$, dann gilt natürlich $\langle A\rangle \subset \langle B\rangle$, weil $\langle B\rangle$ ein Teilraum von V ist, der B und damit auch A enthält und $\langle A\rangle$ der kleinste solche Teilraum ist. Insbesondere ist für ein Erzeugendensystem A für V auch jede Obermenge von A ein Erzeugendensystem. Interessant ist es daher, möglichst kleine Erzeugendensysteme zu finden.

(3) Betrachten wir die Einheitsvektoren $e_i \in \mathbb{K}^n$ für $i = 1, \dots, n$ aus 3.2. Dort haben wir festgestellt, dass ein beliebiger Vektor $(r_1, \dots, r_n) \in \mathbb{K}^n$ als $\sum_i r_i e_i$ geschrieben werden kann. Damit ist die Menge $\{e_1, \dots, e_n\}$ ein Erzeugendensystem für $\mathbb{K}^n$, also ist insbesondere $\mathbb{K}^n$ ein endlich erzeugter $\mathbb{K}$–Vektorraum.

(4) Betrachten wir wie in Beispiel (2) von 2.2 den Körper $\mathbb{C}$ als $\mathbb{R}$–Vektorraum. Dann kann man natürlich jedes Element $z \in \mathbb{C}$ in der Form $z = a + bi$ für $a, b \in \mathbb{R}$ schreiben. Damit ist aber $\{1, i\}$ ein Erzeugendensystem für $\mathbb{C}$ als $\mathbb{R}$–Vektorraum, somit ist $\mathbb{C}$ endlich erzeugt über $\mathbb{R}$.

Betrachten wir analog $\mathbb{R}$ als Vektorraum über $\mathbb{Q}$, dann sieht das ganz anders aus. Wie aus der Einführungsvorlesung bekannt ist, ist die Menge $\mathbb{Q}$ abzählbar, während $\mathbb{R}$ überabzählbar, also echt größer ist. Nun ist leicht zu sehen, dass für endlich viele Elemente $r_1, \dots, r_n \in \mathbb{R}$ auch die Menge aller reellen Zahlen, die in der Form $\sum_i a_i r_i$ für $a_i \in \mathbb{Q}$ geschrieben werden können, ebenfalls abzählbar ist. Damit kann $\mathbb{R}$ über $\mathbb{Q}$ *nicht* endlich erzeugt sein. In Wirklichkeit sieht es noch viel schlimmer aus. Es ist auch leicht zu sehen, dass auch für eine abzählbare Teilmenge $A \subset \mathbb{R}$ der erzeugte $\mathbb{Q}$–Teilraum $\langle A\rangle \subset \mathbb{Q}$ immer noch abzählbar ist. Man kann zeigen, dass jedes Erzeugendensystem für den $\mathbb{Q}$–Vektorraum $\mathbb{R}$ kardinaläquivalent zu $\mathbb{R}$ selbst sein muss, also salopp gesprochen gleich viele Elemente haben muss wie $\mathbb{R}$ selbst.

4.2. Erzeugendensysteme und lineare Abbildungen. Wir können nun untersuchen, wie sich Erzeugendensysteme unter linearen Abbildungen verhalten.

Proposition 4.2. *Seien V und W Vektorräume über $\mathbb{K}$ und $f : V \to W$ eine lineare Abbildung.*

(1) Für eine Teilmenge $A \subset V$ mit linearer Hülle $\langle A\rangle$ ist $\langle f(A)\rangle = f(\langle A\rangle)$. Insbesondere ist für ein Erzeugendensystem A für V das Bild $f(A)$ ein Erzeugendensystem für den Vektorraum $Im(f)$.

(2) Ist $A \subset V$ ein Erzeugendensystem, dann ist die lineare Abbildung $f : V \to W$ durch die Einschränkung $f|_A : A \to W$ eindeutig bestimmt.

Beweis. (1) Nach Satz 4.1 lässt sich jedes Element $v \in \langle A\rangle$ als Linearkombination $\sum_i r_i a_i$ für endliche viele Elemente $a_i \in A$ schreiben. Damit ist aber, $f(v) = \sum_i r_i f(a_i)$, was wiederum nach Satz 4.1 in $\langle f(A)\rangle$ liegt. Damit gilt $f(\langle A\rangle) \subset \langle f(A)\rangle$.

Umgekehrt ist nach Proposition 2.7 $f(\langle A\rangle)$ ein Teilraum von W, der offensichtlich $f(A)$ enthält. Damit enthält er aber auch $\langle f(A)\rangle$, also folgt die Gleichheit.

(2) Für zwei lineare Abbildungen $f, g : V \to W$ gilt $f(v) = g(v)$ genau dann, wenn $(f - g)(v) = 0$ gilt. Damit ist aber $\{v : f(v) = g(v)\}$ genau der Kern der linearen Abbildung $f - g$ und damit ein Teilraum von V. Stimmen f und g auf einem Erzeugendensystem A für V überein, dann enthält dieser Teilraum A, also auch $\langle A\rangle = V$, also ist $f = g$. $\square$

Bemerkung 4.2. Für das Erzeugendensystem $\{e_1, \dots, e_n\} \subset \mathbb{K}^n$ haben wir Teil (2) der Proposition schon in Kapitel 3 beobachtet und benutzt.

4.3. Lineare Unabhängigkeit. Wir konnten das Erzeugnis einer Teilmenge bestimmen indem wir verstehen, welche Vektoren als Linearkombinationen von Elementen der Teilmenge geschrieben werden können. Als Gegenstück dazu werden wir uns jetzt der Frage der Eindeutigkeit von Darstellungen als Linearkombinationen widmen. Ähnlich wie man die Injektivität einer linearen Abbildung daraus schließen kann, dass sie trivialen Kern hat, können wir auch hier die Frage auf Darstellungen des Nullelements zurückführen. Das motiviert folgenden Definition.

Definition 4.3. Sei V ein Vektorraum über $\mathbb{K}$ und seien $v_1, \dots, v_k \in V$.

(1) Man sagt, *die Vektoren* $v_1,\dots,v_k$ *sind linear abhängig*, wenn es Elemente $r_i \in \mathbb{K}$ für $i = 1, \dots, k$ gibt, sodass mindestens ein $r_i \neq 0$ ist und $\sum_i r_i v_i = 0$ gilt.

(2) Die Vektoren $v_1, \dots, v_k$ heißen *linear unabhängig*, wenn sie nicht linear abhängig sind, d.h. wenn $0 = \sum_i r_i v_i$ nur für $r_1 = \cdots = r_k = 0$ gilt.

(3) Mittels (1) und (2) kann man auch sagen, wann eine endliche Teilmenge $A \subset V$ linear unabhängig ist. Eine beliebige Teilmenge $A \subset V$ heißt linear unabhängig, wenn jede endliche Teilmenge $B \subset A$ linear unabhängig ist.

(4) Eine *Basis* für den Vektorraum V ist ein linear unabhängiges Erzeugendensystem für V.

Beispiel 4.3. (1) Betrachten wir eine einelementige Teilmenge $\{v\} \subset V$. Ist diese Menge linear abhängig, dann gibt es nach Definition ein Element $r \in \mathbb{K} \setminus \{0\}$, sodass $rv = 0$ gilt. Dann ist aber $0 = r^{-1}rv = 1v = v$. Umgekehrt zeigt $1 \cdot 0 = 0$ natürlich, dass die Teilmengen $\{0\} \subset V$ linear abhängig ist. Damit ist $\{v\} \subset V$ genau dann linear unabhängig, wenn $v \neq 0$ gilt.

(2) Nach Definition ist eine beliebige Obermenge einer linear abhängigen Menge linear abhängig (man kann ja zusätzlich Elemente immer mit dem Koeffizienten 0 versehen). Insbesondere ist jede Teilmenge $A \subset V$, die den Nullvektor enthält automatisch linear abhängig. Das zeigt umgekehrt auch, dass jede Teilmenge einer linear unabhängigen Menge ebenfalls linear unabhängig ist.

(3) Sei $v \in V$ ein beliebiger Vektor mit $v \neq 0$ und $r \in \mathbb{K}$ beliebig. Dann sind die Vektoren v und rv linear abhängig, weil man $0 = r \cdot v + (-1) \cdot (rv)$ schreiben kann und $-1 \neq 0$ gilt.

Auch hier gilt wieder die Umkehrung: Sind v_1 und v_2 Vektoren, die linear abhängig sind, dann kann man einen der beiden Vektoren als ein Vielfaches des anderen schreiben. Nach Voraussetzung gibt es ja $r_1, r_2 \in \mathbb{K}$ sodass $r_1v_1 + r_2v_2 = 0$ gilt und sodass mindestens eines der r_i ungleich Null ist. Ist $r_1 \neq 0$, dann ist $v_1 = (-(r_1)^{-1}r_2)v_2$ und für $r_2 \neq 0$ ist $v_2 = (-(r_2)^{-1}r_1)v_1$.

(4) Die Vektoren $e_1, \dots, e_n \in \mathbb{K}^n$ aus Beispiel 2.5 sind linear unabhängig. wir wissen ja, dass $\sum_i r_i e_i = (r_1, \dots, r_n)$ gilt, und das ist natürlich nur dann der Nullvektor,

wenn alle r_i Null sind. Gemeinsam mit Beispiel (3) von 4.1 zeigt das, dass die Menge $\{e_1, \dots, e_n\}$ eine Basis für den Vektorraum $\mathbb{K}^n$ bildet. Diese heißt die *Standardbasis von* $\mathbb{K}^n$.

Bemerkung 4.3. (1) Aus Beispiel (3) kann man erkennen, warum wir die Definition der linearen Unabhängigkeit so mühselig (und nicht gleich für Teilmengen) formuliert haben. Die Teilmenge $\{v, rv\}$ ist nämlich nur für $r \neq 1$ linear abhängig, weil ja $\{v, 1\cdot v\} = \{v\}$ gilt.

(2) Noch etwas schwieriger wird es, wenn wir mit unendlichen Mengen umgehen müssen, was aber (zumindest vorübergehend) notwendig ist. Konkret geht es hier um die Frage der "Eindeutigkeit der Darstellung als Linearkombination", die im folgenden Satz auftaucht. Für endlich viele Vektoren $a_1, \dots, a_k$ ist das ein unproblematischer Begriff, man meint einfach, dass für Elemente $r_1, \dots, r_k, s_1, \dots, s_k \in \mathbb{K}$

$$r_1a_1 + \dots + r_ka_k = s_1a_1 + \dots + s_ka_k$$

nur dann gilt, wenn $r_i = s_i$ für alle $i = 1, \dots, k$ gilt.

Kann man ein Element v als Linearkombination von Elementen einer unendlichen Menge A darstellen, dann bedeutet das ja, dass es endlich viele Elemente $a_1, \dots, a_k \in A$ und $r_1, \dots, r_k \in \mathbb{K}$ gibt, sodass $v = r_1a_1 + \dots + r_ka_k$ gilt. Mit Eindeutigkeit der Darstellung als Linearkombination meint man hier Eindeutigkeit bis auf Hinzufügen und/oder Weglassen von Elementen mit Koeffizient 0.

Falls die folgenden Beweise Schwierigkeiten bereiten, lohnt es sich auf jeden Fall, sich den Beweis erst für endliche Mengen klar zu machen und erst dann die technischen Schwierigkeiten im Fall unendlicher Mengen "anzugehen".

Wir können nun lineare Unabhängigkeit einer Teilmenge schön charakterisieren:

Satz 4.3. *Sei V ein Vektorraum über $\mathbb{K}$. Dann sind für eine Teilmenge $A \subset V$ die folgenden Bedingungen äquivalent:*

(1) A ist linear unabhängig.

(2) Für jedes Element $v \in \langle A\rangle \subset V$ gibt es eine eindeutige Darstellung als Linearkombination von Elementen von A.

(3) Für jeden $\mathbb{K}$–Vektorraum W und jede Funktion $f : A \to W$ gibt es eine lineare Abbildung $\tilde{f} : \langle A\rangle \to W$ sodass $\tilde{f}(a) = f(a)$ für alle $a \in A$.

(4) Ist $B \subset A$ eine echte Teilmenge, also $B \neq A$ dann ist auch $\langle B\rangle \neq \langle A\rangle$. Man kann also A nicht verkleinern ohne auch das Erzeugnis zu verkleinern.

Beweis. (1)⇒(2) Nach Satz 4.1 gibt es für jedes $v \in \langle A\rangle$ eine Darstellung als Linearkombination von (endlich vielen) Elementen von A, es bleibt also nur die Eindeutigkeit zu zeigen. Haben wir zwei Darstellungen für v, dann bezeichnen wir mit $\{a_1, \dots, a_k\} \subset A$ die Menge jener Elemente von A, die in einer der beiden Darstellungen vorkommen. Indem wir die Koeffizienten von Elementen, die in den ursprünglichen Darstellungen nicht vorkommen, gleich Null setzen, dürfen wir annehmen, dass $v = \sum_{i=1}^k r_ia_i = \sum_{j=1}^k s_ja_j$ für $r_1, \dots, r_k, s_1, \dots, s_k \in \mathbb{K}$ gilt. Dann ist aber $0 = v - v = \sum_{i=1}^n (r_i - s_i)a_i$. Da A linear unabhängig ist, ist auch $\{a_1, \dots, a_k\}$ linear unabhängig, also muss nach Definition $r_i - s_i = 0$, also $r_i = s_i$ für alle $i = 1, \dots, k$ gelten, also sind die beiden Darstellungen gleich.

(2)⇒(3) Ist $v \in \langle A\rangle$, dann gibt es nach (2) eindeutig bestimmte Elemente $a_1, \dots a_k \in A$ und $r_1, \dots, r_k \in \mathbb{K}$, sodass $v = \sum_{i=1}^k r_ia_i$ gilt, und wir definieren $\tilde{f}(v) := \sum_{i=1}^k r_if(a_i)$.

(Man bemerke, dass sich das nicht ändert, wenn wir Elemente mit Koeffizient 0 hinzufügen oder weglassen.) Für $a \in A$ ist diese Darstellung durch $a = 1 \cdot a$ gegeben, also folgt $\tilde{f}(a) = 1 \cdot f(a) = f(a)$. Damit bleibt zu zeigen, dass $\tilde{f}$ linear ist.

Ist $v = \sum_i r_i a_i$ und $r \in \mathbb{K}$, dann ist $rv = \sum_i (rr_i)a_i$, also gilt

$$\tilde{f}(rv) = \sum_i (rr_i) f(a_i) = r \sum_i r_i f(a_i) = r\tilde{f}(v)$$

Für ein weiteres Element $w \in \langle A \rangle$ können wir durch Hinzufügen mit Koeffizient 0 erreichen, dass wir Elemente $a_1, \dots, a_\ell \in A$ und $r_1, \dots, r_\ell, s_1, \dots, s_\ell \in \mathbb{K}$ finden, für die $v = \sum_{i=1}^{\ell} r_i a_i$ und $w = \sum_{j=1}^{\ell} s_j a_j$ gilt. Dann ist aber $v + w = \sum_{i=1}^{\ell}(r_i + s_i)a_i$ und somit

$$\tilde{f}(v + w) = \sum_{i=1}^{\ell}(r_i + s_i) f(a_i) = \sum_{i=1}^{\ell} r_i f(a_i) + \sum_{j=1}^{\ell} s_j f(a_j) = \tilde{f}(v) + \tilde{f}(w).$$

(3)$\Rightarrow$(1) Für ein Elemet $a \in A$ betrachte die Funktion $\chi_a : A \to \mathbb{K}$, die gegeben ist durch $\chi_a(a) = 1$ und $\chi_a(\tilde{a}) = 0$ für $\tilde{a} \neq a$. Nach Bedingung (3) finden wir für jedes $a \in A$ eine lineare Abbildung $\tilde{\chi}_a : \langle A \rangle \to \mathbb{K}$, deren Einschränkung auf A mit χ_a übereinstimmt.

Seien nun $a_i \in A$ und $r_i \in \mathbb{K}$ für $i = 1, \dots, k$ so, dass $\sum_i r_i a_i = 0$ gilt. Dann gilt für jedes $j = 1, \dots, k$ natürlich

$$0 = \tilde{\chi}_{a_j}\left(\sum_i r_i a_i\right) = \sum_i r_i \chi_{a_j}(a_i) = r_j.$$

Das bedeutet aber nach Definition, dass A linear unabhängig ist.

(4) Wir zeigen, dass eine Teilmenge $A \subset V$ genau dann linear abhängig ist, wenn es eine echte Teilmenge $B \subset A$ gibt, die $\langle B \rangle = \langle A \rangle$ erfüllt.

Ist A linear abhängig, dann finden wir Elemente $a_1, \dots, a_k \in A$ und $r_1, \dots, r_k \in \mathbb{K}$ die nicht alle gleich Null sind, sodass $\sum_{i=1}^{k} r_i a_i = 0$ gilt. Durch umnummerieren können wir annehmen, dass $r_k \neq 0$. Dann erhalten wir aber $-r_k a_k = \sum_{i=1}^{k-1} r_i a_i$ und indem wir durch $-r_k$ dividieren eine Darstellung von a_k als Linearkombination von $a_1, \dots, a_{k-1}$. Betrachten wir nun die echte Teilmenge $B := A \setminus \{a_k\}$ von A, dann gilt natürlich $\langle B \rangle \subset \langle A \rangle$. Wegen $a_1, \dots, a_{k-1} \in B$ gilt aber umgekehrt $a_k \in \langle B \rangle$ und damit $A \subset \langle B \rangle$, also $\langle A \rangle \subset \langle B \rangle$ und die beiden Erzeugnisse stimmen überein.

Ist umgekehrt $B \subset A$ eine echte Teilmenge mit $\langle B \rangle = \langle A \rangle$, dann sei $a \in A \setminus B$. Wegen $a \in \langle B \rangle$ finden wir $b_1, \dots, b_k \in B$ und $r_1, \dots, r_k \in \mathbb{K}$, sodass $a = \sum_i r_i b_i$ gilt. Damit erhalten wir aber $0 = (-1)a + \sum_i r_i b_i$ was zeigt, dass die Menge A linear abhängig ist. □

Basen

Einige grundlegende Eigenschaften von Basen ergeben sich nun ganz einfach. Viel schwieriger zu beweisen, aber von überragender Wichtigkeit ist, dass je zwei Basen eines Vektorraumes gleich viele Elemente haben. Damit wird die Anzahl der Elemente einer Basis eine Zahl, die einem Vektorraum in natürlicher Weise zugeordnet ist. Diese Zahl heißt die Dimension des Vektorraumes.

4.4. Grundlegende Eigenschaften von Basen. Mit Hilfe von Satz 4.3 können wir Basen eines Vektorraumes relativ leicht charakterisieren.

PROPOSITION 4.4. *Für eine Teilmenge $\mathcal{B}$ eines $\mathbb{K}$–Vektorraumes V sind die folgenden Bedingungen äquivalent:*

(i) *$\mathcal{B}$ ist eine Basis für V.*
(ii) *$\mathcal{B}$ ist ein minimales Erzeugendensystem von V, d.h. $\langle \mathcal{B} \rangle = V$ und für jede echte Teilmenge $\mathcal{B}' \subset \mathcal{B}$ ist $\langle \mathcal{B}' \rangle \neq V$.*

(iii) *$\mathcal{B}$ ist eine maximale linear unabhängige Teilmenge von V, d.h. $\mathcal{B}$ ist linear unabhängig, aber jede echte Obermenge von $\mathcal{B}$ ist linear abhängig.*

Beweis. Nach Satz 4.3 ist ein Erzeugendensystem von V genau dann linear unabhängig, wenn keine echte Teilmenge davon ebenfalls ein Erzeugendensystem ist. Das bedeutet aber genau (i)$\Leftrightarrow$(ii).

(i)$\Rightarrow$(iii) Ist $\mathcal{B}$ eine Basis, dann ist $\mathcal{B}$ nach Definition linear unabhängig. Andererseits ist $\langle\mathcal{B}\rangle = V$, also gilt auch für jede echte Obermenge $\mathcal{B}'$ von $\mathcal{B}$, dass $\langle\mathcal{B}'\rangle = V$ ist. Damit kann aber $\mathcal{B}'$ nach Satz 4.3 nicht linear unabhängig sein, weil die echte Teilmenge $\mathcal{B} \subset \mathcal{B}'$, das gleiche Erzeugnis wie $\mathcal{B}'$ hat.

(iii)$\Rightarrow$(i) Sei $\mathcal{B}$ eine maximale linear unabhängige Teilmenge von V und $v \in V$ ein beliebiges Element. Dann ist $\mathcal{B} \cup \{v\}$ linear abhängig, also kann man 0 in nichttrivialer Weise als Linearkombination von Elementen dieser Menge schreiben. Da $\mathcal{B}$ linear unabhängig ist, muss v in dieser Linearkombination vorkommen und einen Koeffizienten ungleich Null haben. Das bedeutet aber, dass es Elemente $b_1, \dots, b_k \in \mathcal{B}$, $r, r_1, \dots, r_k \in \mathbb{K}$ gibt, sodass $r \neq 0$ und $rv + \sum_i r_i b_i = 0$ gilt. Dann ist aber $v = -\frac{1}{r}\sum r_i b_i \in \langle\mathcal{B}\rangle$. Da $v \in V$ beliebig war, ist $\mathcal{B}$ ein Erzeugendensystem für V, also eine Basis. □

Korollar 4.4. *Jeder endlich erzeugte $\mathbb{K}$–Vektorraum V besitzt eine (endliche) Basis*

Beweis. Da V endlich erzeugt ist, finden wir eine endliche Teilmenge $A \subset V$ mit $\langle A\rangle = V$. Ist A linear unabhängig, dann ist A eine Basis und wir sind fertig. Falls nicht, dann gibt es nach Satz 4.3 eine echte Teilmenge $A' \subset A$, die immer noch ein Erzeugendensystem für V ist. Ist A' linear unabhängig, dann sind wir fertig, wenn nicht, dann wiederholen wir den Prozess. Da die Ausgangsmenge A endlich war und wir jedes Mal zu einer echten Teilmenge übergehen, dann müssen wir nach endlich vielen Schritten zu einer Basis (oder zur leeren Menge, die eine Basis für den Vektorraum $\{0\}$ ist) gelangen. □

4.5. Der Austauschsatz. Der Schlüssel zum Vergleich der Größe verschiedener Basen eines Vektorraumes ist der Austauschsatz von Steinitz. Bevor wir diesen Satz formulieren, beweisen wir ein Lemma.

Lemma 4.5 (Austauschlemma). *Sei V ein $\mathbb{K}$–Vektorraum, $A \subset V$ eine beliebige Teilmenge und $v \in \langle A\rangle \subset V$ ein beliebiges Element. Sei $a_0 \in A$ ein Element, sodass es eine Darstellung von v als Linearkombination von Elementen von A gibt, in der a_0 einen Koeffizienten $\neq 0$ hat. Dann erfüllt die Menge $A' := (A \setminus \{a_0\}) \cup \{v\}$, die man erhält indem man in A den Vektor a_0 durch v ersetzt $\langle A'\rangle = \langle A\rangle$.*

Beweis. Nach Voraussetzung ist $v \in \langle A\rangle$, also $A' \subset \langle A\rangle$, also $\langle A'\rangle \subset \langle A\rangle$.

Umgekehrt gibt es nach Voraussetzung Elemente $a_1, \dots, a_n \in A \setminus \{a_0\}$ und Skalare $r_0, \dots, r_n$ mit $r_0 \neq 0$, sodass $v = r_0 a_0 + \sum_{i=1}^n r_i a_i$ gilt. Damit ist aber

$$a_0 = (r_0)^{-1}(v - \textstyle\sum_{i=1}^n r_i a_i) \in \langle A'\rangle.$$

Da offensichtlich $A \setminus \{a_0\} \subset \langle A'\rangle$ gilt, erhalten wir $A \subset \langle A'\rangle$ und daher $\langle A\rangle \subset \langle A'\rangle$. □

Satz 4.5 (Austauschsatz von Steinitz). *Sei V ein endlich erzeugter $\mathbb{K}$–Vektorraum, und sei $\{v_1, \dots, v_n\}$ ein Erzeugendensystem für V. Dann gilt:*

(1) Ist $A \subset V$ eine linear unabhängige Teilmenge, dann ist A endlich und hat höchstens n Elemente. Ist $A = \{a_1, \dots, a_k\}$ mit $k < n$, dann kann man die Elemente v_i so umnummerieren, dass $\{a_1, \dots, a_k, v_{k+1}, \dots, v_n\}$ ein Erzeugendensystem für V ist.

(2) Ist $A \subset V$ ein beliebiges Erzeugendensystem, dann gibt es eine endliche Teilmenge $A' \subset A$ mit höchstens n Elementen, die ebenfalls ein Erzeugendensystem für V ist.

Beweis. (1) Wähle ein beliebiges Element $a_1 \in A$. Da $\{v_1, \dots, v_n\}$ ein Erzeugendensystem für V ist, finden wir Skalare $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, sodass $a_1 = \sum_i \lambda_i v_i$ gilt. Da A linear unabhängig ist, ist $a_1 \neq 0$, also muss zumindest ein $\lambda_i \neq 0$ sein, und durch umnummerieren der v_i erreichen wir $\lambda_1 \neq 0$. Nach dem Austauschlemma ist auch $\{a_1, v_2, \dots, v_n\}$ ein Erzeugendensystem für V. Falls $A = \{a_1\}$ gilt, sind wir fertig.

Falls nicht, dann wählen wir ein beliebiges Element $a_2 \in A \setminus \{a_1\}$. Dann gibt es wiederum Elemente $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, sodass $a_2 = \lambda_1 a_1 + \sum_{i=2}^n \lambda_i v_i$ gilt. Wären alle λ_i für $i \geq 2$ gleich Null, dann würde $a_2 = \lambda_1 a_1$, also $\lambda_1 a_1 + (-1)a_2 = 0$ gelten, was ein Widerspruch zu A linear unabhängig wäre. Somit gibt es ein $i \geq 2$ mit $\lambda_i \neq 0$, und durch umnummerieren erreichen wir $\lambda_2 \neq 0$. Nach dem Austauschlemma ist wieder $\{a_1, a_2, v_3, \dots, v_n\}$ ein Erzeugendensystem für V.

Nehmen wir induktiv an, dass wir für ein $\ell < n$ bereits $\{a_1, \dots, a_\ell\} \subset A$ gefunden haben, sodass (für geeignete Nummerierung der v_i) $\{a_1, \dots, a_\ell, v_{\ell+1}, \dots v_n\}$ ein Erzeugendensystem für V ist. Ist $A = \{a_1, \dots, a_\ell\}$ dann sind wir fertig. Wenn nicht, dann wählen wir ein beliebiges Element $a_{\ell+1} \in A \setminus \{a_1, \dots, a_\ell\}$. Wie zuvor erhalten wir $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ sodass $a_{\ell+1} = \sum_{i=1}^{\ell} \lambda_i a_i + \sum_{i=\ell+1}^{n} \lambda_i v_i$ gilt. Aus der linearen Unabhängigkeit von A folgt wieder, dass es ein $i > \ell$ mit $\lambda_i \neq 0$ geben muss, und durch Umnummerieren erhalten wir $\lambda_{\ell+1} \neq 0$. Nach dem Austauschlemma bildet die Menge $\{a_1, \dots, a_{\ell+1}, v_{\ell+2}, \dots v_n\}$ ein Erzeugendensystem für V.

Nach Induktion können wir nun diesen Prozess fortsetzen, bis wir entweder die ganze Menge A "aufgebraucht" haben, oder wir haben im letzten Schritt Elemente $a_1, \dots, a_n \in A$ gefunden, die ein Erzeugendensystem für V bilden. Im ersten Fall muss natürlich A endlich mit weniger als n Elementen sein, und der letzte Schritt liefert genau die Behauptung des Satzes. Im zweiten Fall ist $\langle \{a_1, \dots, a_n\} \rangle = V = \langle A \rangle$. Da A linear unabhängig ist, muss nach Satz 4.3 $A = \{a_1, \dots, a_n\}$ gelten, also ist A wieder endlich mit höchstens n Elementen.

(2) Ist $A = \{0\}$ dann sind wir fertig, weil wir $A' = \emptyset$ wählen können. Sonst wähle ein Element $0 \neq a_1 \in A$. Wie in (1) können wir a_1 als Linearkombination der v_i schreiben und durch umnummerieren erreichen wir, dass der Koeffizient von v_1 ungleich Null ist. Nach dem Austauschlemma ist dann $\{a_1, v_2, \dots, v_n\}$ ein Erzeugendensystem für V. Ist $\langle \{a_1\} \rangle = V$, dann sind wir fertig. Falls nicht, dann muss es ein Element $a_2 \in A \setminus \langle \{a_1\} \rangle$ geben. Wäre das nicht der Fall, dann wäre $A \subset \langle \{a_1\} \rangle$, ein Widerspruch, weil A ein Erzeugendensystem von V und $\langle \{a_1\} \rangle \neq V$ ist. Dann können wir wieder a_2 als Linearkombination von a_1 und den v_i schreiben und weil $a_2 \notin \langle \{a_1\} \rangle$ muss zumindest ein v_i einen Koeffizienten haben, der ungleich Null ist. Durch Umnummerieren trifft das für v_2 zu, und nach dem Austauschlemma ist $\{a_1, a_2, v_3, \dots, v_n\}$ ein Erzeugendensystem für V.

Nehmen wir induktiv an, dass wir für ein $\ell < n$ schon $a_1, \dots, a_\ell \in A$ gefunden haben, sodass $\{a_1, \dots, a_\ell, v_{\ell+1}, \dots, v_n\}$ ein Erzeugendensystem für V ist. Falls $\{a_1, \dots, a_\ell\}$ ein Erzeugendensystem für V ist, sind wir fertig. Falls nicht, muss es wie oben ein Element $a_{\ell+1} \in A \setminus \langle \{a_1, \dots, a_\ell\} \rangle$ geben, weil sonst A kein Erzeugendensystem sein könnte. Aus dieser Bedingung folgt auch, dass in der Darstellung von $a_{\ell+1}$ als Linearkombination von $a_1, \dots, a_\ell, v_{\ell+1}, \dots, v_n$ zumindest ein v_i einen Koeffizienten ungleich Null hat, und durch umnummerieren ist $v_{\ell+1}$ dieses Element. Nach dem Austauschlemma ist dann $\{a_1, \dots, a_{\ell+1}, v_{\ell+2}, \dots, v_n\}$ ein Erzeugendensystem und (2) folgt mittels Induktion. $\square$

Aus diesem Satz können wir nun die fundamentalsten Tatsachen über endlich erzeugte Vektorräume direkt ablesen:

Korollar 4.5. *Sei V ein endlich erzeugter $\mathbb{K}$–Vektorraum Dann gilt:*

(1) *Jede Basis von V ist endlich, und je zwei Basen haben gleich viele Elemente. Diese Anzahl heißt die Dimension $\dim(V)$ von V über $\mathbb{K}$.*
(2) *Jede linear unabhängige Teilmenge von V kann zu einer Basis erweitert werden.*
(3) *Jedes Erzeugendensystem von V enthält eine Basis für V.*
(4) *Eine linear unabhängige Teilmenge von V kann höchstens $\dim(V)$ viele Elemente haben.*
(5) *Ein Erzeugendensystem für V muss mindestens $\dim(V)$ viele Elemente haben.*
(6) *Eine Teilmenge von V, die $\dim(V)$ viele Elemente enthält und entweder linear unabhängig oder ein Erzeugendensystem ist, ist eine Basis von V.*

Beweis. Nach Korollar 4.4 bestizt V zumindest eine endliche Basis. Sei nun $\mathcal{B} = \{v_1, \dots, v_n\}$ eine beliebige endliche Basis für V. Wenden wir den ersten Teil des Austauschsatzes auf das Erzeugendensystem $\mathcal{B}$ von V an, dann folgt, dass jede linear unabhängige Teilmenge von V endlich ist und höchstens n Elemente hat. Umgekehrt liefert der erste Teil des Austauschsatzes angewandt auf n–elmentige, linear unabhängige Teilmenge $\mathcal{B}$ natürlich, das jedes Erzeugendensystem für V mindestens n Elemente haben muss.

Gemeinsam sagen diese Bedingungen aber, dass jede Basis für V endlich ist und genau n Elemente hat. Damit ist der Beweis von (1) vollständig und da wir nun wissen, dass $\dim(V)$ Sinn macht, sind auch (4) und (5) erledigt. Weiters sehen wir, dass eine linear unabhängige Menge mit $\dim(V)$ Elementen automatisch maximal linear unabhängig, also nach Proposition 4.4 eine Basis ist. Analog ist ein Erzeugendensystem mit $\dim(V)$ Elementen automatisch ein minimales Erzeugendensystem also ebenfalls eine Basis, was den Beweis von (6) vervollständigt.

Für (2) wenden wir Teil (1) des Austauschsatzes auf eine Basis $\{v_1, \dots, v_n\}$ und eine beliebige linear unabhängige Teilmenge A an. Das zeigt, dass A zu einem Erzeugendensystem mit $\dim(V)$ vielen Elementen, also zu einer Basis erweitert werden kann. Analog folgt aus Teil (2) des Austauschsatzes sofort, dass ein beliebiges Erzeugendensystem für V ein Erzeugendensystem mit $\dim(V)$ vielen Elementen, also eine Basis, enthält. □

Bemerkung 4.5. Mit diesem Korollar sehen wir, dass jeder endlich erzeugte Vektorraum über einem Körper $\mathbb{K}$ endliche Dimension hat, wodurch sich der Begriff "endlichdimensional" für diese Räume erklärt. Vektorräume, die kein endliches Erzeugendensystem besitzen, nennt man unendlichdimensional. Man schreibt in diesem Fall formal $\dim(V) = \infty$.

Beispiel 4.5. (1) Aus Beispiel (4) von 4.3 wissen wir, dass für jeden Körper $\mathbb{K}$ und jedes $n \in \mathbb{N}$, der $\mathbb{K}$–Vektorraum $\mathbb{K}^n$ die Standardbasis $\{e_1, \dots, e_n\}$ besitzt. Damit gilt $\dim(\mathbb{K}^n) = n$, was eine der Motivationen für den Dimensionsbegriff ist. Wir werden in Kürze sehen, dass jeder $\mathbb{K}$–Vektorraum der Dimension n isomorph zu $\mathbb{K}^n$ ist.

(2) Für eine Menge X betrachte den Vektorraum $\mathcal{F}(X, \mathbb{R})$ aller Funktionen $f : X \to \mathbb{R}$. Sei $A \subset X$ eine Teilmenge und für $a \in A$ sei $f_a : X \to \mathbb{R}$ gegeben durch $f(a) = 1$ und $f(x) = 0$ für $x \neq a$. Dann ist die Menge $\{f_a : a \in A\} \subset \mathbb{R}^X$ linear unabhängig. Ist nämlich $\lambda_1 f_{a_1} + \dots + \lambda_n f_{a_n} = 0$ für Elemente $a_i \in A$ und Skalare $\lambda_i \in \mathbb{K}$, dann liefert Auswerten im Punkt a_i sofort $0 = \lambda_i$, und das funktioniert für alle $i = 1, \dots, n$. Ist $X = \{x_1, \dots, x_n\}$ endlich, dann ist klarerweise $\{f_{x_1}, \dots, f_{x_n}\}$ eine Basis für $\mathbb{R}^X$, also hat dieser Vektorraum Dimension n. Ist X unendlich, dann bilden

die f_x für $x \in X$ eine unendliche linear unabhängige Teilmenge von $\mathcal{F}(X, \mathbb{R})$, also ist in diesem Fall $\dim(\mathcal{F}(X, \mathbb{R})) = \infty$ nach Teil (1) des Austauschsatzes.

In dem Fall, dass X unendlich ist, bildet $\{f_x : x \in X\}$ zwar eine linear unabhängige Teilmenge von $\mathcal{F}(X, \mathbb{R})$, aber *kein* Erzeugendensystem und somit auch keine Basis. Kann man nämlich eine Funktion $f : X \to \mathbb{R}$ als Linearkombination von Elementen dieser Menge schreiben, dann ist $f = \lambda_1 f_{x_1} + \cdots + \lambda_n f_{x_n}$ für endlich viele Elemente $x_i \in X$ und Skalare $\lambda_i \in \mathbb{K}$. Damit nimmt aber die Funktion f nur in endlich vielen Punkten einen Wert $\neq 0$ an. Insbesondere kann für unendliches X die konstante Funktion $f(x) = 1$ für alle $x \in X$ nicht als Linearkombination von Funktionen der Form f_x geschrieben werden und liegt damit nach Satz 4.1 nicht in $\langle\{f_x : x \in X\}\rangle$.

4.6. Exkurs: Basen für unendlichdimensionale Vektorräume. Dieser Abschnitt ist für das weitere Verständnis der Vorlesung nicht notwendig und dient nur als Zusatzinformation für Interessierte. Wir werden (mit Hilfe des Auswahlaxioms) beweisen, dass jeder Vektorraum über einem Körper $\mathbb{K}$ eine Basis besitzt. Basen in diesem Sinn sind aber nur für sehr wenige Beispiele unendlichdimensionaler Räume wirklich interessant (und für diese Beispiele ist die Existenz von Basen leicht zu sehen). Zum Studium der meisten unendlichdimensionalen Räume (insbesondere von Räumen von Folgen und Funktionen, die in der Analysis interessant sind) muss man zusätzlich zu den Begriffen der linearen Algebra noch Begriffe aus der Topologie wie Stetigkeit und Konvergenz zur Hilfe nehmen.

In seiner ursprünglichen Formulierung besagt das Auswahlaxiom dass man aus jedem Mitglied einer beliebigen Familie $\{A_i : i \in I\}$ von Mengen ein Element auswählen kann, oder formaler, dass es eine Funktion $f : I \to \cup_{i \in I} A_i$ gibt, sodass $f(i) \in A_i$ für alle $i \in I$ gilt (eine sogenannte "Auswahlfunktion").

Es gibt mehrere äquivalente Umformulierungen dieses Axioms. Die außerhalb der Mengenlehre wichtigste dieser Umformulierungen ist das sogenannte Zorn'sche Lemma:

Lemma 4.6 (Zorn'sches Lemma). *Sei M eine Menge und $\leq$ eine Halbordnung auf M, die folgende Bedingung erfüllt: Ist $A \subset M$ eine Teilmenge, sodass $\leq$ eine Totalordnung auf A definiert (also je zwei Elemente von A vergleichbar sind), dann gibt es ein Element $x \in M$, sodass $a \leq x$ für alle $a \in A$ gilt. Dann gibt es ein Element $x_0 \in M$, sodass $x \in M$ und $x_0 \leq x$ schon $x = x_0$ impliziert. ("Falls jede Kette in M eine obere Schranke besitzt, dann existieren maximale Elemente.")*

Mit Hilfe dieses Lemmas können wir nun die Existenz von Basen in beliebigen Vektorräumen beweisen. Wir zeigen ein etwas stärkeres Resultat, das auch gleich die Verallgemeinerung der Teile (2) und (3) von Korollar 4.5 beinhaltet.

Satz 4.6. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, $A \subset V$ eine linear unabhängige Teilmenge und $E \subset V$ ein Erzeugendensystem mit $A \subset E$. Dann gibt es eine Basis $\mathcal{B}$ für V mit $A \subset \mathcal{B} \subset E$.*

Beweis. Betrachte die Familie $\mathcal{M}$ aller linear unabhängigen Teilmengen $C \subset V$, die $A \subset C \subset E$ erfüllen. Wir definieren eine Halbordnung auf dieser Menge durch $C \leq D$ genau dann, wenn $C \subset D$. Sei $\mathcal{A} \subset \mathcal{M}$ eine Teilmenge, auf der dies eine Totalordnung ist und betrachte $X := \cup_{C \in \mathcal{A}} C$. Sind $x_1, \dots, x_n \in X$ und $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ endlich viele Elemente sodass $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$ gilt, dann gibt es $C_1, \dots, C_n \in \mathcal{A}$, sodass $x_i \in C_i$ für jedes i. Da $\mathcal{A}$ durch die Inklusion total geordnet ist, gilt für $i, j = 1, \dots, n$ entweder $C_i \subset C_j$ oder $C_j \subset C_i$. Daraus folgt aber leicht, dass es ein i_0 gibt, sodass $C_i \subset C_{i_0}$ für alle i gilt. Da die Menge C_{i_0} in $\mathcal{A} \subset \mathcal{M}$ liegt ist sie linear unabhängig, also folgt $\lambda_i = 0$

für alle i. Somit ist X linear unabhängig. Offensichtlich gilt $A \subset X \subset E$, weil das für jedes $C \in \mathcal{A}$ gilt, also ist $X \in \mathcal{M}$ eine obere Schranke für A.

Nach dem Zorn'schen Lemma erhalten wir ein maximales Element $X_0 \in \mathcal{M}$. Damit ist $A \subset X_0 \subset E$ und X_0 linear unabhängig. Da X_0 ein maximales Element von $\mathcal{M}$ ist, ist für jedes Element $y \in E \setminus X_0$ die Menge $X_0 \cup \{y\}$ linear abhängig, und man sieht leicht, dass dies nur möglich ist, wenn $y \in \langle X_0 \rangle$ gilt. Damit folgt aber $E \subset \langle X_0 \rangle$, und weil E ein Erzeugendensystem ist, ist auch X_0 ein Erzeugendensystem, also eine Basis. $\square$

Man kann auch für unendlichdimensionale Vektorräume einen sinnvollen Begriff von Dimension einführen. Man kann nämlich zeigen, dass es zwischen je zwei Basen eines Vektorraumes immer eine bijektive Funktion gibt, diese Basen also als Mengen die gleiche Kardinalität besitzen. Diese Kardinalität ist dann eine (unendliche) Kardinalzahl, als die man die Dimension des Raumes definiert. Die einzigen unendlichdimensionalen Räume, die man alleine mit Methoden der linearen Algebra gut in den Griff bekommt, sind Räume abzählbarer Dimension (zum Beispiel der Raum $\mathbb{K}[x]$ aller Polynome über $\mathbb{K}$). Äquivalent zu abzählbarer Dimension ist auch, dass ein Raum ein abzählbares Erzeugendensystem besitzt. Für solche Räume kann man Satz 4.6 auch ohne Auswahlaxiom relativ leicht (mit vollständiger Induktion) beweisen.

Dimensionssätze

Als nächstes beweisen wir einige fundamentale Dimensionssätze.

4.7. Dimension von Teilräumen. Für einen sinnvollen Dimensionsbegriff muss wohl die Dimension eines Teilraumes $W \subset V$ höchstens so groß sein, wie $\dim(V)$. Das ist tatsächlich der Fall, der Beweis enthält aber eine kleine Tücke, weil es nicht automatisch klar ist, dass ein Teilraum eines endlichdimensionalen Vektorraumes selbst endlich erzeugt ist.

PROPOSITION 4.7. *Sei V ein endlichdimensionaler $\mathbb{K}$–Vektorraum und $W \subset V$ ein Teilraum. Dann ist* $\dim(W) \leq \dim(V)$ *und ist* $\dim(W) = \dim(V)$*, dann ist* $W = V$.

BEWEIS. Sei $n := \dim(V)$. Ist $A \subset W$ eine linear unabhängige Teilmenge, dann ist A natürlich auch eine linear unabhängige Teilmenge von V. Damit ist nach Teil (1) des Austauschsatzes 4.5 die Menge A endlich und hat höchstens n Elemente. Damit gibt es eine maximale Größe $k \leq n$ für linear unabhängige Teilmenge von W. Eine linear unabhängige Teilmenge von W mit k Elementen ist aber dann natürlich eine maximale linear unabhängige Teilmenge von W, also nach Proposition 4.4 eine Basis für W. Damit ist aber $\dim(W) = k \leq n = \dim(V)$.

Ist $\dim(W) = n$, dann sei $\mathcal{B}$ eine Basis für W. Dann ist $\mathcal{B} \subset V$ eine linear unabhängige Teilmenge mit $\dim(V)$ vielen Elementen, also eine Basis für V. Damit ist aber $V = \langle \mathcal{B} \rangle = W$. $\square$

BEISPIEL 4.7. Mit diesem Resultat können wir die Beschreibung aller Teilräume von $\mathbb{R}^2$ und $\mathbb{R}^3$, die wir in Kapitel 2 nur mit Mühe zustande gebracht haben, sehr einfach erhalten:

Im Fall von $\mathbb{R}^2$ gibt es außer $\mathbb{R}^2$ nur Teilräume der Dimensionen 0 und 1. Für einen Teilraum $V \subset \mathbb{R}^2$ der Dimension 0 bildet die leere Menge eine Basis, also ist $V = \{0\}$. Ist $V \subset \mathbb{R}^2$ ein Teilraum der Dimension 1, dann finden wir eine Basis $\{v\}$ für V. Da diese Menge linear unabhängig ist, ist $v \neq 0$ und $V = \langle \{v\} \rangle = \{tv : t \in \mathbb{R}\}$ ist die Gerade durch 0 und v.

Für $\mathbb{R}^3$ sind die möglichen Dimensionen für echte Teilräume 0, 1 oder 2, und wie im Fall von $\mathbb{R}^2$ ist der einzige nulldimensionale Teilraum $\{0\}$ und die eindimensionalen Teilräume sind die Geraden durch den Nullpunkt. Im Fall eines zweidimensionalen Teilraumes $V \subset \mathbb{R}^2$ gibt es eine Basis $\{v_1, v_2\}$ für V, und die lineare Unabhängigkeit bedeutet in diesem Fall genau, dass die beiden Vektoren nicht auf einer Geraden durch Null liegen. Damit ist aber $V = \{sv_1 + tv_2 : s, t \in \mathbb{R}\}$, also die Ebene durch Null, die durch v_1 und v_2 bestimmt wird.

4.8. Summen. Als nächstes betrachten wir zwei Teilräume W und W' eines $\mathbb{K}$–Vektorraumes V. Dann wissen wir aus Korollar 2.4, dass der Durchschnitt $W \cap W'$ ebenfalls ein Teilraum von V ist. Wir können aber aus W und W' noch einen weiteren Teilraum konstruieren:

Definition 4.8. Seien W und W' Teilräume eines $\mathbb{K}$–Vektorraumes V. Dann ist die *Summe* $W + W'$ der Teilräume definiert durch $W + W' := \langle W \cup W' \rangle \subset V$.

Wir können nun die Summe zweier Teilräume schön beschreiben, was auch die Notation erklärt, und ihre Dimension bestimmen:

Proposition 4.8. *Seien W und W' Teilräume eines $\mathbb{K}$–Vektorraumes V. Dann gilt:*

(1) Ein Element $v \in V$ liegt genau dann in $W + W'$ wenn es Elemente $w \in W$ und $w' \in W'$ gibt, sodass $v = w + w'$ gilt.

(2) Sind W und W' endlichdimensional, dann ist auch $W + W'$ endlichdimensional und

$$\dim(W + W') = \dim(W) + \dim(W') - \dim(W \cap W').$$

Beweis. (1) Sei $\tilde{W}$ die Menge aller $v \in V$, die in der angegebenen Form geschrieben werden können. Dann gilt natürlich $W \cup W' \subset \tilde{W} \subset W + W'$, also genügt es zu zeigen, dass $\tilde{W} \subset V$ ein Teilraum ist. Nun ist aber $r(w + w') = (rw) + (rw')$ und

$$(w_1 + w_1') + (w_2 + w_2') = (w_1 + w_2) + (w_1' + w_2')$$

und weil W und W' Teilräume sind, liegen die Summanden wieder im jeweiligen Raum.

(2) Da $W \cap W'$ ein Teilraum von W (und von W') ist, ist auch dieser Raum endlichdimensional. Wir können also eine Basis $\{w_1, \dots, w_k\}$ für $W \cap W'$ wählen, wobei $k = \dim(W \cap W')$. Da diese Menge eine linear unabhängige Teilmenge von W ist, finden wir nach Teil (2) von Korollar 4.5 Elemente $w_{k+1}, \dots, w_n \in W$, sodass $\{w_1, \dots, w_n\}$ eine Basis für W ist, und natürlich gilt dann $n = \dim(W)$. Analog finden wir Elemente $w'_{k+1}, \dots, w'_m \in W'$, sodass $\{w_1, \dots, w_k, w'_{k+1}, \dots, w'_m\}$ eine Basis für W' ist, und $m = \dim(W')$.

Wir behaupten nun, dass die Menge $\mathcal{B} := \{w_1, \dots, w_n, w'_{k+1}, \dots, w'_m\}$ eine Basis für den Teilraum $W + W'$ bildet, was den Beweis abschließt, da diese Menge $n + m - k$ Elemente hat. Zum Beweis bemerken wir zunächst, dass $\mathcal{B}$ eine Basis von W enthält, also gilt $W \subset \langle \mathcal{B} \rangle$. Analog folgt $W' \subset \langle \mathcal{B} \rangle$ und damit $W \cup W' \subset \langle \mathcal{B} \rangle$. Da einerseits $\mathcal{B} \subset W \cup W'$ gilt und andererseits $\langle \mathcal{B} \rangle$ ein Teilraum ist, schließen wir $\langle \mathcal{B} \rangle = W + W'$.

Es bleibt also zu zeigen, dass $\mathcal{B}$ linear unabhängig ist. Nehmen wir also an, dass wir Skalare $\lambda_1, \dots, \lambda_n, \mu_{k+1}, \dots, \mu_m \in \mathbb{K}$ finden, sodass

$$\lambda_1 w_1 + \cdots + \lambda_n w_n + \mu_{k+1} w'_{k+1} + \cdots + \mu_m w'_m = 0 \tag{4.1}$$

gilt. Dann erhalten wir

$$\mu_{k+1} w'_{k+1} + \cdots + \mu_m w'_m = -\lambda_1 w_1 + \cdots - \lambda_n w_n.$$

Nun liegen die w' in W' und die $w \in W$, also müssen beide Seiten der Gleichung in $W \cap W'$ liegen. Da $\{w_1, \dots, w_k\}$ eine Basis für diesen Raum bildet, finden wir Skalare $a_1, \dots, a_k$, sodass $\mu_{k+1}w'_{k+1} + \cdots + \mu_m w'_m = a_1w_1 + \cdots + a_kw_k$ gilt. Daraus folgt aber

$$a_1w_1 + \cdots + a_kw_k - \mu_{k+1}w'_{k+1} - \cdots - \mu_m w'_m = 0.$$

Nach Konstruktion bilden aber die Vektoren $\{w_1, \dots, w_k, w'_{k+1}, \dots, w'_m\}$ eine Basis für W', sind also insbesondere linear unabhängig. Damit muss aber $a_i = 0$ für $i = 1, \dots, k$ und $\mu_j = 0$ für $j = k+1, \dots, m$ gelten. Damit liefert (4.1) nun aber $\lambda_1w_1 + \cdots + \lambda_nw_n$ und weil die w_i eine Basis für W bilden folgt $\lambda_i = 0$ für alle $i = 1, \dots, n$. □

Beispiel 4.8. Wir betrachten Proposition 4.8 für Teilräume von $\mathbb{R}^2$ und $\mathbb{R}^3$, wo sie geometrisch sehr anschaulich ist. Im Fall von $\mathbb{R}^2$ sind nur eindimensionale Teilräume interessant, also Geraden durch 0. Sind W und W' solche Geraden, dann kann $W \cap W'$ nur Dimension 1 oder 0 haben, weil es ein Teilraum von W (und von W') ist. Ist $\dim(W \cap W') = 1$, dann muss $W \cap W' = W$ und $W \cap W' = W'$, also $W = W'$ gelten, und dann ist natürlich auch $W + W' = W$. Im interessanten Fall $W \neq W'$ gilt $\dim(W \cap W') = 0$, also $W \cap W' = \{0\}$. Dann folgt aber aus Proposition 4.8 sofort $\dim(W + W') = 2$, also $W + W' = \mathbb{R}^2$ nach Proposition 4.7.

Für $\mathbb{R}^3$ kommen zu den Geraden durch Null noch die Ebenen durch 0 als zweidimensionale Teilräume dazu. Ist W eine Ebene durch 0 und W' eine Gerade durch Null, dann kann $\dim(W \cap W')$ nur 0 oder 1 sein. Im zweiten Fall muss $W \cap W' = W'$, also $W' \subset W$ gelten. Der interessante Fall ist wieder $W \cap W' = \{0\}$. Hier zeigt Proposition 4.8, dass $\dim(W + W') = 3$, also $W + W' = \mathbb{R}^3$ gelten muss.

Sind schließlich $W, W' \subset \mathbb{R}^3$ zwei Ebenen durch den Nullpunkt, dann gilt $\dim(W + W') = 4 - \dim(W \cap W')$ nach Proposition 4.8, und offensichtlich ist $2 = \dim(W) \leq \dim(W + W') \leq 3$. Somit muss in diesem Fall $\dim(W \cap W')$ gleich 1 oder 2 sein. Ist $\dim(W \cap W') = 2$, dann ist wieder $W = W'$. Für $W \neq W'$ muss automatisch $\dim(W \cap W') = 1$ und $\dim(W + W') = 3$ gelten. Damit schneiden zwei verschiedene Ebenen durch Null in $\mathbb{R}^3$ immer in einer Geraden (durch den Nullpunkt) und man kann jedes Element $v \in \mathbb{R}^3$ als Summe von Elementen der beiden Ebenen schreiben.

4.9. Direkte Summen und Komplemente. Eine Verfeinerung der Ideen aus dem letzten Abschnitt führt zum Konzept der direkten Summe und zum Begriff komplementärer Teilräume. Intuitiv betrachtet, wollen wir einen Vektorraum V “in zwei Teile zerlegen”.

Definition 4.9. Sei V ein $\mathbb{K}$–Vektorraum und $W \subset V$ ein Teilraum. Ein Teilraum $W' \subset V$ heißt *komplementär zu* W oder ein *Komplement für* W, wenn $W + W' = V$ und $W \cap W' = \{0\}$ gilt.

In diesem Fall sagt man, V ist die *direkte Summe* der Teilräume W und W' und schreibt $V = W \oplus W'$.

Wir können nun Paare von komplementären Teilräumen eines Vektorraumes V schön charakterisieren:

Proposition 4.9. *Für zwei Teilräume W und W' eines $\mathbb{K}$–Vektorraumes V sind folgende Bedingungen äquivalent:*

(1) $V = W \oplus W'$
(2) $\dim(V) = \dim(W) + \dim(W')$ *und* $W \cap W' = \{0\}$.
(3) *Jedes Element $v \in V$ lässt sich eindeutig in der Form $v = w + w'$ für $w \in W$ und $w' \in W'$ schreiben.*

Beweis. (1)⇒(3) Nach Definition gilt $V = W + W'$ und $W \cap W' = \{0\}$. Wegen der ersten Eigenschaft lässt sich nach Proposition 4.8 jedes Element $v \in V$ in der Form $v = w + w'$ schreiben. Ist $w_1 + w_1' = w_2 + w_2'$ für Element $w_i \in W$ und $w_i' \in W'$, dann gilt natürlich $w_1 - w_2 = w_2' - w_1'$. Nun liegt aber die linke Seite der Gleichung in W und die rechte in W', also beide Seiten in $W \cap W' = \{0\}$. Damit ist $w_1 - w_2 = 0$, also $w_1 = w_2$ und analog ist $w_1' = w_2'$ und die Eindeutigkeit der Darstellung folgt.

(3)⇒(2) Nach Proposition 4.8 folgt aus der Existenz der Darstellung, dass $V = W + W'$ gilt. Ist $v \in W \cap W'$, dann sind $v = v + 0$ und $v = 0 + v$ zwei Darstellungen von v als Summe eines Elements von W und eines Elements von W', also folgt $v = 0$ aus der Eindeutigkeit der Darstellung. Damit ist $W \cap W' = \{0\}$, also auch $\dim(W \cap W') = 0$ und nach Proposition 4.8 folgt

$$\dim(V) = \dim(W + W') = \dim(W) + \dim(W').$$

(2)⇒(1) Wegen $W \cap W' = \{0\}$ folgt aus Proposition 4.8 sofort, dass

$$\dim(W + W') = \dim(W) + \dim(W') = \dim(V)$$

gilt. Damit ist aber $W + W' = V$ nach Proposition 4.7. □

Beispiel 4.9. Meist ist die Bedingung In (2) am leichtesten zu verifizieren. Manchmal kann man aber auch eine Kombination der Bedingungen verifizieren. Betrachten wir zum Beispiel den Teilraum $W \subset \mathbb{R}^n$, der von dem Vektor $(1, \dots, 1)$ erzeugt wird, d.h. $W = \{(t, t, \dots, t) : t \in \mathbb{R}\}$. Wir behaupten, dass der Teilraum $W' := \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 + \dots + x_n = 0\}$ komplementär zu W ist.

Summiert man die Komponenten von $(t, \dots, t)$, dann erhält man natürlich nt, also liegt dieser Vektor nur für $t = 0$ in W'. Damit gilt aber $W \cap W' = \{0\}$. Um unsere Behauptung zu beweisen genügt es also, zu beweisen, dass man einen beliebigen Vektor $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ als Summe eines Elements von W und eines Elements von W' schreiben kann. Sei dazu $a := \frac{1}{n}(x_1 + \dots + x_n)$. Dann gilt natürlich

$$(x_1, \dots, x_n) = (a, a, \dots, a) + (x_1 - a, \dots, x_n - a),$$

also genügt es zu zeigen, dass $(x_1 - a, \dots, x_n - a) \in W'$ liegt. Addiert man aber die Komponenten dieses Vektors auf, dann erhält man $x_1 + \dots + x_n - na = 0$.

Bemerkung 4.9. Es gibt in der linearen Algebra noch ein zweites Konzept von direkten Summen. Seien V_1 und V_2 Vektorräume über dem gleichen Körper $\mathbb{K}$. Dann kann man einfach das kartesische Produkt $V_1 \times V_2$, also die Menge aller geordneten Paare (v_1, v_2) mit $v_i \in V_i$ betrachten. Darauf definiert man eine Addition komponentenweise durch $(v_1, v_2) + (v_1', v_2') := (v_1 + v_1', v_2 + v_2')$. Man verifiziert aus den entsprechenden Bedingungen für V_1 und V_2 sofort, dass diese Addition die Bedingungen (V1)–(V4) erfüllt, wobei $0 = (0, 0)$ und $-(v_1, v_2) = (-v_1, -v_2)$ gilt. Weiters definiert man eine Skalarmultiplikation auf $V_1 \times V_2$ ebenfalls komponentenweise, also durch $r(v_1, v_2) := (rv_1, rv_2)$. Aus den entsprechenden Eigenschaften von V_1 und V_2 verifiziert man nun leicht, dass auch (V5)–(V8) erfüllt sind (siehe Übungen). Damit ist aber $V_1 \times V_2$ ebenfalls ein $\mathbb{K}$–Vektorraum, der meist mit $V_1 \oplus V_2$ bezeichnet wird.

Diese Bezeichnung ist auch sehr natürlich: Aus der Definition ist offensichtlich, dass die Menge $\{(v_1, 0) : v_1 \in V_1\}$ einen Teilraum von $V_1 \times V_2$ bildet, der isomorph zu V_1 ist, und analog bilden die Elemente mit erster Komponente 0 einen Teilraum, der isomorph zu V_2 ist. Aus Bedingung (2) von Proposition 4.9 ist offensichtlich, dass $V_1 \times V_2$ die direkte Summe dieser beiden Teilräume ist.

Gelegentlich wird für das hier verwendete Konzept der Begriff "äußere direkte Summe" und für das Konzept von vorher der Begriff "innere direkte Summe" verwendet, das scheint mir aber mehr in die Unterscheidung hieneinzugeheimnissen als eigentlich drinnen ist.

4.10. Basen und lineare Abbildungen. Bevor wir den fundamentalen Dimensionssatz für lineare Abbildungen beweisen können, der auch wichtige Anwendungen auf lineare Gleichungssysteme hat, müssen wir noch kurz den Zusammenhang von Basen mit linearen Abbildungen weiter studieren. Das wird uns auch sofort zeigen, dass ein endlichdimensionaler Vektorraum durch seine Dimension bis auf Isomorphie eindeutig bestimmt ist.

Satz 4.10. *Seien V und W endlichdimensionale $\mathbb{K}$–Vektorräume.*

(1) Ist $\mathcal{B}$ eine Basis von V, dann kann jede Funktion $\varphi : \mathcal{B} \to W$ eindeutig zu einer linearen Abbildung $f : V \to W$ ausgedehnt werden.

(2) Eine lineare Abbildung $f : V \to W$ ist genau dann surjektiv, wenn für ein (oder äquivalent für jedes) Erzeugendensystem A von V, das Bild $f(A) \subset W$ ein Erzeugendensystem für W ist.

(3) Eine lineare Abbildung $f : V \to W$ ist genau dann injektiv, wenn für beliebige linear unabhängige Vektoren $v_1, \dots, v_k \in V$ auch die Vektoren $f(v_1), \dots, f(v_k)$ linear unabhängig sind. Das ist auch äquivalent dazu, dass die Bilder der Vektoren aus einer Basis $\mathcal{B}$ von V linear unabhängig sind.

(4) Eine lineare Abbildung $f : V \to W$ ist genau dann ein linearer Isomorphismus wenn die Bilder der Vektoren einer (oder äquivalent jeder) Basis $\mathcal{B}$ für V eine Basis für W bilden.

Beweis. (1) Da $\mathcal{B} \subset V$ linear unabhängig ist, und $\langle \mathcal{B} \rangle = V$ gilt, folgt das sofort aus Satz 4.3.

(2) Aus Proposition 4.2 wissen wir, dass für jedes Erzeugendensystem $A \subset V$ das Bilde $f(A)$ ein Erzeugendensystem für $\operatorname{Im}(f)$ ist. Damit sind aber die Bedingungen in (2) offensichtlich äquivalent zu $\operatorname{Im}(f) = W$, also zur Surjektivität von f.

(3) Sei $f : V \to W$ eine injektive lineare Abbildung und seien $v_1, \dots, v_k \in V$ linear unabhängige Vektoren. Sind $r_i \in \mathbb{K}$ Skalare, sodass

$$r_1 f(v_1) + \dots + r_k f(v_k) = 0$$

ist, dann kann man die linke Seite als $f(r_1 v_1 + \dots + r_k v_k)$ schreiben, und wegen der Injektivität von F folgt $r_1 v_1 + \dots + r_k v_k = 0$. Da aber die v_i linear unabhängig sind, folgt daraus $r_1 = \dots = r_k = 0$. Damit sind aber auch die Vektoren $f(v_1), \dots, f(v_k)$ linear unabhängig.

Nehmen wir umgekehrt an, dass für eine Basis $\{v_1, \dots, v_n\}$ von V die Vektoren $f(v_1), \dots, f(v_n)$ linear unabhängig sind und dass $f(v) = 0$ für ein Element $v \in V$ gilt. Dann finden wir Skalare $r_i \in \mathbb{K}$, sodass $v = r_1 v_1 + \dots + r_n v_n$ gilt. Damit rechnen wir

$$0 = f(v) = f(r_1 v_1 + \dots + r_n v_n) = r_1 f(v_1) + \dots + r_n f(v_n).$$

Da die $f(v_i)$ linear unabhängig sind, folgt $r_1 = \dots = r_n = 0$, also $v = 0$. Damit ist aber $\operatorname{Ker}(f) = \{0\}$, also f injektiv.

(4) Das folgt nun sofort aus Teil (2) und (3), weil nach Satz 2.5 ein linearer Isomorphismus einfach nur eine bijektive lineare Abbildung ist. □

Korollar 4.10. *Seien V und W $\mathbb{K}$–Vektorräume mit* $\dim(V) = n$ *und* $\dim(W) = m$. *Dann gilt:*

(1) Es gibt genau dann eine injektive lineare Abbildung $f : V \to W$, wenn $n \leq m$ gilt.

(2) Es gibt genau dann eine surjektive lineare Abbildung $f : V \to W$, wen $n \geq m$ gilt.

(3) Es gibt genau dann einen linearen Isomorphismus $f : V \to W$, wenn $n = m$ gilt. Insbesondere ist jeder n–dimensionale $\mathbb{K}$–Vektorraum isomorph zu $\mathbb{K}^n$.

Beweis. Sei $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis für V. Ist $f : V \to W$ eine surjektive lineare Abbildung, dann ist nach Teil (2) des Satzes $\{f(v_1), \dots, f(v_n)\}$ ein Erzeugendensystem für W, also muss $\dim(W) \leq n$ gelten Analog sind für injektives f die Vektoren $f(v_1), \dots, f(v_n)$ linear unabhängig, also muss $\dim(W) \geq n$ gelten. Damit folgt die Notwendigkeit der Dimensionsbedingungen in allen drei Teilen.

Um zu zeigen, dass diese Bedingungen auch hinreichend sind, wählen wir eine Basis $\{w_1, \dots, w_m\}$ für W. Ist $n \leq m$, dann wissen wir aus Teil (1) des Satzes, dass es eine lineare Abbildung $f : V \to W$ gibt, die $f(v_i) = w_i$ für alle $i = 1, \dots, n$ erfüllt. Da die Bildvektoren als Elemente einer Basis linear unabhängig sind, ist f nach Teil (3) des Satzes injektiv. Im Fall $n = m$ bilden diese Elemente sogar eine Basis für W, also ist f ein linearer Isomorphismus nach Teil (4) des Satzes. Ist schließlich $n > m$, dann betrachten wir die lineare Abbildung $f : V \to W$, die $f(v_i) = w_i$ für $i = 1, \dots, m$ und $f(v_i) = 0$ für $m < i \leq n$ erfüllt. Diese ist surjektiv nach Teil (2) des Satzes. □

Teil (3) zeigt nicht nur, dass jeder n–dimensionale $\mathbb{K}$–Vektorraum isomorph zu $\mathbb{K}^n$ ist, aus dem Beweis sehen wir auch, dass die Wahl einer Basis für V eine Isomorphismus $f : V \to \mathbb{K}^n$ liefert, der die Elemente dieser Basis auf die Standardbasis abbildet. Man kann also V mit $\mathbb{K}^n$ identifizieren, aber diese Identifikation hängt von der Wahl einer Basis ab.

4.11. Der Dimensionssatz für lineare Abbildungen. Nach diesen Vorbereitungen können wir nun den fundamentalen Dimensionssatz über lineare Abbildungen beweisen:

Satz 4.11. *Seien V und W endlichdimensionale $\mathbb{K}$–Vektorräume und sei $f : V \to W$ eine lineare Abbildung. Dann ist* $\dim(\operatorname{Im}(f)) = \dim(V) - \dim(\operatorname{Ker}(f))$.

Beweis. Wähle eine Basis $\{v_1, \dots, v_k\}$ für den Teilraum $\operatorname{Ker}(f) \subset V$. Da die Vektoren v_i linear unabhängig sind, können sie nach Teil (2) von Korollar 4.5 zu einer Basis $\{v_1, \dots, v_n\}$ für den Vektorraum V erweitert werden. Wir behaupten, dass die Vektoren $f(v_{k+1}), \dots, f(v_n) \in W$ eine Basis für den Teilraum $\operatorname{Im}(f) \subset W$ bilden, was natürlich den Satz beweist.

Zunächst wissen wir aus Satz 4.2, dass die Vektoren $f(v_1), \dots, f(v_n)$ ein Erzeugendensystem für $\operatorname{Im}(f)$ bilden. Nach Konstruktion gilt aber $f(v_1) = \dots = f(v_k) = 0$, also bilden auch $f(v_{k+1}), \dots, f(v_n)$ ein Erzeugendensystem. Nehmen wir andererseits an, das $r_{k+1}, \dots, r_n \in \mathbb{K}$ Skalare sind, sodass

$$r_{k+1} f(v_{k+1}) + \dots + r_n f(v_n) = 0$$

gilt. Dann kann man die linke Seite als $f(r_{k+1} v_{k+1} + \dots + r_n v_n)$ schreiben, also gilt $r_{k+1} v_{k+1} + \dots + r_n v_n \in \operatorname{Ker}(f)$. Nach Konstruktion bilden aber die Vektoren $v_1, \dots, v_k$ eine Basis für diesen Teilraum, also finden wir Skalare $s_1, \dots, s_k \in \mathbb{K}$, sodass

$$r_{k+1} v_{k+1} + \dots + r_n v_n = s_1 v_1 + \dots + s_k v_k.$$

Das bedeutet aber gerade, dass

$$s_1 v_1 + \dots + s_k v_k - r_{k+1} v_{k+1} - \dots - r_n v_n = 0$$

und wegen der linearen Unabhängigkeit von $v_1, \dots, v_n$ ist das nur für $s_1 = \dots = s_k = r_{k+1} = \dots = r_n = 0$ möglich. □

Beispiel 4.11. Mit diesem Satz können wir sofort sehen, dass der Teilraum $W' = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 + \dots + x_n = 0\} \subset \mathbb{R}^n$ Dimension $n-1$ hat. Er ist nämlich der Kern der linearen Abbildung $f : \mathbb{R}^n \to \mathbb{R}$, die gegeben ist durch $f(x_1, \dots, x_n) := x_1 + \dots + x_n$. Offensichtlich ist f surjektiv, also $\dim(\operatorname{Im}(f)) = 1$, und damit folgt $\dim(\operatorname{Ker}(f)) = n-1$ sofort aus dem Satz.

4.12. Eine Anwendung auf Polynome. Wir kehren zu den Polynomen zurück, die wir schon kurz einmal in 2.6 betrachtet haben. Dort haben wir Polynome über einem Körper $\mathbb{K}$ als formale Ausdrücke der Form $\sum_{i=0}^N a_i x^i$ kennen gelernt. Wir haben auch gesehen, dass man jedem solchen Polynom eine Funktion $\mathbb{K} \to \mathbb{K}$ zuordnen kann, indem man für x Elemente von $\mathbb{K}$ einsetzt. Wir werden ab sofort die übliche Vereinfachung der Notation benutzen und für ein Polynom $p \in \mathbb{K}[x]$ auch die entsprechende Funktion mit $t \mapsto p(t)$ bezeichnen. Als nächstes betrachten wir die (in Spezialfällen ebenfalls bereits bekannte) Multiplikation von Polynomen, die in beiden Bildern Sinn macht (und gleich aussieht).

Definition 4.12. Für Polynome $p = \sum a_i x^i \in \mathbb{K}_n[x]$ und $q = \sum b_j x^j \in \mathbb{K}_m[x]$ definiert man das *Produkt* $pq = \sum c_k x^k \in \mathbb{K}_{n+m}[x]$ durch $c_k := \sum_{i+j=k} a_i b_j$.

Das ist einfach die übliche Multiplikation, die man erhält indem man $x^i x^j = x^{i+j}$ setzt, und das dann distributiv ausdehnt.

Proposition 4.12. *Die Multiplikation von Polynomen ist assoziativ und kommutativ und distributiv bezüglich der Addition von Polynomen aus 2.6. Das konstante Polynom* 1 *ist ein neutrales Element für die Multiplikation.*

Unter der Funktion $\varphi : \mathbb{K}[x] \to \mathcal{F}(\mathbb{K}, \mathbb{K})$ *aus 2.6 geht das Produkt von Polynomen in das punktweise Produkt von Funktionen über, d.h.* $\varphi(pq)(t) = \varphi(p)(t) \cdot \varphi(q)(t)$ *für* $p, q \in \mathbb{K}[x]$ *und* $t \in \mathbb{K}$, *bzw.* $(pq)(t) = p(t)q(t)$ *in der vereinfachten Notation.*

Beweis. Die Kommutativität und die Tatsache, dass das konstante Polynom 1 ein neutrales Element ist, folgen direkt aus der Definition. Für die Distributivität betrachte $p = \sum a_i x^i$, $\tilde{p} = \sum \tilde{a}_i x^i$ und $q = \sum b_j x^j$. Nach Definition ist $p + \tilde{p} = \sum (a_i + \tilde{a}_i) x^i$. Damit ist der Koeffizient von x^k in $(p + \tilde{p})q$ gegeben durch

$$\textstyle\sum_{i+j=k} (a_i + \tilde{a}_i) b_j = \sum_{i+j=k} a_i b_j + \sum_{i+j=k} \tilde{a}_i b_j,$$

also genau die Summe der entsprechenden Koeffizienten von pq und $\tilde{p}q$. Damit gilt aber $(p + \tilde{p})q = pq + \tilde{p}q$. Die Assoziativität verifiziert man durch eine etwas langwierige aber einfache direkte Rechnung (siehe Übungen).

Die Abbildung φ war ja definiert durch Einsetzen von Elementen von $\mathbb{K}$ für x. Ist also $p = \sum_{i=1}^n a_i x^i$, dann ist ist $p(t) = a_0 + a_1 t + \dots + a_n t^n$. Nun ist natürlich $t^i t^j = t^{i+j}$ für alle $t \in \mathbb{K}$. Für ein weiteres Polynom $q = \sum_{j=1}^m b_j x^j$ ist dann $q(t) = b_0 + b_1 t + \dots + b_m t^m$. Nachdem Distributivgesetz in $\mathbb{K}$ ergibt sich

$$\textstyle p(t) \cdot q(t) = \sum_{i,j} a_i b_j t^{i+j} = \sum_k (\sum_{i+j=k} a_i b_j) t^k = pq(t).$$

□

Wir können nun den Dimensionssatz für lineare Abbildungen benutzen, um elegant zu beweisen, dass man Polynome p konstruieren kann, für die die zugehörige Funktion $\mathbb{K} \to \mathbb{K}$ vorgegebene Eigenschaften hat. Insbesondere interessiert man sich für *Nullstellen* von p, d.h. Elemente $t \in \mathbb{K}$, sodass $p(t) = 0$ gilt. Der folgende Satz liefert eine fundamentale Möglichkeit für die Approximation von beliebigen Funktionen durch Polynomfunktionen:

Satz 4.12. *Sei $\mathbb{K}$ ein beliebiger Körper, und seien $r_0, \dots, r_n \in \mathbb{K}$ verschiedene Elemente. Dann gibt es zu beliebigen Elementen $s_0, \dots s_n \in \mathbb{K}$ ein eindeutiges Polynom $p \in \mathbb{K}_n[x]$ für das $p(r_i) = s_i$ für alle $i = 0, \dots, n$ gilt.*

Insbesondere hat ein Polynom $p \neq 0$ vom Grad $\leq n$ höchstens n verschiedene Nullstellen in $\mathbb{K}$.

Beweis. Für $i = 0, \dots, n$ betrachte die Funktion $f_i : \mathbb{K}_n[x] \to \mathbb{K}$, die definiert ist durch $f_i(p) := p(r_i)$. Man ordnet also jedem Polynom p den Wert der entsprechenden Funktion $\mathbb{K} \to \mathbb{K}$ im Punkt $r_i \in \mathbb{K}$ zu. Aus Proposition 2.6 wissen wir, dass $\varphi : \mathbb{K}[x] \to \mathcal{F}(\mathbb{K}, \mathbb{K})$ linear ist, also $\varphi(p + q) = \varphi(p) + \varphi(q)$ und $\varphi(rp) = r\varphi(p)$ erfüllt. Da Addition und Multiplikation mit Skalaren auf $\mathcal{F}(\mathbb{K}, \mathbb{K})$ punktweise definiert sind, folgt $f_i(p+q) = f_i(p) + f_i(q)$ und $f_i(rp) = rf_i(p)$, also ist jede der Funktionen $f_i : \mathbb{K}_n[x] \to \mathbb{K}$ linear. Damit ist aber auch die Funktion $f : \mathbb{K}_n[x] \to \mathbb{K}^{n+1}$, die gegeben ist durch $f(p) := (f_0(p), \dots, f_n(p))$ linear, siehe 2.5 und 3.2.

Wir müssen zeigen, dass f bijektiv ist. Surjektivität bedeutet nämlich gerade, dass es für jedes Element $(s_0, \dots, s_n) \in \mathbb{K}^{n+1}$ ein Polynom $p \in \mathbb{K}_n[x]$ gibt, das $f(p) = (s_0, \dots, s_n)$, also $p(r_i) = s_i$ für alle $i = 0, \dots, n$ erfüllt. Aus der Injektivität folgt dann, dass dieses Polynom eindeutig bestimmt ist. Insbesondere ist das konstante Polynom 0 dann das einzige Element in $\mathbb{K}_n[x]$, das $p(r_i) = 0$ für alle $i = 0, \dots, n$ erfüllt. Der Dimensionssatz 4.11 kommt dadurch in Spiel, dass wir aus Proposition 2.6 auch schon wissen, dass $\mathbb{K}_n[x]$ isomorph zu $\mathbb{K}^{n+1}$ ist, also die beiden Räume nach Korollar 4.10 die gleiche Dimension haben. Damit genügt es aber zu zeigen, dass f surjektiv ist, denn damit folgt aus $\dim(\mathrm{Im}(f)) = \dim(\mathbb{K}_n[x])$ schon $\dim(\mathrm{Ker}(f)) = 0$.

Die Surjektivität lässt sich aber nun elegant beweisen. Nach Voraussetzung sind die r_i alle verschieden. Damit sind aber die Elemente $r_j - r_i \in \mathbb{K}$ für $j \neq i$ alle $\neq 0$, also ist auch $\lambda_i := \prod_{j \neq i}(r_j - r_i) \neq 0$. Jetzt betrachten wir für $i = 0, \dots, n$ das Polynom

$$q_i := (\lambda_i)^{-1}(x - r_0) \dots (x - r_{i-1})(x - r_{i+1}) \dots (x - r_n)$$

Nachdem man hier eine Zahl und n Polynome vom Grad 1 aufmultipliziert ist $q_i \in \mathbb{K}_n[x]$. Außerdem ist

$$q_i(r_j) = (\lambda_i)^{-1}(r_j - r_0) \dots (r_j - r_{i-1})(r_j - r_{i+1}) \dots (r_j - r_n).$$

Für $j \neq i$ kommt der Faktor $(r_j - r_j) = 0$ vor, also ist $q_i(r_j) = 0$ für $j \neq i$. Für $j = i$ ergibt sich $(\lambda_i)^{-1}\lambda_i = 1$, also $q_i(r_i) = 1$. Das bedeutet aber gerade, dass $f(q_i) = e_{i+1}$ gilt, also $q_0, \dots, q_n$ auf die Standardbasis $e_1, \dots, e_{n+1}$ von $\mathbb{K}^{n+1}$ abgebildet wird. Damit ist f aber surjektiv nach Teil (2) von Satz 4.10. □

Aus dem Beweis folgt natürlich auch, dass die Polynome $q_0, \dots, q_n$ eine Basis für $\mathbb{K}_n[x]$ bilden, und dass das Polynom $q := s_0q_0 + \dots + s_nq_n$ gerade $q(r_i) = s_i$ für $i = 0, \dots, n$ erfüllt.

4.13. Der Rang. Der Dimensionssatz 4.11 zeigt, dass es genügt entweder die Dimension des Kerns oder die Dimension des Bildes einer linearen Abbildung zu kennen. Es stellt sich als vorteilhaft heraus, die Dimension des Bildes als grundlegende Größe zu verwenden.

Definition 4.13. (1) Der *Rang* $\mathrm{rg}(f)$ einer linearen Abbildung $f : V \to W$ zwischen zwei endlichdimensionalen $\mathbb{K}$–Vektorräumen ist die Dimension des Teilraumes $\mathrm{Im}(f) \subset W$.

(2) Der Rang $\mathrm{rg}(A)$ einer Matrix $A \in M_{n,m}(\mathbb{K})$ ist der Rang der linearen Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$ die gegeben ist durch $f(v) := Av$.

Wir können nun sowohl die wichtigsten Eigenschaften des Ranges von Matrizen sofort ablesen. Das zeigt auch, dass wir mit den Methoden aus Kapitel 3 den Rang beliebiger Matrizen algorithmisch bestimmen können.

Proposition 4.13. *Sei $A \in M_{m,n}(\mathbb{K})$ eine Matrix.*

(1) Der Rang von A ist genau die Dimension des von den Spaltenvektoren von A aufgespannten Teilraumes von $\mathbb{K}^m$. Insbesondere ist $\operatorname{rg}(A) = r$ genau dann, wenn es r Spaltenvektoren von A gibt, die linear unabhängig in $\mathbb{K}^m$ sind, aber je $r+1$ Spaltenvektoren von A linear abhängig sind.

(2) Sind $C \in M_m(\mathbb{K})$ und $D \in M_n(\mathbb{K})$ invertierbar, dann ist $\operatorname{rg}(A) = \operatorname{rg}(CA) = \operatorname{rg}(AD)$.

(3) Hat die Matrix A Zeilenstufenform mit Indizes r, $j_1, \dots, j_r$, dann gilt $\operatorname{rg}(A) = r$.

Beweis. (1) Ist $f : \mathbb{K}^n \to \mathbb{K}^m$ die lineare Abbildung $f(v) = Av$, dann sind die Spaltenvektoren von A gerade die Vektoren $f(e_1), \dots, f(e_n)$, wobei die e_i die Elemente der Standardbasis von $\mathbb{K}^n$ sind (siehe 3.2). Nach Satz 4.2 bilden diese Vektoren eine Erzeugendensystem für $\operatorname{Im}(f)$, also folgt die erste Behauptung sofort. Nach Korollar 4.5 enthält dieses Erzeugendensystem eine Basis für diesen Teilraum, also folgt auch die zweite Behauptung.

(2) Sei wieder $f : \mathbb{K}^n \to \mathbb{K}^m$ die lineare Abbildung $f(v) = Av$, und $g : \mathbb{K}^m \to \mathbb{K}^m$ gegeben durch $g(v) = Cv$. Dann ist g nach Voraussetzung ein linearer Isomorphismus. Natürlich bildet g den Teilraum $\operatorname{Im}(f) \subset \mathbb{K}^m$ auf $\operatorname{Im}(g \circ f) \subset \mathbb{K}^m$ ab, und $g : \operatorname{Im}(f) \to \operatorname{Im}(g \circ f)$ ist ebenfalls bijektiv. Damit ist aber $\operatorname{rg}(CA) = \dim(\operatorname{Im}(g \circ f)) = \dim(\operatorname{Im}(f)) = \operatorname{rg}(A)$. Analog verifiziert man $\operatorname{rg}(AD) = \operatorname{rg}(A)$ (siehe Übungen).

(3) Wir können A durch elementare Zeilenoperationen auf reduzierte Zeilenstufenform mit den gleichen Indizes bringen. Nach 3.6 entspricht das einer Multiplikation von links mit einer invertierbaren Matrix und ändert damit nach Teil (2) den Rang nicht. Also können wir ohne Beschränkung der Allgemeinheit annehmen, dass A reduzierte Zeilenstufenform hat. Dann ist aber für $i = 1, \dots, r$ der j_i–te Spaltenvektor von A gerade e_i, also hat A mindestens r linear unabhängige Spaltenvektoren und $\operatorname{rg}(A) \geq r$. Andererseits sind nach Definition ab der $(r+1)$–ten Zeile alle Zeilen von A gleich Null. Das bedeutet aber, dass das Bild der Abbildung $v \mapsto Av$ im Teilraum $\mathbb{K}^r \subset \mathbb{K}^m$ liegt und damit Dimension $\leq r$ haben muss. □

Dieser Satz sagt uns auch, wie man den Rang einer Matrix effektiv bestimmen kann: Man bringt A durch elementare Zeilenoperationen auf Zeilenstufenform. Das entspricht einer Multiplikation von links mit einer invertierbaren Matrix, ändert also nach Teil (2) den Rang nicht. Für die resultierende Matrix kann man aber den Rang sofort aus Teil (3) des Satzes ablesen.

4.14. Anwendung auf lineare Gleichungssysteme. Mit Hilfe des Ranges können wir nun das Lösungsverhalten von linearen Gleichungssystemen ziemlich genau beschreiben. Zunächst können wir mit Hilfe des Ranges ein einfaches Kriterium für die Lösbarkeit eines linearen Gleichungssystems angeben.

Proposition 4.14. *Sei $A \in M_{m,n}(\mathbb{K})$ eine Matrix und $b \in \mathbb{K}^m$ ein Vektor. Dann hat das lineare Gleichungssystem $Ax = b$ genau dann eine Lösung, wenn die Koeffizientenmatrix A und die erweiterte Koeffizientenmatrix $(A|b)$ den gleichen Rang haben.*

Beweis. Seien $a_1, \dots, a_n \in \mathbb{K}^m$ die Spaltenvektoren von A, und sei V der von ihnen aufgespannte Teilraum von $\mathbb{K}^m$. Da die Spaltenvektoren von $(A|b)$ gerade die a_i und b sind, enthält $\langle a_1, \dots, a_n, b\rangle$ natürlich V. Nun ist $\operatorname{rg}(A) = \dim(V)$, also $\operatorname{rg}(A) = rg((A|b))$

äquivalent zu $\langle a_1, \dots, a_n, b\rangle = V$. Das ist aber offensichtlich äquivalent zu $b \in V$, also zur Lösbarkeit von $Ax = b$, siehe Proposition 3.5. □

Diese Charakterisierung ist allerdings von eher theoretischer Bedeutung, weil man zur Bestimmung der Ränge ohnehin den Gaußschen Algorithmus benötigt. Wichtiger sind die folgenden allgemeinen Aussagen über das Lösungsverhalten, die wir nun treffen können:

Satz 4.14. *Sei $Ax = b$ ein System von m linearen Gleichungen in n–Variablen und sei r der Rang der Matrix $A \in M_{m,n}(\mathbb{K})$.*

(1) Das System besitzt genau dann eine Lösung für jedes $b \in \mathbb{K}^m$, wenn $r = m$ gilt. Insbesondere kann das nur für $n \geq m$ erfüllt sein.

(2) Existierende Lösungen des Systems sind genau dann eindeutig bestimmt, wenn $r = n$ gilt. Insbesondere kann das nur für $n \leq m$ erfüllt sein.

(3) Im Fall $n = m$ (also so viele Gleichungen wie Unbekannte) sind äquivalent:

(i) $r = n$
(ii) *Das System $Ax = b$ ist für jedes $b \in \mathbb{K}^n$ lösbar.*
(iii) *Es gibt ein $b \in \mathbb{K}^n$ für das die Lösung von $Ax = b$ eindeutig bestimmt ist.*
(iv) *Die Lösung von $Ax = b$ ist für jedes $b \in \mathbb{K}^n$ eindeutig bestimmt.*
(v) *Die $n \times n$–Matrix A ist invertierbar.*

Beweis. Sei $f : \mathbb{K}^n \to \mathbb{K}^m$ die lineare Abbildung $f(v) = Av$. Dann ist $r = \dim(\mathrm{Im}(f))$. Da $\mathrm{Im}(f) \subset \mathbb{K}^m$ gilt, ist $r \leq m$ und nach Satz 4.11 ist $\dim(\mathrm{Ker}(f)) = n - r$, also $r \leq n$. Nun ist $r = m$ natürlich äquivalent zu $\mathrm{Im}(f) = \mathbb{K}^m$ (siehe Proposition 4.7) und damit folgt (1) sofort. Analog ist $r = n$ äquivalent zu $\mathrm{Ker}(f) = \{0\}$ und damit folgt (2) sofort aus Proposition 3.5.

Aus diesen Überlegungen folgen die Implikationen (i)⇔(ii) und (iii)⇒(i)⇒(iv) von Teil (3). Nachdem (iv)⇒(iii) offensichtlich gilt, sind die ersten vier Bedingungen äquivalent. Nach Satz 4.11 ist (i) äquivalent zur Bijektivität von f also zu (v). □

Matrizenkalkül für allgemeine Vektorräume

Wir können nun die Überlegungen über Basen dazu verwenden, lineare Abbildungen zwischen beliebigen endlichdimensionalen Vektorräumen durch Matrizen zu beschreiben. Die Beschreibung hängt von der Wahl von Basen ab, diese Abhängigkeit ist aber gut kontrollierbar.

4.15. Koordinatenvektoren und Matrizen. Die Idee für Koordinatenvektoren ist einfach. Sei V ein $\mathbb{K}$–Vektorraum und sei $n = \dim(V)$. Sei $\mathcal{B}$ eine geordnete Basis für V, die aus den Vektoren $v_1, \dots, v_n$ besteht. Dann können wir jeden Vektor $v \in V$ eindeutig als $v = \sum_{i=1}^n r_i v_i$ für $r_1, \dots, r_n \in \mathbb{K}$ schreiben. Damit ist v natürlich eindeutig durch $[v]_{\mathcal{B}} := (r_1, \dots, r_n) \in \mathbb{K}^n$ bestimmt, und man nennt $[v]_{\mathcal{B}}$ den *Koordinatenvektor von v bezüglich der Basis $\mathcal{B}$.*

Sei nun V' ein weiterer $\mathbb{K}$–Vektorraum mit $\dim(V') = m$ und $\mathcal{B}'$ eine geordnete Basis für V', die aus den Vektoren $v'_1, \dots, v'_m$ besteht. Dann können wir auch hier die Koordinatenvektoren $[v']_{\mathcal{B}'}$ für Elemente $v' \in V'$ betrachten. Wenn wir eine lineare Abblidung $f : V \to V'$ gegeben haben, dann können wir diese explizit beschrieben, indem wir erklären, wie man für ein Element $v \in V$ den Koordinatenvektor $[f(v)]_{\mathcal{B}'}$ aus dem Koordinatenvektor $[v]_{\mathcal{B}}$ berechenen kann. Wie wir gleich beweisen werden, ist dieser Übergang durch Multiplikation mit einer $m \times n$–Matrix gegeben. Diese wird mit $[f]^{\mathcal{B}}_{\mathcal{B}'}$ bezeichnet und heißt die *Matrixdarstellung von f bezüglich der Basen $\mathcal{B}$ und $\mathcal{B}'$.*

Mit den Werkzeugen, die wir bereits entwickelt haben, können wir diese Aussagen ziemlich effizient beweisen, zugleich erhalten wir auch explizite Beschreibungen der vorkommenden Objekte. Für V und $\mathcal{B}$ wie oben, gibt es nach Teil (1) von Satz 4.10 eine eindeutige lineare Abbildung $\varphi = \varphi_{\mathcal{B}} : V \to \mathbb{K}^n$, die $\varphi(v_i) = e_i$ für alle $i = 1, \dots, n$ erfüllt und nach Teil (5) dieses Satzes ist φ ein linearer Isomorphismus.

Proposition 4.15. *(1) Sei V ein $\mathbb{K}$–Vektorraum, $\dim(V) = n$ und $\mathcal{B}$ eine goordnete Basis für V. Dann ist der eindeutig bestimmte lineare Isomorphismus $\varphi = \varphi_{\mathcal{B}} : V \to \mathbb{K}^n$, der $\mathcal{B}$ auf die Standardbasis abbildet gegeben durch $\varphi(v) = [v]_{\mathcal{B}}$.*

(2) Sei V' ein weiterer $\mathbb{K}$–Vektorraum, $m = \dim(V')$, $\mathcal{B}'$ eine geordnete Basis für V' und $\varphi' : \mathbb{V} \to \mathbb{K}^m$ der entsprechende lineare Isomorphismus. Sei $f : V \to V'$ eine lineare Abbildung und sei $[f]^{\mathcal{B}}_{\mathcal{B}'} \in M_{m,n}(\mathbb{K})$ die Matrix, die der linearen Abbildung $\varphi' \circ f \circ \varphi^{-1} : \mathbb{K}^n \to \mathbb{K}^m$ entspricht. Dann gilt

$$[f(v)]_{\mathcal{B}'} = [f]^{\mathcal{B}}_{\mathcal{B}'} \cdot [v]_{\mathcal{B}}.$$

(3) Besteht $\mathcal{B}$ aus den Vektoren $v_1, \dots, v_n$ dann sind die Spaltenvektoren von $[f]^{\mathcal{B}}_{\mathcal{B}'}$ gerade die Koordinatenvektoren $[f(v_i)]_{\mathcal{B}'}$ der Bilder der Elemente von $\mathcal{B}$.

Beweis. (1) Ist $v = \sum_{i=1}^n r_i v_i$, dann gilt nach Definition $\varphi(v) = \sum_{i=1}^n r_i \varphi(v_i) = \sum_{i=1}^n r_i e_i = [v]_{\mathcal{B}}$.

(2) Für $x \in \mathbb{K}^n$ und $v := \varphi^{-1}(x) \in V$ gilt nach Teil (1) $x = \varphi(v) = [v]_{\mathcal{B}}$. Nach Definition der Matrix zu einer linearen Abbilding (siehe 3.2) folgt nun

$$[f]^{\mathcal{B}}_{\mathcal{B}'} \cdot [v]_{\mathcal{B}} = (\varphi' \circ f \circ \varphi^{-1})(x) = \varphi'(f(v)) = [f(v)]_{\mathcal{B}'}.$$

(3) folgt sofort aus (2) und der Beschreibung der Matrix zu einer linearen Abbildung $\mathbb{K}^n \to \mathbb{K}^m$ in Proposition 3.2. □

Auch "theoretische" Eigenschaften der Matrixdarstellungen können wir nun leicht beweisen.

Satz 4.15. *Seien V und V' endlichdimensionale $\mathbb{K}$–Vektorräume, mit $\dim(V) = n$ und $\dim(V') = m$ und seien $\mathcal{B}$ und $\mathcal{B}'$ geordnete Basen. Dann definiert die Abbildung $f \mapsto [f]^{\mathcal{B}}_{\mathcal{B}'}$ einen linearen Isomorphismus $L(V, W) \cong M_{m,n}(\mathbb{K})$. Insbesondere gilt $\dim(L(V, W)) = \dim(V)\dim(W)$.*

Beweis. Seien $\varphi : V \to \mathbb{K}^n$ und $\varphi' : V' \to \mathbb{K}^m$ die linearen Isomorphismen, die durch $\mathcal{B}$ und $\mathcal{B}'$ bestimmt werden. Wir zeigen, dass die Abbildung $f \mapsto \varphi' \circ f \circ \varphi^{-1}$ einen linearen Isomorphismus $L(V, W) \to L(\mathbb{K}^n, \mathbb{K}^m)$ definiert. Dann folgt das Resultat, weil nach Proposition 3.2 der Übergang zur Matrix einen linearen Isomorphismus $L(\mathbb{K}^n, \mathbb{K}^m) \to M_{m,n}(\mathbb{K})$ definiert. Nun ist

$$\begin{aligned}\varphi' \circ (f + g) \circ \varphi^{-1}(v) &= \varphi'((f + g)(\varphi^{-1}(v))) = \varphi'(f(\varphi^{-1}(v)) + g(\varphi^{-1}(v))) \\ &= \varphi' \circ f \circ \varphi^{-1}(v) + \varphi' \circ g \circ \varphi^{-1}(v),\end{aligned}$$

also $\varphi' \circ (f + g) \circ \varphi^{-1} = (\varphi' \circ f \circ \varphi^{-1}) + (\varphi' \circ f \circ \varphi^{-1})$. Ganz analog folgt $\varphi' \circ rf \circ \varphi^{-1} = r(\varphi' \circ f \circ \varphi^{-1})$ für $r \in \mathbb{K}$. Damit ist aber unsere Abbildung linear. Da offensichtlich $g \mapsto \varphi'^{-1} \circ g \circ \varphi$ eine Inverse definiert, ist unsere Abbildung bijektiv, also ein linearer Isomorphismus. □

4.16. Komposition und Matrizenmultiplikation. Auch für die allgemeineren Matrixdarstellungen wird die Komposition von linearen Abbildungen wieder durch die Matrizenmultiplikation beschrieben. Man muss hier aber aufpassen, dass die Basen "zusammenpassen".

Satz 4.16. *Seien V, V' und V'' endlichdimensionale $\mathbb{K}$–Vektorräume mit geordneten Basen $\mathcal{B}$, $\mathcal{B}'$ und $\mathcal{B}''$, und seien $f : V \to V'$ und $g : V' \to V''$ lineare Abbildungen. Dann gilt*

(1) $[g \circ f]^{\mathcal{B}}_{\mathcal{B}''} = [g]^{\mathcal{B}'}_{\mathcal{B}''}[f]^{\mathcal{B}}_{\mathcal{B}'}$

(2) Sei $\dim(V) = \dim(V')$. *Dann ist f genau dann ein linearer Isomorphismus, wenn die Matrix $[f]^{\mathcal{B}}_{\mathcal{B}'}$ invertierbar ist. Wenn das der Fall ist, dann ist $[f^{-1}]^{\mathcal{B}'}_{\mathcal{B}}$ die inverse Matrix zu $[f]^{\mathcal{B}}_{\mathcal{B}'}$.*

Beweis. Seien $\varphi : V \to \mathbb{K}^n$, $\varphi' : V' \to \mathbb{K}^m$ und $\varphi'' : V'' \to \mathbb{K}^k$ die linearen Isomorphismen, die durch unsere Basen bestimmt werden. Dann ist

$$\varphi'' \circ (g \circ f) \circ \varphi^{-1} = \varphi'' \circ g \circ (\varphi')^{-1} \circ \varphi' \circ f \circ \varphi^{-1}$$

und (1) folgt sofort, indem man Satz 3.3 auf die linearen Abbildungen $\varphi'' \circ g \circ (\varphi')^{-1} : \mathbb{K}^m \to \mathbb{K}^k$ und $\varphi' \circ f \circ \varphi^{-1} : \mathbb{K}^n \to \mathbb{K}^m$ und die zugehörigen Matrizen anwendet.

(2) Ist f invertierbar, dann sieht man sofort, dass $\varphi \circ f^{-1} \circ (\varphi')^{-1}$ invers zu $\varphi' \circ f \circ \varphi^{-1}$ ist. Damit sind auch die zugehörigen Matrizen invers zueinander nach Satz 3.4.

Ist umgekehrt die Matrix $[f]^{\mathcal{B}}_{\mathcal{B}'}$ invertierbar, dann ist nach Satz 3.4 $\varphi' \circ f \circ \varphi^{-1}$ ein linearer Isomorphismus. Damit ist aber auch $(\varphi')^{-1} \circ (\varphi' \circ f \circ \varphi^{-1}) \circ \varphi = f$ ein linearer Isomorphismus. □

4.17. Verhalten bei Basiswechsel. Durch Wahl von Basen $\mathcal{B}$ und $\mathcal{C}$ für zwei Vektorräume V und W haben wir lineare Abbildungen durch Matrizen beschreiben können. Natürlich müssen wir uns als nächstes überlegen, wie diese Beschreibung von der Wahl der Basen abhängt. Das kann man aber im wesentlichen direkt aus Satz 4.16 ablesen.

Definition 4.17. Sei V ein $\mathbb{K}$–Vektorraum, und seien $\mathcal{B}$ und $\tilde{\mathcal{B}}$ geordnete Basen für V, die aus den Vektoren $v_1, \dots, v_n$ bzw. $\tilde{v}_1, \dots, \tilde{v}_n$ bestehen. Dann ist die *Matrix zum Basiswechsel von $\mathcal{B}$ nach $\tilde{\mathcal{B}}$* die Matrix $[\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}} \in M_n(\mathbb{K})$, also die Matrixdarstellung der Identitätsabbildung bezüglich der beiden Basen.

Da die Identitätsabbildung $\mathrm{id} : V \to V$ ein linearer Isomorphismus ist, ist die Matrix $[\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ zum Basiswechsel nach Satz 4.16 invertierbar. Aus Proposition 4.15 können wir sofort ablesen, wie man diese Matrix konkret bestimmt: Da $\tilde{\mathcal{B}}$ eine Basis für V ist, können wir jeden der Vektoren der Basis $\mathcal{B}$ in der Basis $\tilde{\mathcal{B}}$ entwickeln. Die Koeffizienten a_{ij} von $[\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ sind dann bestimmt durch $v_j = \sum_i a_{ij}\tilde{v}_i$.

Haben wir einen zweiten Vektorraum W (über dem gleichen Körper $\mathbb{K}$) mit Basen $\mathcal{C}$ und $\tilde{\mathcal{C}}$ gegeben, dann können wir analog die Matrix $[\mathrm{id}_W]^{\mathcal{C}}_{\tilde{\mathcal{C}}}$ zum entsprechenden Basiswechsel bilden. Mit Hilfe von Satz 4.16 können wir nun die Abhängigkeit der Matrixdarstellung von der Wahl der Basen für V und W leicht beschreiben:

Korollar 4.17 (zu Satz 4.16). *Seien V und W Vektorräume über einem Körper $\mathbb{K}$, $\mathcal{B}$ und $\tilde{\mathcal{B}}$ Basen von V und $\mathcal{C}$ und $\tilde{\mathcal{C}}$ Basen von W. Dann gilt*

(1) *Die Matrizen $[\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ und $[\mathrm{id}_V]^{\tilde{\mathcal{B}}}_{\mathcal{B}}$ sind invers zueinander.*

(2) *Ist $f : V \to W$ eine beliebige lineare Abbildung, dann ist*

$$[f]^{\tilde{\mathcal{B}}}_{\tilde{\mathcal{C}}} = [id_W]^{\mathcal{C}}_{\tilde{\mathcal{C}}}[f]^{\mathcal{B}}_{\mathcal{C}}[\mathrm{id}_V]^{\tilde{\mathcal{B}}}_{\mathcal{B}} = [id_W]^{\mathcal{C}}_{\tilde{\mathcal{C}}}[f]^{\mathcal{B}}_{\mathcal{C}}([\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}})^{-1}.$$

Beweis. (1) Nach Satz 4.16 ist $[\mathrm{id}_V]^{\tilde{\mathcal{B}}}_{\mathcal{B}}[\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}} = [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}}$. Nachdem die Identitätsabbildung id_V jedes Element von $\mathcal{B}$ auf sich selbst abbildet ist $[\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{B}}$ die Einheitsmatrix $\mathbb{I}_n$.

(2) Natürlich gilt $\mathrm{id}_W \circ f \circ \mathrm{id}_V = f$, und damit folgt die erste Gleichung direkt aus Satz 4.16, und mit Teil (1) folgt auch die zweite Gleichung. □

Beispiel 4.17. Betrachten wir den Raum $\mathbb{R}_2[x]$ der reellen Polynome vom Grad ≤ 2, mit der Basis $\mathcal{B}$, die aus den Polynomen 1, x, und x^2 besteht. Wenden wir die Überlegungen aus 4.12 auf $r_0 = -1$, $r_1 = 0$ und $r_2 = 1$, dann erhalten wir eine andere Basis $\tilde{\mathcal{B}}$ für $\mathbb{R}_2[x]$, die aus den Polynomen $p_1 = \frac{1}{2}x^2 - \frac{1}{2}x$, $p_2 = -x^2+1$, und $p_3 = \frac{1}{2}x^2 + \frac{1}{2}x$ besteht. Nach Definition bedeutet das gerade

$$[\mathrm{id}]^{\tilde{\mathcal{B}}}_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix}$$

Schreibt ein beliebiges Polynom p als $a_1p_1 + a_2p_2 + a_3p_3$ dann sind die a_i gerade die Werte der Polynomfunktion $p : \mathbb{R} \to \mathbb{R}$ in den Punkten -1, 0 und 1. (Das alles kann man in diesem Fall auch problemlos direkt nachrechnen.) Damit sehen wir, dass die Matrix zum umgekehrten Basiswechsel gegeben ist durch

$$[\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

und man verifiziert sofort, dass diese Matrix invers zur obigen ist.

Betrachten wir nun die eindeutige lineare Abbildung $D : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$, die $D(x^2) = 2x$, $D(x) = 1$ und $D(1) = 0$ erfüllt, also die Ableitung von Polynomen. Dann gilt natürlich $[D]^{\mathcal{B}}_{\mathcal{B}} = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{smallmatrix}\right)$, also

$$[D]^{\tilde{\mathcal{B}}}_{\tilde{\mathcal{B}}} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} & 2 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -2 & \frac{3}{2} \end{pmatrix}$$

Diese Matrix berechnet also für eine Polynomfunktion $p : \mathbb{R} \to \mathbb{R}$ die Werte der Ableitung $p' : \mathbb{R} \to \mathbb{R}$ in den Punkten -1, 0, und 1 aus den Werten des Polynoms p in diesen Punkten! Andererseits sieht man aus diesem Beispiel auch gut, dass "mit freiem Auge" kaum eine Ähnlichkeit zwischen den Matrixdarstellungen einer linearen Abbildung bezüglich verschiedener Basen zu erkennen ist.

4.18. Ähnlichkeit von Matrizen. Wir können nun relativ einfach abklären, wann zwei gegebene Matrizen als Matrixdarstellungen der gleichen linearen Abbildung bezüglich verschiedener Basen interpretiert werden können. Als ersten Schritt zeigen wir, dass jede invertierbare Matrix als Matrix zu einem Basiswechsel aufgefasst werden kann.

Lemma 4.18. *Sei V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $\mathcal{B}$ eine geordnete Basis für V. Dann definierte die Abbildung $\tilde{\mathcal{B}} \mapsto [\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ eine Bijektion zwischen der Menge aller Basen von V und der Menge aller invertierbaren $n \times n$–Matrizen über $\mathbb{K}$.*

Beweis. Es bleibt nur die Bijektivität zu verifizieren. Seien $v_1, \dots, v_n$ die Vektoren der Basis $\mathcal{B}$. Kennt man die Matrix $[\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$, dann ist $[\mathrm{id}_V]^{\tilde{\mathcal{B}}}_{\mathcal{B}}$ einfach die Inverse dieser Matrix. Bezeichnet man die Eintragungen der letzteren Matrix mit a_{ij}, dann wissen wir aus 4.17, dass der j–te Vektor in $\tilde{\mathcal{B}}$ gerade durch $\sum a_{ij}v_i$ gegeben ist. Damit können wir $\tilde{\mathcal{B}}$ aus $[\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ zurückgewinnen, also ist die Abbildung $\tilde{\mathcal{B}} \mapsto [\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ injektiv.

Sei andererseits $A \in M_n(\mathbb{K})$ eine beliebige invertierbare $n \times n$–Matrix. Dann definiert $f(x) := A^{-1}x$ einen linearen Isomorphismus $f : \mathbb{K}^n \to \mathbb{K}^n$. Sein $\varphi : V \to \mathbb{K}^n$ der eindeutige lineare Isomorphismus, der $\varphi(v_i) = e_i$ für $i = 1, \dots, n$ erfüllt. Dann ist auch

$g := \varphi^{-1} \circ f \circ \varphi : V \to V$ ein linearer Isomorphismus, also bilden nach Satz 4.10 die Vektoren $g(v_1), \dots, g(v_n)$ eine Basis $\tilde{\mathcal{B}}$ für V. Sind b_{ij} die Eintragungen von A^{-1}, dann ist nach Definition $[g(v_j)]_{\mathcal{B}}$ gerade der j–te Spaltenvektor von A^{-1}, also $g(v_j) = \sum_i b_{ij} v_i$. Damit ist aber $A^{-1} = [\mathrm{id}_V]^{\tilde{\mathcal{B}}}_{\mathcal{B}}$, also $A = [\mathrm{id}_V]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$, was die Surjektivität unserer Abbildung beweist. □

Definition 4.18. Man sagt, zwei $m \times n$–Matrizen $A, B \in M_{m,n}(\mathbb{K})$ heißen *ähnlich* und schreibt $A \sim B$, wenn es invertierbare Matrizen $C \in M_m(\mathbb{K})$ und $D \in M_n(\mathbb{K})$ gibt, sodass $B = CAD^{-1}$ gilt.

Bemerkung 4.18. (1) Dass man in der Definition $B = CAD^{-1}$ statt $B = CAD$ verwendet ist natürlich formal unerheblich, es wird aber für den späteren Begriff für quadratische Matrizen günstiger sein.

(2) Aus der Definition folgt sofort, dass Ähnlichkeit von Matrizen eine Äquivalenzrelation ist: Weil $A = \mathbb{I}_m A \mathbb{I}_n$ ist gilt $A \sim A$. Ist $B = CAD^{-1}$, dann ist $A = C^{-1}BD$, also folgt aus $A \sim B$ automatisch $B \sim A$. Ist schließlich $B = CAD^{-1}$ und $E = \tilde{C}B\tilde{D}^{-1}$, dann ist $E = (\tilde{C}C)A(\tilde{D}D)^{-1}$, also folgt aus $A \sim B$ und $B \sim E$ auch $A \sim E$.

(3) Seien V und W $\mathbb{K}$–Vektorräume mit $\dim(V) = n$ und $\dim(W) = m$ und Basen $\mathcal{B}$ und $\mathcal{B}'$, $f : V \to W$ eine lineare Abbildung und $A = [f]^{\mathcal{B}}_{\mathcal{B}'}$. Dann ist eine Matrix $B \in M_{m,n}(\mathbb{K})$ genau dann ähnlich zu A, wenn es Basen $\tilde{\mathcal{B}}$ von V und $\tilde{\mathcal{B}}'$ von W gibt, sodass $B = [f]^{\tilde{\mathcal{B}}}_{\tilde{\mathcal{B}}'}$ gilt.

Zwei Matrizen sind also genau dann ähnlich, wenn man sie als Matrixdarstellung der gleichen linearen Abbildung bezüglich verschiedener Basen interpretieren kann.

Satz 4.18. *Für $r \leq m, n$ sei $\mathbb{J}_r \in M_{m,n}(\mathbb{K})$ die Matrix, deren erste r Spalten die Einheitsvektoren $e_1, \dots, e_r$ sind, während alle weiteren Spalten nur aus Nullen bestehen.*

Dann ist eine Matrix $A \in M_{m,n}(\mathbb{K})$ genau dann ähnlich zu $\mathbb{J}_r$, wenn sie Rang r hat. Insbesondere sind zwei Matrizen $A, B \in M_{m,n}(\mathbb{K})$ genau dann ähnlich, wenn sie den gleichen Rang haben.

Beweis. Nach allem was wir schon beobachtet haben, ist das nur noch eine einfache Übersetzung von Tatsachen über lineare Abbildungen, die wir bereits verifiziert haben. Aus Proposition 4.13 folgt sofort, dass ähnliche Matrizen den gleichen Rang haben.

Hat $A \in M_{m,n}(\mathbb{K})$ Rang r, dann betrachten wir die lineare Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$, die gegeben ist durch $f(x) = Ax$. Dann ist $\dim(\mathrm{Im}(f)) = r$ und $\dim(\mathrm{Ker}(f)) = n - r$. Wie im Beweis des Dimensionssatzes 4.11 können wir eine Basis $\mathcal{B}$ für $\mathbb{K}^n$ wählen, die aus Vektoren $v_1, \dots, v_n$ besteht, sodass $v_{r+1}, \dots, v_n$ eine Basis für $\mathrm{Ker}(f)$ bilden. Dann sind die Vektoren $w_1 = f(v_1), \dots, w_r = f(v_r)$ eine Basis für $\mathrm{Im}(f)$ und können durch Vektoren $w_{r+1}, \dots, w_m$ zu einer Basis $\mathcal{C}$ für W erweitert werden. Nach Konstruktion gilt $f(v_i) = w_i$ für $i = 1, \dots, r$ und $f(v_i) = 0$ für $i > r$, also $[f]^{\mathcal{B}}_{\mathcal{C}} = \mathbb{J}_r$, also $A \sim J_r$. Ist B eine weitere Matrix mit Rang r, dann gilt analog $B \sim \mathbb{J}_r$ und da $\sim$ eine Äquivalenzrelation ist, folgt $A \sim B$. □

In gewissem Sinn haben wir mit diesem Resultat das Studium von linearen Abbildungen zwischen verschiedenen $\mathbb{K}$–Vektorräumen abgeschlossen und alles gesagt, was man über solche Abbildungen sagen kann. Insbesondere ist der Rang die einzige wesentliche Kenngröße so einer linearen Abbildung, weitere Unterscheidungsmöglichkeiten gibt es nicht, weil ja nach geschickter Wahl von Basen alle linearen Abbildungen vom Rang r gleich aussehen.

Ein Problem, dass noch offen bleibt und das ein zentrales Thema für die weiteren Teile der Vorlesung sein wird, ist das Studium von linearen Abbildungen von einem

$\mathbb{K}$–Vektorraum V auf sich selbst. In diesem Fall wird man natürlich nur Matrixdarstellungen der Form $[f]^{\mathcal{B}}_{\mathcal{B}}$ betrachten (und nicht Matrixdarstellungen bezüglich verschiedener Basen ansehen). Dass dies eine feinere Unterscheidung von Abbildungen erlaubt ist anhand eines Beispiels sofort einzusehen. Betrachten wir den $\mathbb{R}$–Vektorraum $\mathbb{R}^n$ und für eine Zahl $\lambda \neq 0$ die Abbildung $f(x) = \lambda x$. Dann ist natürlich $[f]^{\mathcal{B}}_{\mathcal{B}} = \lambda \mathbb{I}_n$ für jede Basis $\mathcal{B}$. Man kann also diese Abbildungen für verschiedene λ unterscheiden, obwohl sie alle Rang n haben.

KAPITEL 5

Einige Konstruktionen und Anwendungen

In diesem Kapitel besprechen wir zunächst eine Anwendung der linearen Algebra in der Codierungstheorie und dann die Anwendung der bisherigen Konzepte auf geometrische Fragestellungen. Am Ende des Kapitels besprechen wir noch zwei wichtige allgemeine Konstruktionen mit Vektorräumen.

Exkurs: Lineare Codes

Wir wollen als nächstes kurz eine Anwendung der linearen Algebra über endlichen Körpern besprechen, nämlich die sogenannten linearen Codes. Der Einfachheit halber werden wir uns auf den Körper $\mathbb{Z}_2$ beschränken und hauptsächlich ein konkretes Beispiel (nämlich den sogenannten Hemming–(7,4)–Code) besprechen, anstatt allgemeine Theorie zu entwickeln.

5.1. Bei Codes sollte man nicht an Geheimbotschaften denken, sondern daran, dass man eine Botschaft übertragen möchte, wobei man damit rechnen muss, dass bei der Übertragung Fehler auftreten können. Der Einfachheit halber treffen wir die (ziemlich realistische) Annahme, dass unsere Botschaft einfach eine Folge von "Worten" einer fixen Länge k in den "Buchstaben" $\{0, 1\}$ ist. Wir können also ein solches Wort als ein Element $(x_1, \dots, x_k) \in \mathbb{Z}_2^k$ betrachten. Die einzige Rechenregel in $\mathbb{Z}_2$, die nicht daraus folgt, das 0 additiv neutral und 1 multiplikativ neutral ist, ist $1 + 1 = 0$.

Wir wollen nun annehmen, dass wir unsere Botschaft auf irgend eine Art und Weise übertragen, und das vereinzelt Fehler auftreten können, also vereinzelt Nullen in Einsen und Einsen in Nullen verwandelt werden. Wir wollen solche Fehler erkennbar und wenn möglich sogar korrigierbar machen, indem wir zusätzliche Informationen übertragen (in die bei der Übertragung natürlich auch Fehler hinein kommen können). Damit so etwas praktisch nutzbar ist, sollte natürlich die Codierung (also das Hinzufügen der zusätzlichen Information) und die Decodierung (also das Zurückgewinnen der ursprünglichen Information) möglichst einfach sein.

Eine einfacher Schritt in dieser Richtung ist die Übertragung eines sogenannten Prüf– oder Paritätsbits. Man macht einfach aus einem Wort $(x_1, \dots, x_k)$ der Länge k ein Wort der Länge $k + 1$, indem man x_{k+1} so wählt, dass in dem Wort $(x_1, \dots, x_{k+1})$ eine gerade Anzahl von Einsen vorkommt, d.h. $x_{k+1} = x_1 + \dots + x_k$ (in $\mathbb{Z}_2$!). Empfängt man ein Wort der Länge $k+1$, dann bildet man $x_1 + \dots + x_{k+1}$. Ist diese Summe gleich Null, dann nimmt man einfach die ersten k Elemente als übertragenes Wort. Ist die Summe gleich Eins, dann weiß man, dass in der Übertragung ein Fehler aufgetreten ist (aber nicht an welcher Stelle). Treten zwei Fehler auf, dann kann man das mit dieser Methode nicht mehr erkennen.

Man kann diese Methode so interpretieren, dass man auf die Datenworte aus $\mathbb{Z}_2^k$ zunächst eine lineare Abbildung $\mathbb{Z}_2^k \to \mathbb{Z}_2^{k+1}$ anwendet und dann das Bild überträgt. Dann kann man mit einer linearen Abbildung $\mathbb{Z}_2^{k+1} \to \mathbb{Z}_2$ feststellen ob ein Fehler aufgetreten ist, und wenn das nicht passiert ist, das ursprüngliche Wort durch eine lineare Abbildung $\mathbb{Z}_2^{k+1} \to \mathbb{Z}_2^k$ rekonstruieren.

Wichtig für die Qualität eines Codes ist nicht die lineare Abbildung, die zum Bilden der Codeworte verwendet wird, sondern nur ihr Bild. Deshalb definiert man einen *linearen* (n, k)*–Code über* $\mathbb{Z}_2$ als einen k–dimensionalen Teilraum C von $\mathbb{Z}_2^n$. Wir wollen nun einen (sehr guten) linearen $(7, 4)$–Code, den sogenannten *Hemming* $(7, 4)$*–Code* beschreiben. Bei diesem Code fügt man zu 4 Datenbits 3 Prüfbits, womit bis zu 2 Fehler sicher erkannt und Einzelfehler sicher korrigiert werden können. (Warum gerade die Kombination $(7, 4)$ günstig ist, werden wir im Verlauf der Konstruktion sehen.)

Wir suchen also einen 4–dimensionalen Teilraum C von $\mathbb{Z}_2^7$, dessen Element die *gültigen Codeworte* sind. Eine einfache Methode, so einen Teilraum anzugeben, ist als Kern einer surjektiven linearen Abbildung $\varphi : \mathbb{Z}_2^7 \to \mathbb{Z}_2^3$, deren zugehörige Matrix K die *Kontrollmatrix* des Codes heißt. Somit ist ein (empfangenes) Codewort x genau dann gültig, wenn $Kx = 0$ gilt. Das Auftreten eines Fehlers in der Übertragung kann man so interpretieren, dass zu einem gültigen Codewort x ein *Fehlerwort* y addiert wird. (In $\mathbb{Z}_2$ entspricht ja Addition von 1 genau dem Vertauschen von 0 und 1.) Die Anzahl der Übertragungsfehler ist genau die Anzahl der Einsen in y. Wegen der Linearität gilt nun $K(x + y) = Kx + Ky = Ky$, also können Fehler genau dann erkannt werden, wenn die Fehlerworte nicht im Kern von φ liegen. Wollen wir also Einzelfehler erkennen, dann muss zumindest $Ke_i \neq 0$ für $i = 1, \dots, 7$ gelten.

Wir wollen aber Einzelfehler nicht nur erkennen, sondern sogar korrigieren können. Um das zu tun, müssen wir die Einzelfehler von einander unterscheiden können, also muss $Ke_i \neq Ke_j$ für $i \neq j$ gelten. Dies lässt uns für K aber im Wesentlichen nur noch eine Möglichkeit, weil $\mathbb{Z}_2^3$ genau $2^3 = 8$ Elemente, also genau 7 Elemente ungleich Null hat. Wir wählen also $K = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$. Diese Wahl ist so getroffen, dass Ke_i gerade die Darstellung von i als Binärzahl ist. Wir erhalten aber direkt noch eine wichtige Konsequenz: Da in $\mathbb{Z}_2^3$ immer $x = -x$ gilt, und $Ke_i \neq Ke_j$ für $i \neq j$ gilt, ist auch $K(e_i + e_j) = Ke_i + Ke_j \neq 0$ für $i \neq j$. Das bedeutet aber, dass auch Doppelfehler noch sicher erkannt werden.

Somit gehen wir folgendermaßen vor: Empfangen wir ein Wort x, dann bilden wir Kx. Ist das gleich Null, dann ist kein (erkennbarer) Übertragungsfehler aufgetreten, und x ist das richtige Codewort. Ist $Kx \neq 0$, dann ist sicher ein Übertragungsfehler aufgetreten. War das ein Einzelfehler (die wahrscheinlichste Möglichkeit) dann erhält man das ursprüngliche Codewort, indem man Kx als Binärdarstellung von j interpretiert und dann die j–te Eintragung von x ändert (also $x + e_j$ bildet).

Nun wollen wir noch eine konkrete Codierung angeben, also eine lineare Abbildung $\mathbb{Z}_2^4 \to \operatorname{Ker}(\varphi)$, deren zugehörige Matrix die *Generatormatrix* G des Codes heißt. Die einfachste Möglichkeit ist natürlich, an ein Wort $(x_1, \dots, x_4)$ einfach drei geeignete Prüfbits (x_5, x_6, x_7) anzuhängen. Um das zu realisieren müssten die ersten 4 Zeilen der 7×4–Matrix G die 4×4–Einheitsmatrix $\mathbb{I}_4$ bilden. Das stellt auch sicher, dass G einer injektiven linearen Abbildung entspricht, also genügt es zusätzlich sicher zu stellen, dass $KG = 0$ gilt, denn dann liefert G aus Dimensionsgründen einen linearen Isomorphismus $\mathbb{Z}_2^4 \to \operatorname{Ker}(\varphi)$. Nun verifiziert man leicht durch Lösen eines linearen Gleichungssystems, dass diese beiden Bedingungen G eindeutig festlegen. Bestimmt man G explizit, dann sieht man, dass die Prüfbits gegeben sind durch $x_5 = x_2 + x_3 + x_4$, $x_6 = x_1 + x_3 + x_4$ und $x_7 = x_1 + x_2 + x_4$.

Beispiel. Nehmen wir an, das ursprüngliche Datenwort ist $(1, 0, 1, 0)$. Nach der Vorschrift zum Bilden der Prüfbits wird das zu $x = (1, 0, 1, 0, 1, 0, 1)$ kodiert. Wird dieses

Codewort korrekt übertragen, dann erhält man $Kx = 0$, und gewinnt das ursprüngliche Wort $(1, 0, 1, 0)$ als die ersten vier Stellen. Tritt etwa an der zweiten Stelle ein Fehler auf, dann empfangen wir statt x das Wort $x' = (1, 1, 1, 0, 1, 0, 1)$. Nachrechnen ergibt $Kx' = (0, 1, 0)$. Also sehen wir, dass ein Fehler aufgetreten ist, und zwar vermutlich an der zweiten Stelle. Also sollte das übertragene Wort $(1, 0, 1, 0, 1, 0, 1)$ sein, dessen erste vier Stellen das richtige Datenwort $(1, 0, 1, 0)$ liefern. Nehmen wir nun an, dass zwei Übertragungsfehler auftreten, etwa an der zweiten und der letzten Stelle. Dann empfangen wir das Wort $x'' = (1, 1, 1, 0, 1, 0, 0)$ und $Kx'' = (1, 0, 1)$. Wir stellen also (richtigerweise) fest, dass ein Übertragungsfehler aufgetreten ist, der wahrscheinlichste Übertragungsfehler wäre aber eine Änderung in der fünften Stelle (also einem Prüfbit), und wir würden (fälschlicherweise) vermuten, dass das ursprüngliche Wort $(1, 1, 1, 0)$ war.

Einige Grundideen der affinen und projektiven Geometrie

Als nächstes werden wir illustrieren, wie man Methoden der linearen Algebra auf geometrische Fragestellungen anwenden kann. Die Art von Geometrie, die wir hier betreiben, ist die sogenannte affine Geometrie, in der man keine Längen oder Winkel messen kann. Trotzdem erhält man viele interessante Konzepte und auch einige interessante Resultate. Man kann affine Geometrie problemlos über allgemeinen Körpern betreiben, der Einfachheit halber werden wir aber in diesem Abschnitt nur über $\mathbb{R}$ arbeiten.

5.2. $\mathbb{A}^n$ und affine Teilräume. Für Anwendungen in der Geometrie sind Vektorräume nicht ganz der richtige Rahmen, wobei das Hauptproblem die ausgezeichnete Rolle des Nullpunktes eines Vektorraumes ist. (Geometrisch gesehen sollten natürlich alle Punkte "gleichberechtigt" sein.) Versucht man, die ausgezeichnete Rolle des Nullpunktes zu eliminieren, dann führt das zum Begriff des affinen Raumes. ähnlich wie bei Vektorräumen kann man diesen Begriff allgemein abstrakt einführen. Um die Dinge einfach zu halten, werden wir uns hier auf eine konkrete Realisierung eines n–dimensionalen affinen Raumes einschränken.

Im folgenden betrachten wir $\mathbb{R}^n$ sowohl als "eigenständigen" Vektorraum, als auch als den Teilraum $\{x = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} = 0\}$.

Definition 5.2. (1) Für zwei Punkte x, y in einem Vektorraum V definieren wir den *Verbindungsvektor* $\overrightarrow{xy} \in V$ zwischen x und y durch $\overrightarrow{xy} = y - x$. Nach Definition gilt also $y = x + \overrightarrow{xy}$.

(2) Der *n–dimensionale affine Raum* ist die Teilmenge

$$\mathbb{A}^n := \{x = (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} = 1\} \subset \mathbb{R}^{n+1}.$$

Für zwei Punkte $x, y \in \mathbb{A}^n$ gilt dann $\overrightarrow{xy} \in \mathbb{R}^n \subset \mathbb{R}^{n+1}$.

(3) Eine nichtleere Teilmenge $A \subset \mathbb{A}^n$ heißt ein *affiner Teilraum* von $\mathbb{A}^n$, wenn die Teilmenge $\overrightarrow{A} := \{\overrightarrow{ab} : a, b \in A\} \subset \mathbb{R}^n$ ein Teilraum ist, und für beliebiges $a \in A$ und $v \in \overrightarrow{A}$ auch $a + v$ in A liegt. Man nennt dann $\overrightarrow{A}$ den *modellierenden Teilraum* zu A.

(4) Die Dimension $\dim(A)$ eines affinen Teilraumes $A \subset \mathbb{A}^n$ ist definiert als die Dimension $\dim(\overrightarrow{A})$ des modellierenden Teilraumes. Insbesondere ist $\dim(A) \leq n$. Eindimensionale affine Teilräume heißen *affine Geraden*, zweidimensionale *affine Ebenen.*

Bemerke, dass für $x \in \mathbb{A}^n$ und $v \in \mathbb{R}^n$ natürlich $x + v \in \mathbb{A}^n$ gilt. Somit ist $\mathbb{A}^n$ selbst ein affiner Teilraum von $\mathbb{A}^n$. Fixiert man eine Punkt $a_0 \in \mathbb{A}^n$, dann definiert $f(v) := a_0 + v$ eine Funktion $f : \mathbb{R}^n \to \mathbb{A}^n$. Man verifiziert sofort, dass $a \mapsto \overrightarrow{a_0a}$ invers zu dieser Funktion ist, also ist f bijektiv. Das drückt die intuitive Idee aus, dass man $\mathbb{A}^n$

aus $\mathbb{R}^n$ erhält, indem man vergisst, wo der Ursprung liegt. Für einen affinen Teilraum $A \subset \mathbb{A}^n$ ist natürlich $\dim(A) = \dim(\overrightarrow{A}) \leq n$ und man verifiziert sofort, dass für eine Punkt $a_0 \in A$ die Abbildung $v \mapsto a_0 + v$ eine Bijektion $\overrightarrow{A} \to A$ definiert. Insbesondere folgt aus $\dim(A) = n$ sofort $A = \mathbb{A}^n$.

In $\mathbb{A}^n$ kann man natürlich im allgemeinen keine Linearkombinationen bilden, weil $\mathbb{A}^n$ kein Teilraum von $\mathbb{R}^{n+1}$ ist. Sind $a_1, \dots, a_k \in \mathbb{A}^n$ Punkte und $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ Skalare, dann können wir natürlich in $\mathbb{R}^{n+1}$ den Vektor $\lambda_1 a_1 + \cdots + \lambda_k a_k$ bilden. Nachdem jedes a_i als letzte Koordinate 1 hat, ist die letzte Koordinate dieses Punktes gerade $\lambda_1 + \cdots + \lambda_k$. Damit erhalten wir aber einen wohldefinierten Punkt $\sum \lambda_i a_i \in \mathbb{A}^n$, sofern wir voraussetzen, dass $\sum_i \lambda_i = 1$ gilt. Man nennt dann $\sum_i \lambda_i a_i$ eine *affine Kombination* der Punkte $a_1, \dots, a_k$.

Die Tatsache, dass man $\sum_i \lambda_i = 1$ verlangen muss hängt nicht damit zusammen, dass wir $\mathbb{A}^n$ als Menge der Punkte mit letzter Koordinate 1 realisiert haben. Solche Kombinationen machen auch in abstrakten affinen Räumen Sinn. Zum Verständnis ist es hilfreich, den Fall $k = 2$ zu betrachten, indem man $\lambda_1 + \lambda_2 = 1$ als $\lambda_2 = 1 - \lambda_1$ umformulieren kann. Dann erhält man $\lambda_1 a_1 + (1 - \lambda_1) a_2 = a_2 + \lambda_1(a_1 - a_2) = a_2 + \lambda_1 \overrightarrow{a_2 a_1}$. Das beschreibt genau die Gerade durch a_1 und a_2, die natürlich unabhängig von der Lage des Nullpunktes Sinn macht.

Wir können nun affine Teilräume schön mittels affiner Kombinationen charakterisieren.

PROPOSITION 5.2. *Eine nichtleere Teilmenge $A \subset \mathbb{A}^n$ ist genau dann eine affiner Teilraum, wenn für beliebige Punkte $a_1, \dots, a_k \in A$ und Skalare $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ mit $\sum_i \lambda_i = 1$ auch $\sum_i \lambda_i a_i$ in A liegt.*

BEWEIS. ($\Rightarrow$) Sei $A \subset \mathbb{A}^n$ ein affiner Teilraum und seien $a_1, \dots, a_k \in A$ beliebig und $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ so, dass $\sum_i \lambda_i = 1$ gilt. Für $k = 1$ nichts zu beweisen, also sei $k > 1$. Dann ist $\lambda_1 = 1 - \sum_{i=2}^k \lambda_i$ und somit

$$\lambda_1 a_1 + \cdots + \lambda_k a_k = a_1 + \lambda_2(a_2 - a_1) + \cdots + \lambda_k(a_k - a_1).$$

Für $i = 2, \dots, k$ gilt $a_i - a_1 = \overrightarrow{a_1 a_i} \in \overrightarrow{A}$. Nach Voraussetzung ist $\overrightarrow{A}$ ein Teilraum von $\mathbb{R}^n$, also ist $\lambda_2(a_2 - a_1) + \cdots + \lambda_k(a_k - a_1) \in \overrightarrow{A}$ und wiederum nach Voraussetzung landet man in A, wenn man diesen Vektor zu $a_1 \in A$ addiert.

($\Leftarrow$) Da A nichtleer ist, gibt es einen Punkt $a \in A$, und damit ist $0 = \overrightarrow{aa} \in \overrightarrow{A}$. Ist $v \in \overrightarrow{A}$ und $a \in A$, dann gibt es nach Voraussetzung $b, c \in \mathbb{A}$, sodass $v = \overrightarrow{bc}$ gilt. Für $\lambda \in \mathbb{R}$ ist aber $a + \lambda v = a - \lambda b + \lambda c$ und das ist das eine affine Kombination und liegt somit nach Voraussetzung in A. Damit folgt einerseits $a + v \in A$ und andererseits $\lambda v \in \overrightarrow{A}$. Ist $w \in \overrightarrow{A}$ ein weiterer Vektor, dann ist $(a + v) + w = a + (v + w) \in A$, also $v + w \in \overrightarrow{A}$. □

BEMERKUNG 5.2. Man kann völlig analog affine Teilräume von in reellen Vektorräumen definieren (und $\mathbb{A}^n$ ist selbst ein affiner Teilraum von $\mathbb{R}^{n+1}$). Wir haben auch schon Beispiele für affine Teilräume in diesem Sinn kennen gelernt: Seien V und W Vektorräume über $\mathbb{R}$, $f : V \to W$ eine lineare Abbildung und $b \in \mathrm{Im}(f) \subset W$ ein Punkt. Dann können wir $f^{-1}(b) = \{x \in V : f(x) = b\} \subset V$ betrachten. Für $x, y \in f^{-1}(b)$ gilt $f(y - x) = f(y) - f(x) = 0$, also $\overrightarrow{xy} \in \mathrm{Ker}(f)$. Ist umgekehrt $v \in \mathrm{Ker}(f)$, dann ist $f(x + v) = f(x) + f(v) = b + 0$, also ist $x + v \in f^{-1}(b)$. Damit ist aber $f^{-1}(b)$ ein affiner Teilraum von V mit modellierendem Teilraum $\mathrm{Ker}(f)$. Insbesondere bilden die Lösungsmengen von linearen Gleichungssystemen (über $\mathbb{R}$) immer reelle affine Teilräume in diesem Sinn.

Die Charakterisierung affiner Teilräume über affine Kombinationen zeigt leicht, dass man analog zum erzeugten Teilraumes auch einen Begriff des erzeugten affine Teilraumes erhält. Damit ergibt sich natürlich auch automatisch der Begriff eines affinen Erzeugendensystems.

KOROLLAR 5.2. *(1) Der Durchschnitt über eine beliebige Familie affiner Teilräume von $\mathbb{A}^n$ ist selbst ein affiner Teilraum.*

(2) Für eine nichtleere Teilmenge $B \subset \mathbb{A}^n$ gibt einen eindeutigen affinen Teilraum $\langle B \rangle_{aff} \subset \mathbb{A}^n$, das affine Erzeugnis *von B, sodass $B \subset \langle B \rangle_{aff}$ gilt und jeder affine Teilraum $A \subset \mathbb{A}^n$, der B enthält auch $\langle B \rangle_{aff}$ enthält.*

BEWEIS. (1) Das folgt sofort aus der Proposition.

(2) Man definiert $\langle B \rangle_{\text{aff}}$ als den Durchschnitt aller affinen Teilräume von $\mathbb{A}^n$, die B enthalten. Nach (1) ist das ein affiner Teilraum, der offensichtlich die gewünschte Eigenschaft hat. Das diese Eigenschaft das affine Erzeugnis eindeutig charakterisiert folgt sofort wie im Fall von Teilräumen. □

Man kann leicht sehen, dass für einen affinen Teilraum $A \subset \mathbb{A}^n$ die Punkte $a_0, \dots, a_k$ genau dann ein affines Erzeugendensystem bilden, wenn $\overrightarrow{a_0a_1}, \dots, \overrightarrow{a_0a_k}$ ein Erzeugendensystem für den Teilraum $\overrightarrow{A} \subset \mathbb{R}^n$ bilden. Insbesondere besteht ein affines Erzeugendensystem aus mindestens $\dim(A) + 1$ Punkten.

5.3. Affine Abbildungen. Affine Abbildungen sind gerade die strukturerhaltenden Abbildungen zwischen affinen Räumen. Die Definition sollte nicht überraschend sein.

DEFINITION 5.3. (1) Eine Funktion $f : \mathbb{A}^n \to \mathbb{A}^m$ heißt eine *affine Abbildung*, wenn es eine lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ gibt, sodass für je zwei Punkte $a, b \in \mathbb{A}^n$ die Gleichung $f(b) = f(a) + \varphi\left(\overrightarrow{ab}\right)$ gilt.

(2) Ein affiner Isomorphismus ist eine bijektive affine Abbildung. Man sieht leicht, dass dann auch die inverse Funktion affin ist.

Bemerke, dass man die Gleichung in der Definition der affinen Abbildung auch als $\overrightarrow{f(a)f(b)} = \varphi\left(\overrightarrow{ab}\right)$ lesen kann. Sei $f : \mathbb{A}^n \to \mathbb{A}^m$ eine affine Abbildung mit zugehöriger linearer Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ und seien $a \in \mathbb{A}^n$ und $\tilde{a} \in \mathbb{A}^m$ beliebig gewählte Punkte. Dann können wir diese Punkte wie in 5.1 benutzen um die affinen Räume mit den modellierenden Vektorräumen zu identifizieren. Die Abbildung $\mathbb{R}^n \to \mathbb{R}^m$, die in dieser Identifikation der affinen Abbildung f entspricht ist $v \mapsto f(a+v) - \tilde{a} = \varphi(v) + w$, wobei $w := \overrightarrow{\tilde{a}f(a)} \in \mathbb{R}^m$. Eine affine Abbildung entspricht also einer linearen Abbildung und einer Translation.

Man kann affine Abbildungen wiederum über Verträglichkeit mit affinen Kombinationen charakterisieren. Auch affine Isomorphismen erlauben eine einfache Charakterisierung.

PROPOSITION 5.3. *(1) Eine Funktion $f : \mathbb{A}^n \to \mathbb{A}^m$ ist genau dann eine affine Abbildung, wenn für beliebige Punkte $a_1, \dots, a_k \in \mathbb{A}^n$ und Skalare $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, so dass $\sum_i \lambda_i = 1$*

$$f(\lambda_1 a_1 + \dots + \lambda_k a_k) = \lambda_1 f(a_1) + \dots + \lambda_k f(a_k)$$

gilt.

(2) Eine affine Abbildung $f : \mathbb{A}^n \to \mathbb{A}^m$ ist genau dann ein affiner Isomorphismus, wenn die zugehörige lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ ein linearer Isomorphismus ist. Insbesondere ist das nur für $n = m$ möglich.

Beweis. (1): Sei f eine affine Abbildung und seien a_i und λ_i gegeben. Für $k = 1$ ist nichts zu beweisen, also nehmen wir $k \geq 2$ an. Dann haben wir schon im Beweis von Proposition 5.2 gesehen, dass $\lambda_1 = 1 - \sum_{i=2}^{k} \lambda_i$ und damit

$$\lambda_1 a_1 + \cdots + \lambda_k a_k = a_1 + \lambda_2 \overrightarrow{a_1 a_2} + \cdots + \lambda_k \overrightarrow{a_1 a_k}$$

gilt. Damit ist aber

$$\begin{aligned} f(\textstyle\sum_i \lambda_i a_i) &= f(a_1) + \varphi(\lambda_2 \overrightarrow{a_1 a_2} + \cdots + \lambda_k \overrightarrow{a_1 a_k}) = f(a_1) + \lambda_2 \varphi(\overrightarrow{a_1 a_2}) + \cdots + \lambda_k \varphi(\overrightarrow{a_1 a_k}) \\ &= f(a_1) + \lambda_2 (f(a_2) - f(a_1)) + \cdots + \lambda_k (f(a_k) - f(a_1)) = \textstyle\sum_i \lambda_i f(a_i). \end{aligned}$$

Sei umgekehrt $f : \mathbb{A}^n \to \mathbb{A}^m$ verträglich mit affinen Kombinationen. Fixiere einen Punkt $a_0 \in \mathbb{A}^n$ und definiere $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ durch $\varphi(v) := f(a_0 + v) - f(a_0)$. Für $v \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$, setze $a := a_0 + v$ und betrachte die affine Kombination $(1 - \lambda)a_0 + \lambda a = a_0 + \lambda(a - a_0) = a_0 + \lambda v$. Die rechte Seite der Gleichung zeigt, dass $f((1 - \lambda)a_0 + \lambda a) = f(a_0) + \varphi(\lambda v)$ gilt. Aus der Verträglichkeit mit affinen Kombinationen folgt aber

$$f((1 - \lambda)a_0 + \lambda a) = (1 - \lambda)f(a_0) + \lambda f(a) = f(a_0) + \lambda(f(a) - f(a_0)) = f(a_0) + \lambda\varphi(v),$$

und somit $\varphi(\lambda v) = \lambda\varphi(v)$. Für einen weiteren Vektor $w \in \mathbb{R}^n$ setze $b = a_0 + w$ und $c = a_0 + v + w$. Dann gilt natürlich $f(c) = f(a_0) + \varphi(v + w)$. Andererseits ist $c = a + b - a_0$ und die rechte Seite ist eine affine Kombination, also ist

$$f(c) = f(a) + f(b) - f(a_0) = f(a_0) + \varphi(v) + f(a_0) + \varphi(w) - f(a_0) = f(a_0) + \varphi(v) + \varphi(w).$$

Damit folgt $\varphi(v + w) = \varphi(v) + \varphi(w)$ also ist φ linear. Für beliebige Punkte $a, b \in A$ folgt dann $\overrightarrow{a_0 b} = \overrightarrow{a_0 a} + \overrightarrow{ab}$ und damit

$$f(b) = f(a_0) + \varphi(\overrightarrow{a_0 b}) = f(a_0) + \varphi(\overrightarrow{a_0 a}) + \varphi(\overrightarrow{ab}) = f(a) + \varphi(\overrightarrow{ab}).$$

(2) Wir haben schon bemerkt, dass $\overrightarrow{f(a)f(b)} = \varphi\left(\overrightarrow{ab}\right)$ für alle $a, b \in \mathbb{A}^n$ gilt. Damit ist aber $f(a) = f(b)$ äquivalent zu $\varphi(\overrightarrow{ab}) = 0$ und somit Injektivität von f äquivalent zu Injektivität von φ. Die Gleichung zeigt auch, dass Surjektivität von f Surjektivität von φ impliziert, und weil $w \mapsto f(a) + w$ eine surjektive Abbildung $\mathbb{R}^m \to \mathbb{A}^m$ ist, gilt auch die Umkehrung.

□

Bemerkung 5.3. Aus der Charakterisierung in Teil (1) folgt sofort, dass Bilder und Urbilder von affinen Teilräumen unter affinen Abbildungen wieder affine Teilräume sind. Affine Teilräume sind ja gerade Teilmengen die abgeschlossen unter affinen Kombinationen sind, und affine Abbildungen vertauschen mit affinen Kombinationen. Im Fall von affinen Isomorphismen bleiben sogar die Dimensionen erhalten, also sind insbesondere Bilder und Urbilder von affinen Geraden (bzw. Ebenen) unter affinen Isomorphismen wieder affine Geraden (bzw. Ebenen).

5.4. Beschreibungen über den umgebenden Raum. Bis jetzt haben wir die Realisierung von $\mathbb{A}^n$ als Teilmenge von $\mathbb{R}^{n+1}$ noch nicht wirklich benutzt, die bisherigen Beweise funktionieren auch in abstrakten affinen Räumen. Jetzt wollen wir die Realisierung benutzen, um affine Konzepte auf eine weitere Art mit Vektorraumkonzepten in Verbindung zu bringen.

Als ersten Schritt in diese Richtung machen wir folgende Beobachtung: Für einen Punkt $x \in \mathbb{A}^n$ ist $\{x\} \subset \mathbb{A}^n$ ein affiner Teilraum mit modellierendem Teilraum $\{0\}$, also mit Dimension Null. Als Element von $\mathbb{R}^{n+1}$ betrachtet ist aber $x \neq 0$, und bestimmt damit einen eindimensionalen Teilraum $V_x \subset \mathbb{R}^{n+1}$, der explizit als $V_x = \{tx : t \in \mathbb{R}\}$

gegeben ist. Nun ist nach Definition $x = (x_0, \dots, x_n, 1)$ also hat tx als letzte Koordinate gerade t. Damit ist aber $V_x \cap \mathbb{R}^n = \{0\}$ und $V_x \cap \mathbb{A}^n = \{x\}$.

Sei umgekehrt $V \subset \mathbb{R}^{n+1}$ ein eindimensionaler Teilraum, der $\mathbb{R}^n$ nur in 0 schneidet. Wählen wir ein Element $v \in V$ mit $v \neq 0$, dann ist $V = \{tv : t \in \mathbb{R}\}$. Die letzte Koordinate von v muss ungleich Null sein, sonst wäre $V \subset \mathbb{R}^n$. Damit hat aber V einen eindeutigen Schnittpunkt mit $\mathbb{A}^n$. Damit sehen wir, dass wir $\mathbb{A}^n$ mit der Menge aller eindimensionalen Teilräume von $\mathbb{R}^{n+1}$ identifizieren können, die $\mathbb{R}^n$ nur in 0 schneiden.

Wir können analog affine Teilräume und affine Abbildungen beschreiben:

Proposition 5.4. *(1) Eine nichtleere Teilmenge $A \subset \mathbb{A}^n$ ist genau dann ein affiner Teilraum, wenn es einen Teilraum $V \subset \mathbb{R}^{n+1}$ gibt, der $V \cap \mathbb{A}^n = A$ erfüllt. Ist das der Fall, dann ist V eindeutig bestimmt, $\overrightarrow{A} = V \cap \mathbb{R}^n$ und $\dim(A) = \dim(V) - 1$.*

(2) Für eine nichtleere Teilmenge $B \subset \mathbb{A}^n$ gilt $\langle B \rangle_{aff} = \langle B \rangle \cap \mathbb{A}^n$, wobei $\langle B \rangle$ den von B erzeugten Teilraum von $\mathbb{R}^{n+1}$ bezeichnet.

(3) Eine Funktion $f : \mathbb{A}^n \to \mathbb{A}^m$ ist genau dann eine affine Abbildung, wenn es eine lineare Abbildung $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ gibt, die $\Phi|_{\mathbb{A}^n} = f$ (und damit insbesondere $\Phi(\mathbb{A}^n) \subset \mathbb{A}^m$) erfüllt. In diesem Fall ist Φ eindeutig bestimmt und f ist genau dann ein affiner Isomorphismus, wenn Φ ein linearer Isomorphismus ist.

Beweis. (1) Ist A ein k–dimensionaler affiner Teilraum, dann ist nach Definition $\overrightarrow{A}$ ein k–dimensionaler Teilraum von $\mathbb{R}^n \subset \mathbb{R}^{n+1}$. Wähle einen Punkt $a_0 \in A$ und eine Basis $\{v_1, \dots, v_k\}$ für $\overrightarrow{A}$. Dann haben die v_i als letzte Koordinate 0, während a_0 als letzte Koordinate 1 hat. Ist $0 = \lambda_0 a_0 + \lambda_1 v_1 + \cdots + \lambda_k v_k$ für Skalare $\lambda_i \in \mathbb{R}$, dann folgt durch Betrachten der letzten Koordinate sofort $\lambda_0 = 0$ und dann $\lambda_i = 0$ für alle i wegen der linearen Unabhängigkeit der v_i. Damit erzeugen aber a_0, v_1, …, v_k einen $(k+1)$–dimensionalen Teilraum $V \subset \mathbb{R}^{n+1}$ mit $V \cap \mathbb{R}^n = \overrightarrow{A}$.

Durch Betrachten der letzten Koordinate folgt auch sofort, dass ein Punkt $x = \lambda_0 a_0 + \lambda_1 v_1 + \cdots + \lambda_k v_k$ genau dann in $\mathbb{A}^n$ liegt, wenn $\lambda_0 = 1$ gilt. Damit ist aber $x = a_0 + v$, mit $v = \lambda_1 v_1 + \cdots + \lambda_k v_k \in \overrightarrow{A}$. Da wir in 5.2 schon festgestellt haben, dass $v \mapsto a_0 + v$ eine Bijektion $\overrightarrow{A} \to A$ definiert, folgt $A = V \cap \mathbb{A}^n$.

Sei umgekehrt $V \subset \mathbb{R}^{n+1}$ ein Teilraum. Dann bemerken wir zunächst, dass $V \cap \mathbb{A}^n = \emptyset$ äquivalent zu $V \subset \mathbb{R}^n$ ist, weil man aus einem Punkt dessen letzte Koordinate ungleich Null ist, durch Multiplikation mit einem Skalar sofort einen Punkt mit letzter Koordinate 1 erhält. Ist V nicht in $\mathbb{R}^n$ enthalten, dann ist $V \cap \mathbb{R}^n$ ein Teilraum von $\mathbb{R}^{n+1}$, und wir behaupten, dass die Dimension dieses Teilraumes $\dim(V) - 1$ ist. Ist nämlich $a_0 \in V \cap \mathbb{A}^n$, dann betrachte den eindimensionalen Teilraum $\mathbb{R} \cdot a_0 \subset V$. Offensichtlich ist $\mathbb{R}^{n+1} = \mathbb{R}^n \oplus \mathbb{R} \cdot a_0$ und daraus folgt sofort $V = (V \cap \mathbb{R}^n) \oplus \mathbb{R} \cdot a_0$ und damit mit Proposition 4.9 die Behauptung.

Setze nun $A := V \cap \mathbb{A}^n$. Für $a, b \in A$ ist $\overrightarrow{ab} = y - x \in V \cap \mathbb{R}^n$, und umgekehrt folgt wie oben $x + v \in V \cap \mathbb{A}^n$ für $x \in A$ und $v \in V \cap \mathbb{R}^n$. Damit folgt aber sofort, dass $A \subset \mathbb{A}^n$ ein affiner Teilraum mit modellierendem Teilraum $\overrightarrow{A} = V \cap \mathbb{R}^n$ ist und damit $\dim(A) = \dim(V) - 1$ gilt. Man schließt auch sofort, dass V dann durch die Konstruktion von oben aus A entsteht, was den Beweis von (1) vervollständigt.

(2) Ist B gegeben, dann gibt es nach Teil (1) einen Teilraum $V \subset \mathbb{R}^{n+1}$, der $\langle B \rangle_{\text{aff}} = V \cap \mathbb{A}^n$ erfüllt. Dann gilt natürlich $B \subset V$, also $\langle B \rangle \subset V$ nach Satz 4.1 und damit auch $\langle B \rangle \cap \mathbb{A}^n \subset \langle B \rangle_{\text{aff}}$. Umgekehrt ist nach Teil (1) $\langle B \rangle \cap \mathbb{A}^n$ ein affiner Teilraum von $\mathbb{A}^n$, der natürlich B enthält. Damit folgt aber $\langle B \rangle_{\text{aff}} \subset \langle B \rangle \cap \mathbb{A}^n$ und damit folgt die Behauptung.

(3) Sei zunächst $f : \mathbb{A}^n \to \mathbb{A}^m$ eine affine Abbildung und $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ die zugehörige lineare Abbildung, d.h. $f(b) = f(a) + \varphi(\overrightarrow{ab})$ für alle $a, b \in \mathbb{A}^n$. Für die Standardbasis $\{e_1, \dots, e_{n+1}\}$ von $\mathbb{R}^{n+1}$ gilt $e_1, \dots, e_n \in \mathbb{R}^n$ und $e_{n+1} \in \mathbb{A}^n$. Sei nun $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ die eindeutige lineare Abbildung, die $\Phi(e_i) = \varphi(e_i)$ für $i = 1, \dots, n$ und $\Phi(e_{n+1}) = f(e_{n+1})$ erfüllt. Dann gilt natürlich $\Phi(v) = \varphi(v)$ für $v \in \mathbb{R}^n$. Für $a \in \mathbb{A}^n$ gilt $a = e_{n+1} + \overrightarrow{e_{n+1}a}$ und damit

$$\Phi(a) = \Phi(e_{n+1}) + \Phi(\overrightarrow{e_{n+1}a}) = f(e_{n+1}) + \varphi(\overrightarrow{e_{n+1}a}) = f(a).$$

Ist umgekehrt $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ linear mit $\Phi(\mathbb{A}^n) \subset \mathbb{A}^m$, dann sei $f := \Phi|_{\mathbb{A}^n} : \mathbb{A}^n \to \mathbb{A}^m$. Für $a, b \in \mathbb{A}^n$ gilt dann $f(b) - f(a) = \Phi(b) - \Phi(a) = \Phi(b - a)$. Das zeigt, dass $\Phi(\mathbb{R}^n) \subset \mathbb{R}^m$ gilt und für $\varphi := \Phi|_{\mathbb{R}^n}$ erhalten wir $f(b) - f(a) = \varphi(\overrightarrow{ab})$, also ist f affin.

Zum Beweis der letzten Bemerkung bemerken wir, dass $\Phi(\mathbb{A}^n) \subset \mathbb{A}^m$ impliziert, dass für jedes $x \in \mathbb{R}^{n+1}$ die letzte Koordinate von $\Phi(x)$ gleich der letzten Koordinate von x ist. Damit ist aber $\mathrm{Ker}(\Phi) \subset \mathbb{R}^n$ und somit $\mathrm{Ker}(\Phi) = \mathrm{Ker}(\varphi)$. Daraus folgt die Behauptung leicht. □

5.5. Der Dimensionssatz für affine Teilräume. Die etwas mühsamen Überlegungen über den Zusammenhang von affinen Teilräumen und Teilräumen erlauben uns nun den Dimensionssatz für affine Teilräume sehr einfach zu beweisen. Der Satz setzt für zwei affine Teilräume $A, B \subset \mathbb{A}^n$ die Dimensionen von A, B, $A \cap B$, und dem Analogon $\langle A \cup B \rangle_{\text{aff}}$ der Summe von Teilräumen in Beziehung. Die Situation ist etwas komplizierter als im Fall von gewöhnlichen Teilräumen. Das Problem ist, dass affine Teilräume auch leeren Schnitt haben können (und dann macht die Dimension des Schnitts keinen Sinn). Das kann einerseits bei *parallelen* affinen Teilräumen passieren, wenn also $\overrightarrow{A} = \overrightarrow{B}$ gilt, oder allgemeiner einer der modellierenden Teilräume im anderen enthalten ist. Das Beispiel windschiefer Geraden in $\mathbb{R}^3$ zeigt aber, dass das auch für nicht parallele affine Teilräume passieren kann.

Satz 5.5 (Dimensionsformel für affine Teilräume). *Seien $A, B \subset \mathbb{A}^n$ affine Teilräume und sei $C := \langle A \cup B \rangle_{\mathit{aff}}$. Dann gilt:*
(1) Ist $A \cap B \neq \emptyset$, dann ist $\dim(C) = \dim(A) + \dim(B) - \dim(A \cap B)$.
(2) Ist $A \cap B = \emptyset$, dann ist $\dim(C) = \dim(A) + \dim(B) - \dim(\overrightarrow{A} \cap \overrightarrow{B}) + 1$.

Beweis. Nach Proposition 5.4 können wir $A = V \cap \mathbb{A}^n$ und $B = W \cap \mathbb{A}^n$ für passende Teilräume $V, W \subset \mathbb{R}^{n+1}$ schreiben. Analog kann man C als Schnitt $\langle A \cup B \rangle \cap \mathbb{A}^n$ schreiben. Nach Konstruktion ist gilt aber $V, W \subset U$ und damit $V + W \subset U$. Umgekehrt ist $(V + W) \cap \mathbb{A}^n$ ein affiner Teilraum, der A und B enthält, also folgern wir $C = (V + W) \cap \mathbb{A}^n$.

Somit gilt aber nach dem Dimensionssatz 4.8

$$\begin{aligned}\dim(C) &= \dim(V + W) - 1 = \dim(V) + \dim(W) - 1 - \dim(V \cap W) \\ &= \dim(A) + \dim(B) - (\dim(V \cap W) - 1).\end{aligned}$$

Nun ist aber offensichtlich $A \cap B = (V \cap W) \cap \mathbb{A}^n$. Ist dieser Durchschnitt nichtleer, dann ist er ein affiner Teilraum der Dimension $\dim(V \cap W) - 1$ und (1) folgt. Ist der Durchschnitt $A \cap B$ leer, dann ist $V \cap W \subset \mathbb{R}^n$ (würde der Schnitte einen Punkt mit letzter Koordinate $\neq 0$ enthalten, dann auch einen Punkt mit letzter Koordinate 1). Damit ist aber $V \cap W = (V \cap \mathbb{R}^n) \cap (W \cap \mathbb{R}^n) = \overrightarrow{A} \cap \overrightarrow{B}$. □

Beispiel 5.5. (1) Betrachten wir affine Teilräume von $\mathbb{A}^2$, so ist nur der Fall $\dim(A) = \dim(B) = 1$, $A \neq B$, also der Fall zweier verschiedener affiner Geraden, interessant. Setzen wir wieder $C = \langle A \cup B\rangle_{\text{aff}}$. Ist $A \cap B = \emptyset$, dann folgt aus Teil (2) des Satzes wegen $\dim(C) \leq 2$ sofort $\dim(\overrightarrow{A} \cap \overrightarrow{B}) \geq 1$. Somit ist aber nur der Fall $\overrightarrow{A} = \overrightarrow{B}$ möglich, also der Falle verschiedener paralleler Geraden, und dann gilt automatisch $\dim(C) = 2$, also erzeugen diese Geraden die ganze Ebene. Insbesondere haben je zwei nicht parallele Gerade in der Ebene mindestens einen Schnittpunkt. Ist $A \cap B \neq \emptyset$, dann gilt $\dim(A \cap B) = 0$ wegen $A \neq B$, also $A \cap B = \{x_0\}$ für einen Punkt $x_0 \in \mathbb{A}^2$. Somit schneiden sich zwei nicht parallele Geraden in genau einem Punkt, und nach Teil (1) des Satzes ist $\dim(C) = 2$, also erzeugen die beiden Geraden wieder die ganze Ebene.

(2) Betrachten wir den Fall von zwei verschiedenen affinen Geraden in $\mathbb{A}^3$, dann gibt es neben den beiden Fällen die wir schon aus der Ebene kennen noch den Fall $A \cap B = \emptyset$, $\dim(\overrightarrow{A} \cap \overrightarrow{B}) = 0$. Somit sind die Geraden nicht parallel und haben trotzdem keinen Schnittpunkt, also sind sie windschief. Teil (2) des Satzes sagt dann, dass $C = \mathbb{A}^3$ gilt.

Ist $A \subset \mathbb{A}^3$ eine affine Ebene und $B \subset \mathbb{A}^3$ eine affine Gerade mit $B \not\subset A$, dann gilt entweder $A \cap B = \emptyset$ oder $\dim(A \cap B) = 0$. Im ersten Fall folgt aus Teil (2) des Satzes $\overrightarrow{B} \subset \overrightarrow{A}$, also ist B parallel zu A und $C = \mathbb{A}^3$. Im zweiten Fall erhalten wir einen eindeutigen Schnittpunkt und $C = \mathbb{A}^3$.

Sind schließlich $A, B \subset \mathbb{A}^3$ verschiedene affine Ebenen, dann muss immer $C = \mathbb{A}^3$ gelten. Ist $A \cap B = \emptyset$, dann ist nach Teil (2) des Satzes $\dim(\overrightarrow{A} \cap \overrightarrow{B}) = 2$, also $\overrightarrow{A} = \overrightarrow{B}$ also sind die Ebenen parallel. Für $A \cap B \neq \emptyset$ muss $\dim(A \cap B) = 1$ gelten, also schneiden sich zwei nicht parallele affine Ebenen in $\mathbb{A}^3$ genau in einer affinen Geraden.

5.6. Kollinearität und Teilverhältnis. Im weitesten Sinn ist affine Geometrie das Studium von Eigenschaften von Teilmengen von $\mathbb{A}^n$, die invariant unter affinen Isomorphismen sind. Ein affiner Teilraum bestimmter Dimension zu sein ist natürlich offensichtlich eine solche Eigenschaft. Auch die Tatsache, ob zwei affine Teilräume parallel sind oder nicht ist eine solche Eigenschaft. Ist nämlich $f : \mathbb{A}^n \to \mathbb{A}^n$ ein affiner Isomorphismus mit zugehöriger linearer Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, dann ist für einen affinen Teilraum $A \subset \mathbb{A}^n$ natürlich $\overrightarrow{f(A)} = \varphi(\overrightarrow{A})$ und damit überträgt sich die Eigenschaft $\overrightarrow{A} \subset \overrightarrow{B}$ auf die Bilder.

Allgemein ist es aber gar nicht so leicht, solche Eigenschaften zu finden. Einem einzelnen Punkt $a \in \mathbb{A}^n$ kann man sicher nichts besonderes ansehen. Sind nämlich $a, b \in \mathbb{A}^n$ beliebige Punkte, dann definiert nach Proposition 5.3 die Funktion $f(x) = b + \overrightarrow{ax}$ einen affinen Isomorphismus $f : \mathbb{A}^n \to \mathbb{A}^n$ (mit $\varphi = \mathrm{id}$), der $f(a) = b$ erfüllt. Somit sind alle Punkte von $\mathbb{A}^n$ "gleichberechtigt", was ja von Anfang an erwünscht war. Die einzige geometrische Eigenschaft die eine Menge von zwei Punkten haben kann, ist ob die beiden Punkte gleich oder verschieden sind. Sind nämlich $a_1 \neq a_2$ und $b_1 \neq b_2$ zwei Paare von verschiedenen Punkten, dann sind $v := \overrightarrow{a_1a_2}$ und $w := \overrightarrow{b_1b_2}$ Vektoren in $\mathbb{R}^n$ die beide $\neq 0$ sind. Damit gibt es aber einen linearen Isomorphismus $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, der $\varphi(v) = w$ erfüllt. (Die Mengen $\{v\}$ und $\{w\}$ sind linear unabhängig, können also zu Basen erweitert werden, und dann gibt es einen eindeutigen linearen Isomorphismus, der die erste Basis auf die zweite Basis abbildet.) Dann erfüllt aber $f(a) := b_1 + \varphi(\overrightarrow{a_1a})$ gerade $f(a_1) = b_1$ und $f(a_2) = b_1 + \varphi(v) = b_2$.

Für drei Punkte wird es aber interessant. Die Punkte können entweder *kollinear* sein, also auf einer affinen Gerade liegen, oder nicht. Falls drei Punkte nicht kollinear sind, dann sagt man, sie befinden sich in *allgemeiner Lage*. Wie oben zeigt man, dass es für zwei Mengen $\{a_1, a_2, a_3\}$ und $\{a'_1, a'_2, a'_3\}$ von je drei Punkten in allgemeiner Lage einen affinen Isomorphismus gibt, der $f(a_i) = a'_i$ für $i = 1, 2, 3$ erfüllt. Damit haben aber drei Punkte in allgemeiner Lage wieder keine geometrischen Eigenschaften. Insbesondere sind in der affinen Geometrie je zwei Dreiecke isomorph, es gibt also keine affine Geometrie von Dreiecken.

Drei verschiedene kollineare Punkte bestimmen aber eine geometrische Größe, nämlich das sogenannte Teilverhältnis, das eine entscheidende affine Invariante darstellt. Sind nämlich $a, b \in \mathbb{A}^n$ zwei verschiedene Punkte, dann bestimmen sie eine eindeutige affine Gerade $\mathcal{G} = \mathcal{G}_{ab}$, die durch $\mathcal{G} = \{a + t\overrightarrow{ab} : t \in \mathbb{R}\}$ gegeben ist. Ist $c \in \mathbb{A}^n$ ein weiterer Punkt, sodass a, b und c kollinear sind, dann gibt es also ein eindeutiges (weil $\overrightarrow{ab} \neq 0$ gilt) Element $t \in \mathbb{R}$, sodass $c = a + t\overrightarrow{ab}$, also $\overrightarrow{ac} = t\overrightarrow{ab}$ gilt. Diese Zahl heißt das *Teilverhältnis* von $\{c, b, a\}$ und wird mit $TV(c, b, a)$ oder mit $\frac{\overrightarrow{ac}}{\overrightarrow{ab}}$ oder mit $\frac{c-a}{b-a}$ bezeichnet.

Ist $f : \mathbb{A}^n \to \mathbb{A}^n$ ein affiner Isomorphismus und $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ der zugehörige lineare Isomorphismus, dann ist $\overrightarrow{f(a)f(b)} = \varphi(\overrightarrow{ab})$ und damit

$$f(c) = f(a) + \varphi(\overrightarrow{ac}) = f(a) + t\varphi(\overrightarrow{ab}),$$

also folgt sofort, dass $\frac{f(c)-f(a)}{f(b)-f(a)} = \frac{c-a}{b-a}$ gilt. Damit ist das Teilverhältnis eine affine Invariante, also eine geometrische Größe. Aus der Definition ist offensichtlich, dass für kollineare Punkte a, b, c mit $a \neq b$ folgendes gilt: c liegt genau dann zwischen a und b, wenn $0 \leq \frac{c-a}{b-a} \leq 1$ gilt, das Teilverhältnis ist genau für $c = a$ Null und für $c = b$ Eins. $\frac{c-a}{b-a} < 0$ gilt genau dann, wenn c auf der anderen Seite von a als b liegt, und $\frac{c-a}{b-a} > 1$ ist äquivalent dazu, dass c auf der anderen Seite von b als a liegt.

Die Bruchschreibweise hat einerseits den Vorteil, dass man sich die Abhängigkeit von der Reihenfolge der Punkte nicht merken muss. Andererseits zeigt sie aber auch gut, wie man mit Teilverhältnissen rechnen kann. Sind nämlich a, b, c verschiedene kollineare Punkte in A und ist d ein weiterer Punkt, der auf der Geraden durch a, b und c liegt, dann ist nach Definition $\overrightarrow{ad} = \frac{d-a}{c-a}\overrightarrow{ac}$ und $\overrightarrow{ac} = \frac{c-a}{b-a}\overrightarrow{ab}$. Damit folgt aber $\overrightarrow{ad} = \frac{d-a}{c-a}\frac{c-a}{b-a}\overrightarrow{ab}$ und somit wiederum nach Definition $\frac{d-a}{b-a} = \frac{d-a}{c-a}\frac{c-a}{b-a}$. Insbesondere folgt daraus, dass $\frac{b-a}{c-a}$ tatsächlich der Reziprokwert von $\frac{c-a}{b-a}$ ist, weil offensichtlich $\frac{b-a}{b-a} = 1$ gilt.

5.7. Der Strahlensatz. Als erstes Beispiel für einen grundlegenden Satz der affinen Geometrie betrachten wir den Strahlensatz. Wir benutzen diesen Satz auch, um eine der fundamentalen Beweismethoden der affinen Geometrie zu illustrieren. Die Idee dieser Methode ist, dass man sich zunächst überlegt, dass die Aussage des Satzes "invariant unter affinen Isomorphismen" ist. Das bedeutet, dass der Satz einer gegeben Situation genau dann gilt, wenn er nach Anwenden eines affinen Isomorphismus gültig ist. Dann kann man einen affinen Isomorphismus konstruieren, der eine gegebene Situation auf einen Modellfall abbildet, in dem die Aussage leicht zu beweisen ist.

Satz 5.7 (Strahlensatz). *Seien $\mathcal{G}, \mathcal{G}' \subset \mathbb{A}^n$ zwei affine Geraden, die sich entweder schneiden, oder parallel sind. Seien $p, q \in \mathcal{G}$ und $p', q' \in \mathcal{G}'$ jeweils verschiedene Punkte, die im Fall schneidender Geraden auch alle verschieden vom Schnittpunkt o sind. Sei $\mathcal{G}_{pp'}$ die affine Gerade durch p und p' und $\mathcal{G}_{qq'}$ die affine Gerade durch q und q'. Dann gilt:*

(1) Ist $\mathcal{G} \cap \mathcal{G}' \neq \emptyset$, dann gilt $\mathcal{G}_{pp'}$ und $\mathcal{G}_{qq'}$ sind parallel $\iff \frac{q-o}{p-o} = \frac{q'-o}{p'-o} \iff \frac{q-p}{o-p} = \frac{q'-p'}{o-p'}$.

(2) Sind $\mathcal{G}$ und $\mathcal{G}'$ parallel, dann sind $\mathcal{G}_{pp'}$ und $\mathcal{G}_{qq'}$ genau dann parallel, wenn $\overrightarrow{pp'} = \overrightarrow{qq'}$ gilt.

Beweis. Ist $f : \mathbb{A}^n \to \mathbb{A}^n$ ein affiner Isomorphismus, dann sind natürlich $f(\mathcal{G})$ und $f(\mathcal{G}')$ affine Geraden und $f(p), f(q) \in f(\mathcal{G})$ und $f(p'), f(q') \in f(\mathcal{G}')$ sind verschiedene Punkte, und im Fall schneidender Geraden auch alle verschieden vom Schnittpunkt $f(o)$. Natürlich ist die affine Gerade durch $f(p)$ und $f(p')$ genau $f(\mathcal{G}_{pp'})$ und analog für die zweite Gerade. Damit ist die Situation des Satzes invariant unter affinen Isomorphismen. Nachdem affine Isomorphismen verträglich mit Parallelität, Teilverhältnissen und Verbindungsvektoren sind, ist auch die Aussage des Satzes affin invariant.

(1) Im Fall schneidender Geraden sind die Richtungsvektoren von $\mathcal{G}$ und $\mathcal{G}'$ linear unabhängig, also kann man die Verbindungsvektoren $\overrightarrow{op}$ und $\overrightarrow{op'}$ durch einen linearen Isomorphismus auf e_1 und e_2 transformieren. Dann fügen wir eine passende Translation dazu, um auch noch $f(o) = e_{n+1}$ zu erreichen. Dann ist $f(\mathcal{G}) = \{e_{n+1} + te_1 : t \in \mathbb{R}\}$ und $f(\mathcal{G}') = \{e_{n+1} + se_2 : s \in \mathbb{R}\}$ und $f(p) = e_1 + e_{n+1}$, $f(p') = e_2 + e_{n+1}$, also $\overrightarrow{f(p)f(p')} = e_2 - e_1$. Da $f(q)$ auf $f(\mathcal{G})$ liegt, muss $f(q) = e_{n+1} + \frac{q-o}{p-o}e_1$ gelten und analog ist $f(q') = e_{n+1} + \frac{q'-o}{p'-o}e_2$ und somit $\overrightarrow{f(q)f(q')} = \frac{q'-o}{p'-o}e_2 - \frac{q-o}{p-o}e_1$. Nun sind die Geraden $\mathcal{G}_{f(p)f(p')}$ und $\mathcal{G}_{f(q)f(q')}$ natürlich genau dann parallel, wenn die beiden Vektoren $\overrightarrow{f(p)f(p')}$ und $\overrightarrow{f(q)f(q')}$ proportional zueinander sind, und das ist offensichtlich äquivalent zu $\frac{q'-o}{p'-o} = \frac{q-o}{p-o}$. Bezeichnen wir diese Zahl mit λ, dann folgt aus $\overrightarrow{oq} = \lambda\overrightarrow{op}$ sofort, dass $\overrightarrow{pq} = \overrightarrow{po} + \overrightarrow{oq} = (-1-\lambda)\overrightarrow{po}$ und umgekehrt. Das geht analog für die anderen Punkte, also folgt die Behauptung.

(2) In diesem Fall wählt man f so, dass $f(p) = e_{n+1}$, $f(q) = e_{n+1} + e_1$ und $f(p') = e_{n+1} + e_2$ gilt. Da $f(\mathcal{G}')$ parallel zu $f(\mathcal{G})$ ist, muss $f(q') = e_{n+1} + e_2 + te_1$ für eine reelle Zahl t sein. Nun ist aber $\overrightarrow{f(p)f(p')} = e_2$ und $\overrightarrow{f(q)f(q')} = e_2 + (t-1)e_1$. Die Geraden $\mathcal{G}_{f(p)f(p')}$ und $\mathcal{G}_{f(q)f(q')}$ sind genau dann parallel, wenn diese beiden Vektoren proportional sind. Das ist aber nur für $t = 1$ möglich und dann sind die beiden Vektoren gleich. □

5.8. Die Sätze von Desargues und Pappos–Pascal. Unter den vielen Beispielen klassischer Sätze der affinen Geometrie greifen wir noch zwei sogenannte Schließungsätze heraus.

Proposition 5.8 (Satz von Desargues). *Seien $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3 \subset \mathbb{A}^2$ paarweise verschiedene affine Geraden, die sich entweder in einem Punkt o schneiden, oder alle parallel sind. Für $i = 1, 2, 3$ wähle jeweils Punkte $p_i, q_i \in \mathcal{G}_i$, die im Fall schneidender Gerader alle verschieden von o sind. Dann gilt: Sind $\mathcal{G}_{p_1p_2}$ und $\mathcal{G}_{q_1q_2}$ sowie $\mathcal{G}_{p_1p_3}$ und $\mathcal{G}_{q_1q_3}$ jeweils parallel, dann sind auch $\mathcal{G}_{p_2p_3}$ und $\mathcal{G}_{q_2q_3}$ parallel.*

Beweis. Betrachten wir zunächst den Fall schneidender Geraden. Dann ist nach dem Strahlensatz 5.7(1) die Voraussetzung äquivalent zu $\frac{q_1-o}{p_1-o} = \frac{q_2-o}{p_2-o}$ sowie zu $\frac{q_1-o}{p_1-o} = \frac{q_3-o}{p_3-o}$. Damit ist aber auch $\frac{q_2-o}{p_2-o} = \frac{q_3-o}{p_3-o}$, und die Behauptung folgt aus dem Strahlensatz 5.7(1).

Im Fall von parallelen Geraden liefert die Voraussetzung nach 5.7(2) $\overrightarrow{p_1p_2} = \overrightarrow{q_1q_2}$ und $\overrightarrow{p_1p_3} = \overrightarrow{q_1q_3}$. Subtrahieren der ersten Gleichungen von der zweiten liefert dann $\overrightarrow{p_2p_3} = \overrightarrow{q_2q_3}$ und damit die Behauptung. □

Die Bezeichnung "Schließungssatz" ist folgendermaßen zu verstehen: Nehmen wir an, das wir die Punkte p_1, p_2, p_3 und q_1 gewählt haben. Dann können wir die Geraden $\mathcal{G}_{p_1p_2}$ und $\mathcal{G}_{p_1p_3}$ durch parallel durch q_1 verschieben und q_2 und q_3 als Schnittpunkte. Der Satz besagt dann, dass wenn man die Gerade $\mathcal{G}_{p_2p_3}$ parallel durch einen der Punkte

q_2 oder q_3 verschiebt, auch der andere der beiden Punkte auf der Geraden liegt, also insgesamt eine geschlossene Figur entsteht.

Satz 5.8 (Satz von Pappos–Pascal). *Seien $\mathcal{G}, \mathcal{G}' \subset \mathbb{A}^2$ verschiedene affine Geraden. Seien $p_1, p_2, p_3 \in \mathcal{G}$ und $p_1', p_2', p_3' \in \mathcal{G}'$ jeweils paarweise verschiedene Punkte, die im Fall schneidender Geraden auch noch alle verschieden vom Schnittpunkt sind. Dann gilt: Sind $\mathcal{G}_{p_1p_3'}$ und $\mathcal{G}_{p_3p_1'}$ sowie $\mathcal{G}_{p_1p_2'}$ und $\mathcal{G}_{p_2p_1'}$ jeweils parallel, dann sind auch $\mathcal{G}_{p_2p_3'}$ und $\mathcal{G}_{p_3p_2'}$ parallel.*

Beweis. Betrachten wir zunächst den Fall von schneidenden Geraden, wobei wir den Schnittpunkt mit o bezeichnen. Nach dem Strahlensatz 5.7(1) ist die Voraussetzung über Parallelität äquivalent zu $\frac{p_3-o}{p_1-o} = \frac{p_1'-o}{p_3'-o}$ und $\frac{p_2-o}{p_1-o} = \frac{p_1'-o}{p_2'-o}$. Die zweite Gleichung kann man natürlich auch als $\frac{p_1-o}{p_2-o} = \frac{p_2'-o}{p_1'-o}$ schreiben. Multiplizieren wir die linke Seite dieser Gleichung mit der linken Seite der ersten Gleichung, dann erhalten wir $\frac{p_3-o}{p_1-o}\frac{p_1-o}{p_2-o} = \frac{p_3-o}{p_2-o}$ (siehe 5.6), und für die rechten Seiten ergibt das Produkt $\frac{p_1'-o}{p_3'-o}\frac{p_2'-o}{p_1'-o} = \frac{p_2'-o}{p_3'-o}$. Damit erhalten wir $\frac{p_3-o}{p_2-o} = \frac{p_2'-o}{p_3'-o}$, was nach dem Strahlensatz 5.7(1) wiederum impliziert, dass $\mathcal{G}_{p_3p_2'}$ und $\mathcal{G}_{p_2p_3'}$ parallel sind.

Sind $\mathcal{G}$ und $\mathcal{G}'$ parallel, dann ist mit Hilfe des Strahlensatzes 5.7(2) die Voraussetzung äquivalent zu $\overrightarrow{p_1p_3'} = \overrightarrow{p_3p_1'}$ und $\overrightarrow{p_1p_2'} = \overrightarrow{p_2p_1'}$. Aus der zweiten Gleichung folgt sofort $p_1' = p_2 + p_2' - p_1$, also $\overrightarrow{p_2p_1'} = \overrightarrow{p_1'p_2'}$. Addiert man diese Gleichung zur ersten Gleichung von oben, dann folgt $\overrightarrow{p_2p_3'} = \overrightarrow{p_3p_2'}$ was wiederum nach dem Strahlensatz 5.7(2) die Behauptung impliziert. □

Die Interpretation als Schließungssatz ist ähnlich wie zuvor: Haben wir p_1 und die p_i' gewählt, dann verschieben wir die Gerade $\mathcal{G}_{p_1p_2'}$ parallel durch p_1' und definieren den Schnittpunkt mit $\mathcal{G}$ als p_2 und analog konstruieren wir $p_3 \in \mathcal{G}$. Der Satz sagt nun, dass man eine geschlossene Figur erhält, wenn man $\mathcal{G}_{p_2p_3'}$ parallel durch p_2' verschiebt, sich als Schnittpunkt mit $\mathcal{G}$ also gerade p_3 ergibt.

5.9. Bemerkungen zur projektiven Geometrie. Wir haben in 5.4 den affinen Raum $\mathbb{A}^n$ mit der Menge aller eindimensionalen Teilräume von $\mathbb{R}^{n+1}$ identifiziert, die nicht in $\mathbb{R}^n$ enthalten sind. Dann konnten wir affine Abbildungen als Einschränkungen von passenden linearen Abbildungen realisieren.

Betrachten wir zunächst den Fall von $\mathbb{A}^1 \subset \mathbb{R}^2$. Hier identifizieren wir $\mathbb{A}^1$ mit der Menge aller 1–dimensionalen Teilräume von $\mathbb{R}^2$ mit Ausnahme des Teilraumes $\mathbb{R} \subset \mathbb{R}^2$. Wandert man in $\mathbb{A}^1$ (egal in welcher der beiden Richtungen) nach außen, dann werden die entsprechenden Geraden immer flacher, nähern sich also der "fehlenden" Geraden an. Nun definiert man die *projektive Gerade* $\mathbb{P}^1$ als die Menge aller eindimensionalen Teilräume in $\mathbb{R}^2$, und interpretiert sie als die affine Gerade $\mathbb{A}^1$, zu der man noch einen "Fernpunkt" hinzugefügt hat.

Betrachte man dann $\mathbb{A}^2$, so entspricht das allen eindimensionalen Teilräumen in $\mathbb{R}^3$, die nicht im Teilraum $\mathbb{R}^2$ liegen. Definiert man also $\mathbb{P}^2$ als die Menge aller eindimensionalen Teilräume in $\mathbb{R}^3$ so hat man zu $\mathbb{A}^2$ gerade noch eine projektive Ferngerade hinzugefügt, die gerade die verschiedenen Richtungen kodiert, wie man in der Ebene gegen unendlich gehen kann, wobei man auf einer Geraden immer in beiden Richtungen zum gleichen Fernpunkt gelangt.

Allgemein definiert man dann den projektiven Raum $\mathbb{P}^n$ als die Menge aller eindimensionalen Teilräume von $\mathbb{R}^{n+1}$. Ein $(k+1)$–dimensionalen Teilraum $V \subset \mathbb{R}^{n+1}$ bestimmt dann eine Teilmenge in $\mathbb{P}^n$ (nämlich jene Geraden, die in V enthalten sind) und diese Teilmengen sind genau die *projektiven Teilräume* der Dimension k. Die Situation ist einfacher als im affinen Fall. Zwei verschiedene projektive Gerade in $\mathbb{P}^2$ entsprechen gerade zwei verschiedenen zweidimensionalen Teilräumen in $\mathbb{R}^3$. Aus dem Dimensionssatz folgt sofort, dass der Durchschnitt solcher Teilräume immer eindimensional sein muss. Damit schneiden sich als je zwei verschiedene projektive Geraden in $\mathbb{P}^2$ in genau einem Punkt. Zwei verschiedene projektive Geraden in $\mathbb{P}^3$ entsprechen zwei verschiedenen zweidimensionalen Teilräumen $V, W \subset \mathbb{R}^4$. Aus der Dimensionsformel aus 4.8 folgt sofort, dass entweder $\dim(V \cap W) = 1$ und $\dim(V + W) = 3$ oder $\dim(V \cap W) = 0$ und $\dim(V + W) = 4$ gilt. Es gibt also genau die beiden Möglichkeiten, dass die Geraden in einer projektiven Ebene enthalten sind und einen eindeutigen Schnittpunkt haben, oder nicht in einer solchen Ebene liegen und disjunkt sind. (Es fallen also alle Sonderfälle von Parallelität weg.) Ein allgemeiner Dimensionssatz für projektive Teilräume folgt sofort aus dem entsprechenden Dimensionssatz 4.8 für Teilräume von $\mathbb{R}^{n+1}$.

Für einen linearen Isomorphismus $\Phi : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ und einen Teilraum $\ell \in \mathbb{P}^n$ (also eine Gerade durch 0) ist auch $\Phi(\ell)$ eine Gerade durch 0, also liefert Φ eine Abbildung $f : \mathbb{P}^n \to \mathbb{P}^n$. Abbildungen dieser Form heißen *projektive Transformationen*. Im weitesten Sinn beschreibt projektive Geometrie gerade Eigenschaften von Teilmengen von $\mathbb{P}^n$, die invariant unter projektiven Transformationen sind.

Wieder haben einzelne Punkte und Paare von Punkten keine solchen Eigenschaften. Für Tripel von Punkten gibt es nur die Unterscheidung zwischen "kollinear" und "in allgemeiner Lage", aber es gibt kein projektiv invariantes Teilverhältnis. Erst vier kollinearen Punkten kann man eine projektiv invariante Zahl zuordnen, nämlich das sogenannten *Doppelverhältnis*, das die entscheidende projektive Invariante darstellt.

Wichtiger als die projektive Geometrie in diesem Sinn ist, dass man gewissen Polynomen in sinnvoller Weise eine Nullstellenmenge in $\mathbb{P}^n$ zuordnen kann. Das Studium dieser Nullstellenmengen ist die *algebraische Geometrie*, die ein großes Teilgebiet der Mathematik darstellt.

Quotientenräume

In 4.9 haben wir den Begriff komplementärer Teilräume kennen gelernt. Zu einem Teilraum W in einem endlichdimensionalen $\mathbb{K}$–Vektorraum V gibt es im Allgemeinen viele mögliche Komplemente. Der "Quotientenraum von V nach W" repräsentiert in gewissem Sinn alle diese Komplemente in einem Vektorraum, der in natürlicher Weise aus V und W konstruiert wird, allerdings ist das Resultat kein Teilraum von V. Die Konstruktion stellt eines der Grundprinzipien der Algebra dar. Sie ist zwar formal eigentlich einfach, macht aber erfahrungsgemäß vielen AnfängerInnen Schwierigkeiten.

5.10. Der Begriff des Quotientenraumes. Betrachten wir als Beispiel die Ebene $W := \mathbb{R}^2 \subset \mathbb{R}^3 =: V$, also die Menge aller Punkte, deren letzte Koordinate Null ist. Aus Proposition 4.9 wissen wir, dass die komplementären Teilräume zu dieser Ebene genau die Geraden durch Null sind, die diese Ebene nur im Nullpunkt schneiden, die also nicht horizontal sind. Nun können wir die Ebene natürlich parallel verschieben und erhalten (in der Terminologie von 5.2) genau die affinen Teilräume von V mit modellierendem Teilraum W. Jeder Teilraum von V, der komplementär zu W ist, schneidet jede dieser parallel verschobenen Ebenen in genau einem Punkt. Die parallel verschobenen Ebenen bestehen aber genau aus all jenen Punkten, für die die letzte Koordinate einen fixen Wert hat. Sind E_1 und E_2 zwei solche Ebenen und $x_1 \in E_1$ und $x_2 \in E_2$ zwei beliebige Punkte, dann hängt die letzte Koordinate von $x_1 + x_2$ bzw. die parallele Ebene durch diesen Punkt nur von E_1 und E_2 und nicht von der Wahl von x_1 und x_2 ab. Genau so hängt für $r \in \mathbb{R}$ die letzte Koordinate von rx_1 bzw. die parallele Ebene durch diesen Punkt nur von E_1 und r ab, und nicht von der Wahl von x_1. Damit kann man auf der Menge aller zu W parallelen Ebenen eine Addition und eine Skalarmultiplikation definieren, die genau der Addition und Skalarmultiplikation auf einer beliebigen komplementären Gerade entspricht.

Ganz analog kann man den Quotientenraum allgemein definieren:

Definition 5.10. Sei V ein $\mathbb{K}$–Vektorraum und $W \subset V$ ein Teilraum.

(1) Für $v \in V$ definiere die *Nebenklasse von v bezüglich W* als

$$v + W := \{v + w : w \in W\} \subset V.$$

(2) Die Menge $V/W := \{v + W : v \in V\}$ aller Nebenklassen heißt der *Quotient von V nach W*.

(3) Die *kanonische Quotientenabbildung* ist die Funktion $\pi : V \to V/W$, die jedem Element von V seine Nebenklasse zuordnet, also durch $\pi(v) := v + W$ definiert ist.

Nach Definition ist die Nebenklasse $v + W$ gerade der affine Teilraum von V, den man erhält indem man W parallel durch v verschiebt.

Satz 5.10. *Sei V ein $\mathbb{K}$–Vektorraum und $W \subset V$ ein Teilraum. Dann gilt:*

(1) Für zwei Elemente $v_1, v_2 \in V$ gilt $v_1 + W = v_2 + W$ genau dann, wenn $v_1 - v_2 \in W$ gilt. Insbesondere sind zwei Nebenklassen entweder gleich oder disjunkt.

(2) Man kann den Quotienten V/W in natürlicher Weise zu einem $\mathbb{K}$–Vektorraum machen. Diese Vektorraumstruktur ist dadurch charakterisiert, dass die kanonische Quotientenabbildung $\pi : V \to V/W$ eine lineare Abbildung ist.

(3) Ist V endlichdimensional, dann ist auch V/W endlichdimensional und hat Dimension $\dim(V/W) = \dim(V) - \dim(W)$.

(4) Ein Teilraum $W' \subset V$ ist genau dann komplementär zu W, wenn die Einschränkung $\pi|_{W'} : W' \to V/W$ ein linearer Isomorphismus ist.

Beweis. (1) Ist $v_1 - v_2 \in W$, dann ist $v_1 = v_2 + (v_1 - v_2) \in v_2 + W$ und damit auch $v_1 + W \subset v_2 + W$. Da W ein Teilraum ist, folgt aber aus $v_1 - v_2 \in W$ auch $v_2 - v_1 \in W$ und damit die umgekehrte Inklusion.

Umgekehrt zeigen wir, dass aus $(v_1 + W) \cap (v_2 + W) \neq \emptyset$ schon $v_1 - v_2 \in W$ folgt, was den Beweise von (1) vervollständigt. Ist $v \in (v_1 + W) \cap (v_2 + W)$, dann gibt es nach Definition Elemente $w_1, w_2 \in W$, sodass $v_1 + w_1 = v = v_2 + w_2$ gilt. Damit ist aber $v_1 - v_2 = w_2 - w_1$ und das liegt in W, weil W ein Teilraum ist.

(2) Wir definieren eine Addition auf V/W durch $(v_1+W)+(v_2+W) := (v_1+v_2)+W$ und eine Skalarmultiplikation durch $r(v+W) := (rv)+W$. Wir behaupten, dass das nur von den Nebenklassen und nicht von der Wahl von v_1 uns v_2 abhängt. Ist $\tilde{v}_i+W = v_i+W$ für $i = 1, 2$, dann gilt $\tilde{v}_1 - v_1 \in W$ und $\tilde{v}_2 - v_2 \in W$ nach Teil (1). Damit ist aber auch

$$(\tilde{v}_1 + \tilde{v}_2) - (v_1 + v_2) = (\tilde{v}_1 - v_1) + (\tilde{v}_2 - v_2) \in W,$$

also $(\tilde{v}_1 + \tilde{v}_2) + W = (v_1 + v_2) + W$. Analog ist $r\tilde{v}_1 - rv_1 = r(\tilde{v}_1 - v_1) \in W$, also $(r\tilde{v}_1) + W = (rv_1) + W$.

Nun folgen die Vektorraumeigenschaften (V1)–(V8) aus 2.2 für V/W direkt aus den entsprechenden Eigenschaften für V. Insbesondere ist das Nullelement in V/W gerade die Nebenklasse $0 + W = W$ und $-(v + W) = (-v) + W$. Die Definitionen von Addition und Skalarmultiplikation sind offensichtlich äquivalent dazu, dass $\pi(v_1 + v_2) = \pi(v_1) + \pi(v_2)$ und $\pi(rv_1) = r\pi(v_1)$ gelten, und das vervollständigt den Beweis von (2).

(3) Für $v \in V$ gilt $v \in \mathrm{Ker}(\pi)$ genau dann, wenn $v + W = 0 + W$ gilt, also nach Teil (1) genau dann wenn $v \in W$ gilt. Damit ist $\mathrm{Ker}(\pi) = W$ und offensichtlich ist

$\operatorname{Im}(\pi) = V/W$. Damit folgt das Resultat sofort aus dem Dimensionssatz 4.11 für lineare Abbildungen. Alternativ kann man auch eine Basis $v_1, \dots, v_k$ für W wählen und zu einer Basis $v_1, \dots, v_n$ für V erweitern und dann direkt zeigen, dass die Vektoren $v_{k+1} + W$, $\dots$, $v_n + W$ eine Basis für V/W bilden. (Das ist eine gute Übung.)

(4) Für jeden Teilraum $W' \subset W$ ist die Einschränkung $\pi|_{W'} : W' \to V/W$ natürlich eine lineare Abbildung, also müssen wir nur zeigen, dass $\pi|_{W'}$ genau dann bijektiv ist, wenn W' ein Komplement zu W ist. Aus dem Beweis von (3) wissen wir, dass $\operatorname{Ker}(\pi) = W$ gilt. Damit ist $\operatorname{Ker}(\pi|_{W'}) = W \cap W'$, also Injektivität von π äquivalent zu $W \cap W' = \{0\}$. Ist das erfüllt, dann ist π genau dann surjektiv, wenn $\dim(W') = \dim(V/W) = \dim(V) - \dim(W)$ gilt, was nach Proposition 4.9 den Beweis vervollständigt. □

5.11. Die universelle Eigenschaft von Quotienten. Durch die kanonische Quotientenabbildung $\pi : V \to V/W$ kann man jeder linearen Abbildung $\varphi : V/W \to Z$ in einen beliebigen $\mathbb{K}$–Vektorraum Z die lineare Abbildung $\varphi \circ \pi : V \to Z$ zuordnen. Da π surjektiv ist, ist die Abbildung φ durch $\varphi \circ \pi$ eindeutig bestimmt. Man kann also lineare Abbildung von V/W nach Z als "spezielle lineare Abbildungen" von V nach Z betrachten. Welche linearen Abbildungen man in dieser Weise erhält, können wir nun auch sofort verstehen:

Proposition 5.11. *Sei V ein $\mathbb{K}$–Vektorraum, $W \subset V$ ein Teilraum, V/W der Quotient von V nach W und $\pi : V \to V/W$ die kanonische Quotientenabbildung. Weiters sei Z ein beliebiger Vektorraum über $\mathbb{K}$.*

(1) Für jede lineare Abbildung $\varphi : V/W \to Z$ ist $W \subset \operatorname{Ker}(\varphi \circ \pi)$.

(2) Ist $f : V \to Z$ eine beliebige lineare Abbildung, sodass $W \subset \operatorname{Ker}(f)$ gilt, dann gibt es eine eindeutige lineare Abbildung $\underline{f} : V/W \to Z$, sodass $f = \underline{f} \circ \pi$ gilt.

Beweis. (1) Für $w \in W$ gilt $\pi(w) = 0$, also auch $\varphi(\pi(w)) = 0$, weil φ linear ist.

(2) Seien $v, \tilde{v} \in V$ mit $v + W = \tilde{v} + W$. Dann gilt $v - \tilde{v} \in W$ nach Teil (1) von Satz 5.10. Da $W \subset \operatorname{Ker}(f)$ gilt, erhalten wir $0 = f(v - \tilde{v}) = f(v) - f(\tilde{v})$, also ist $f(v) = f(\tilde{v})$. Damit können wir aber $\underline{f} : V/W \to Z$ durch $\underline{f}(v + W) := f(v)$ definieren. Aus dieser Definition und der Linearität von f folgt sofort, dass $\underline{f}$ linear ist, und offensichtlich gilt $f(v) = \underline{f}(\pi(v))$, also $f = \underline{f} \circ \pi$. □

Daraus erhalten wir sofort den sogenannten *ersten Isomorphiesatz* für Vektorräume. Sobald man die Definitionen verdaut hat, ist er eigentlich eine ziemlich banale Beobachtung.

Korollar 5.11. *Sei $f : V \to W$ eine lineare Abbildung zwischen $\mathbb{K}$–Vektorräumen. Dann induziert f einen linearen Isomorphismus $V/\operatorname{Ker}(f) \cong \operatorname{Im}(f)$.*

Beweis. Nach der Proposition induziert f eine lineare Abbildung $\underline{f} : V/\operatorname{Ker}(f) \to W$, die $f = \underline{f} \circ \pi$ für die kanonische Quotientenabbildung $\pi : V \to V/\operatorname{Ker}(f)$ erfüllt. Das bedeutet gerade, dass $\underline{f}(v + W) = f(v)$ gilt, also ist $\operatorname{Im}(\underline{f}) = \operatorname{Im}(f)$. Andererseits zeigt das auch, dass $v + W \in \operatorname{Ker}(\underline{f})$ nur für $f(v) = 0$, also $v \in \operatorname{Ker}(f)$, also $v + W = 0 + W$ gilt. Damit ist aber $\underline{f}$ injektiv, definiert also einen linearen Isomorphismus $V/\operatorname{Ker}(f) \to \operatorname{Im}(f)$. □

Dualräume und Dualität

Auch hier handelt es sich um ein Thema, das formal eigentlich gar nicht so kompliziert ist, aber wegen seiner (scheinbaren?) Abstraktheit vielen AnfängerInnen Schwierigkeiten macht.

5.12. Dualraum und duale Abbildung. Wir haben schon festgestellt, dass für Vektorräume V und W über $\mathbb{K}$ auch die Menge $L(V,W)$ aller linearen Abbildungen von V nach W mit den punktweisen Operationen ein Vektorraum über $\mathbb{K}$ ist. Nun kann man versuchen, zu einem gegebenen Vektorraum V weitere Vektorräume als Räume von linearen Abbildung von oder nach speziellen $\mathbb{K}$–Vektorräumen zu betrachten. Einfache Kandidaten für solche speziellen Vektorräume sind natürlich $\{0\}$ und $\mathbb{K}$. Nun gibt es aber weder interessante lineare Abbildungen $\{0\} \to V$ noch $V \to \{0\}$ und man verifiziert sofort, dass $f \mapsto f(1)$ einen linearen Isomorphismus $L(\mathbb{K},V) \cong V$ induziert. Bleibt nur noch der Raum $L(V,\mathbb{K})$ zu betrachten, der sehr interessant ist.

Im Fall des Vektorraumes $\mathbb{K}^n$ können wir lineare Abbildungen $\mathbb{K}^n \to \mathbb{K}$ einfach durch $(1 \times n)$–Matrizen, also durch Zeilenvektoren der Länge n beschreiben. In dieser Sprache ist der Wert einer Abbildung y auf einem Spaltenvektor x dadurch gegeben, dass man x als $(n \times 1)$–Matrix interpretiert und das Matrizenprodukt yx bildet, dass dann eine (1×1)–Matrix, also einfach ein Element von $\mathbb{K}$ ist.

Sei nun $A \in M_{m,n}(\mathbb{K})$ eine Matrix. Dann können wir A nicht nur als lineare Abbildung $\mathbb{K}^n \to \mathbb{K}^m$ interpretieren, die durch $x \mapsto Ax$ gegeben ist, sondern auch für einen Zeilenvektor $y \in M_{1,m}$ der Länge m das Produkt $yA \in M_{1,n}$ bilden. Die Matrix A ordnet also nicht nur jedem Vektor in $\mathbb{K}^n$ einen Vektor in $\mathbb{K}^m$ zu, sondern gleichzeitig auch jeder linearen Abbildung $\mathbb{K}^m \to \mathbb{K}$ eine lineare Abbildung $\mathbb{K}^n \to \mathbb{K}$. Bezeichnen wir mit f die lineare Abbildung $f(x) = Ax$ und mit f^* die Abbildung $f^*(y) = yA$, dann impliziert die Assoziativität der Matrizenmultiplikation sofort, dass $y(Ax) = (yA)x$ also $(f^*(y))(x) = y(f(x)) = (y \circ f)(x)$ gilt. Das macht aber für allgemeine Vektorräume Sinn.

Erinnern wir uns auch noch daran, dass man für endlichdimensionales V und eine Basis $\{v_1, \dots, v_n\}$ für V, lineare Abbildungen von V nach $\mathbb{K}$ leicht beschreiben kann: Nach Proposition 4.2 ist jede lineare Abbildung $\varphi : V \to \mathbb{K}$ eindeutig durch die Elemente $\varphi(v_1), \dots, \varphi(v_n) \in \mathbb{K}$ bestimmt, und nach Satz 4.10 gibt es für jede Wahl dieser Element auch tatsächlich eine lineare Abbildung φ.

Definition 5.12. (1) Ein *lineares Funktional* auf einem $\mathbb{K}$–Vektorraum V ist eine lineare Abbildung $\varphi : V \to \mathbb{K}$.

(2) Der Vektorraum $L(V,\mathbb{K})$ aller linearen Funktionale auf V wird der *Dualraum von V* genannt und mit V^* bezeichnet.

(3) Für zwei $\mathbb{K}$–Vektorräume V und W und eine lineare Abbildung $f : V \to W$ definiert man die *duale Abbildung* $f^* : W^* \to V^*$ durch $f^*(\varphi) := \varphi \circ f$. Das macht Sinn, weil $f^*(\varphi)$ als Komposition linearer Abbildungen selbst linear ist.

(4) Sei V endlichdimensional und $\mathcal{B}$ eine geordnete Basis von V, die aus den Elementen $v_1, \dots, v_n$ besteht. Dann ist die *duale Basis* $\mathcal{B}^*$ zu $\mathcal{B}$ die geordnete Teilmenge von V^*, die aus den Funktionalen $\varphi_1, \dots, \varphi_n$ besteht, die charakterisiert sind durch

$$\varphi_i(v_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

(5) Für eine Matrix $A = (a_{ij}) \in M_{m,n}(\mathbb{K})$ ist die *transponierte Matrix* $A^t = (b_{ij}) \in M_{n,m}(\mathbb{K})$ gegeben durch $b_{ij} = a_{ji}$. Die Matrix A^t entsteht also aus A durch “Vertauschen von Zeilen und Spalten” bzw. durch “Spiegeln an der Hauptdiagonale”.

Satz 5.12. *Seien U, V und W Vektorräume, $f : V \to W$ und $g : U \to V$ lineare Abbildungen. Dann gilt:*

(1) Die duale Abbildung $f^* : W^* \to V^*$ ist ebenfalls linear.

(2) Es gilt $(f \circ g)^* = g^* \circ f^* : W^* \to U^*$.

(3) Ist V endlichdimensional, und $\mathcal{B}$ eine Basis für V, dann ist die duale Basis $\mathcal{B}^*$ eine Basis für V^*. Insbesondere ist auch V^* endlichdimensional und $\dim(V^*) = \dim(V)$.

(4) Seien V und W endlichdimensional, $\mathcal{B}$ eine Basis für V und $\mathcal{C}$ eine Basis für W mit dualen Basen $\mathcal{B}^*$ und $\mathcal{C}^*$. Dann ist die Matrixdarstellung von f^* bezüglich der dualen Basen gegeben durch $[f^*]^{\mathcal{C}^*}_{\mathcal{B}^*} = ([f]^{\mathcal{B}}_{\mathcal{C}})^t$.

Beweis. (1) Nach Definition ist die Vektorraumstruktur auf $W^* = L(W, \mathbb{K})$ durch die punktweisen Operationen gegeben, d.h. für $\varphi_1, \varphi_2 \in W^*$, $r \in \mathbb{K}$ und $w \in W$ gilt $(\varphi_1 + \varphi_2)(w) = \varphi_1(w) + \varphi_2(w)$ und $(r\varphi)(w) = r\varphi(w)$. Wiederum nach Definition ist $f^*(\varphi_i) : V \to \mathbb{K}$ gegeben durch $(f^*(\varphi_i))(v) = \varphi_i(f(v))$ für $i = 1, 2$. Damit ist aber

$$\begin{aligned}(f^*(\varphi_1) + f^*(\varphi_2))(v) &= (f^*(\varphi_1))(v) + (f^*(\varphi_2))(v) = \varphi_1(f(v)) + \varphi_2(f(v)) \\ &= (\varphi_1 + \varphi_2)(f(v)) = (f^*(\varphi_1 + \varphi_2))(v),\end{aligned}$$

also $f^*(\varphi_1) + f^*(\varphi_2) = f^*(\varphi_1 + \varphi_2)$. Analog ist

$$(f^*(r\varphi_1))(v) = (r\varphi_1)(f(v)) = r(\varphi_1(f(v))) = r((f^*(\varphi_1))(v)) = (r(f^*(\varphi_1)))(v),$$

also $f^*(r\varphi_1) = r(f^*(\varphi_1))$, also ist f^* linear.

(2) Für $\varphi : W \to \mathbb{K}$ und $u \in U$ ist $((f \circ g)^*\varphi)(u) = \varphi((f \circ g)(u)) = \varphi(f(g(u)))$. Andererseits ist $(g^* \circ f^*)(\varphi) = g^*(f^*(\varphi))$, und das bildet $u \in U$ auf $(f^*(\varphi))(g(u)) = \varphi(f(g(u)))$ ab.

(3) Nach Satz 4.15 ist $\dim(V^*) = \dim(V)\dim(\mathbb{K}) = \dim(V)$. Nach Korollar 4.5 genügt es daher zu zeigen, dass die Elemente φ_i der dualen Basis linear unabhängig sind. Sind aber $r_1, \dots, r_n \in \mathbb{K}$ so, dass $r_1\varphi_1 + \dots + r_n\varphi_n = 0$ gilt, dann gilt für die Elemente v_i von $\mathcal{B}$ natürlich

$$0 = (r_1\varphi_1 + \dots + r_n\varphi_n)(v_i) = r_1\varphi_1(v_i) + \dots + r_n\varphi_n(v_i) = r_i.$$

(4) Auch das ist ziemlich einfach, wenn man die Definitionen verdaut hat. Seien v_i, w_j, φ_k und ψ_ℓ die Elemente von $\mathcal{B}$, $\mathcal{C}$, $\mathcal{B}^*$ und $\mathcal{C}^*$. Entwickelt man ein Element $v \in V$ als $v = a_1v_1 + \dots + a_nv_n$, dann gilt nach Definition $\varphi_i(v) = a_1\varphi_i(v_1) + \dots + a_n\varphi_i(v_n) = a_i$. Das zeigt aber gerade, dass der Koordinatenvektor $[v]_{\mathcal{B}}$ gegeben ist durch $(\varphi_1(v), \dots, \varphi_n(v))$.

Ganz analog kann man $\varphi \in V^*$ als $b_1\varphi_1 + \dots + b_n\varphi_n$ entwickeln und erhält $\varphi(v_i) = b_1\varphi_1(v_i) + \dots + b_n\varphi_n(v) = b_i$. Damit ist aber $[\varphi]_{\mathcal{B}^*} = (\varphi(v_1), \dots, \varphi(v_n))$.

Ist nun $[f]^{\mathcal{B}}_{\mathcal{C}} = A = (a_{ij})$, dann ist der j–te Spaltenvektor von A gerade $[f(v_j)]_{\mathcal{C}}$, also $a_{ij} = \psi_i(f(v_j))$. Analog ist für $[f^*]^{\mathcal{C}^*}_{\mathcal{B}^*} = B = (b_{ij})$ der j–te Spaltenvektor gerade $[f^*(\psi_j)]_{\mathcal{B}^*}$, also

$$b_{ij} = (f^*(\psi_j))(v_i) = \psi_j(f(v_i)) = a_{ji}.$$

□

Die Resultate über den Zusammenhang zwischen der dualen Basis und Koordinatenvektoren bzw. Matrixdarstellungen von linearen Abbildungen aus dem Beweis von Teil (4) sind natürlich auch allgemeiner nützlich. Die Symmetrie zwischen Raum und Dualraum, die in diesem Beweisteil erkennbar wird, ist ein erster Hinweis auf das Dualitätsprinzip.

5.13. Annihilatoren und Dualität. Grundlage für die Dualität ist ein Zusammenhang zwischen Teilräumen eines Vektorraumes V und seines Dualraumes V^*. Um das effizient nutzen zu können ist es notwendig, auch den Dualraum des Dualraumes zu beachten. Daher werden wir als ersten Schritt zeigen, das wir den sogenannten *Bidualraum* $(V^*)^*$ im Fall endlichdimensionaler Vektorräume mit dem ursprünglichen Raum V identifizieren können.

Lemma 5.13. *Sei V ein $\mathbb{K}$–Vektorraum. Dann definiert man für jedes Element $v \in V$ eine Funktion $i(v) : V^* \to \mathbb{K}$ durch $i(v)(\varphi) := \varphi(v)$. Das liefert eine lineare Abbildung $i = i_V : V \to (V^*)^*$ die für jeden endlichdimensionalen Vektorraum V ein linearer Isomorphismus ist.*

Ist W ein weiterer $\mathbb{K}$–Vektorraum mit Dualraum W^* und Bidualraum $(W^*)^*$ und $f : V \to W$ eine lineare Abbildung mit dualer Abbildung $f^* : W^* \to V^*$, dann gilt für $(f^*)^* : (V^*)^* \to (W^*)^*$ die Gleichung $(f^*)^* \circ i_V = i_W \circ f$. Im Fall endlichdimensionaler Räume kann man also $(f^*)^*$ mit f identifizieren.

Beweis. Nach Definition gilt $i(v)(0) = 0$ und

$$i(v)(r\varphi_1 + \varphi_2) = (r\varphi_1 + \varphi_2)(v) = r\varphi_1(v) + \varphi_2(v) = ri(v)(\varphi_1) + i(v)(\varphi_2),$$

also ist $i(v)$ linear. Außerdem ist $i(0) = 0$ und

$$i(rv_1+v_2)(\varphi) = \varphi(rv_1+v_2) = r\varphi(v_1)+\varphi(v_2) = ri(v_1)(\varphi)+i(v_2)(\varphi) = (ri(v_1)+i(v_2))(\varphi),$$

also ist $i : V \to (V^*)^*$ linear. Ist $v \neq 0$ ein beliebiger Vektor in V, dann ist $\{v\} \subset V$ linear unabhängig, kann also zu einer Basis für V erweitert werden. Das erste Element φ der dualen Basis erfüllt dann $i(v)(\varphi) = \varphi(v) = 1$, also ist $i(v) \neq 0$ und damit i injektiv. Für endlichdimensionales V gilt $\dim(V) = \dim((V^*)^*)$ nach Satz 5.12, also ist i ein linearer Isomorphismus.

Für $\alpha \in (V^*)^*$ ist nach Definition $(f^*)^*(\alpha) = \alpha \circ f^*$, also $(f^*)^*(\alpha)(\psi) = \alpha(f^*(\psi))$ für alle $\psi \in W^*$. Ist aber $\alpha = i_V(v)$, dann ist

$$\alpha(f^*(\psi)) = (f^*(\psi))(v) = \psi(f(v)) = i_W(f(v))(\psi),$$

also folgt auch die letzte Behauptung. □

Die grundlegende Verbindung zwischen Teilmengen (und insbesondere Teilräumen) von V und V^* erhalten wir nun wie folgt:

Definition 5.13. Sei V ein $\mathbb{K}$–Vektorraum mit Dualraum V^*. Für eine Teilmenge $A \subset V$ definieren wir den *Annihilator* $A^\circ := \{\varphi \in V^* : \varphi(a) = 0 \quad \forall a \in A\}$

Im Fall eines endlichdimensionalen $\mathbb{K}$–Vektorraumes V können wir nach dem Lemma den Dualraum von V^* wieder mit V identifizieren. Somit können wir für eine Teilmenge $B \subset V^*$ den Annihilator B als Teilmenge von V betrachten. Insbesondere können wir also für eine Teilmenge $A \subset V$ den Annihilator $(A^\circ)^\circ$ wieder als Teilmenge von V betrachten. Damit können wir die grundlegenden Eigenschaften der Annihilatorabbildung beweisen:

Satz 5.13. *Sei V ein endlichdimensionaler $\mathbb{K}$–Vektorraum mit Dualraum V^*.*

(1) Für jede Teilmenge $A \subset V$ ist der Annihilator $A^\circ \subset V^*$ ein Teilraum. Ist $A_1 \subset A_2$, dann ist $A_1^\circ \supset A_2^\circ$.

(2) Für eine Teilmenge $A \subset V$ mit linearer Hülle $\langle A \rangle \subset V$ ist $A^\circ = (\langle A \rangle)^\circ$. Identifiziert man den Bidualraum $(V^*)^*$ mit V, so gilt $(A^\circ)^\circ = \langle A \rangle$.

(3) Für einen Teilraum $W \subset V$ ist $W^\circ \cong (V/W)^*$. Insbesondere ist $\dim(W^\circ) = \dim(V) - \dim(W)$.

(4) Für Teilräume $W_1, W_2 \subset V$ gilt

$$(W_1 \cap W_2)^\circ = W_1^\circ + W_2^\circ$$
$$(W_1 + W_2)^\circ = W_1^\circ \cap W_2^\circ$$

(5) Ist Z ein weiterer, endlichdimensionaler $\mathbb{K}$–Vektorraum und $f : V \to Z$ eine lineare Abbildung mit dualer Abbildung $f^* : Z^* \to V^*$ dann gilt $\operatorname{Im}(f^*) = \operatorname{Ker}(f)^\circ$, $\operatorname{Ker}(f^*) = \operatorname{Im}(f)^\circ$ *und* $\operatorname{rg}(f^*) = \operatorname{rg}(f)$.

Beweis. Für eine beliebige Teilmenge $A \subset V$ und das Nullfunktional $0 \in V^*$ gilt natürlich $0 \in A^\circ$. Sind $\varphi_1, \varphi_2 \in A^\circ$ und $r \in \mathbb{K}$, dann gilt für $a \in A$ natürlich

$$(r\varphi_1 + \varphi_2)(a) = r\varphi_1(a) + \varphi_2(a) = 0,$$

also $r\varphi_1 + \varphi_2 \in A^\circ$, also ist A° immer ein Teilraum. Ist $A_1 \subset A_2$ und $\varphi \in A_2^\circ$, dann gilt nach Definition $\varphi(a) = 0$ für alle $a \in A_2$, also insbesondere auch für alle $a \in A_1$. Damit ist $\varphi \in A_1^\circ$, also $A_2^\circ \subset A_1^\circ$ und das vervollständigt den Beweis von (1).

Ist $f : V \to Z$ linear, dann gilt nach Definition der dualen Abbildung $f^*(\varphi) = \varphi \circ f$ für $\varphi \in Z^*$. Damit ist aber $f^*(\varphi) = 0$ äquivalent zu $\varphi(f(v)) = 0$ für alle $v \in V$, also zu $\varphi(z) = 0$ für alle $z \in \operatorname{Im}(f) \subset Z$. Damit erhalten wir $\operatorname{Ker}(f^*) = \operatorname{Im}(f)^\circ$. Andererseits gilt für $v \in \operatorname{Ker}(f)$ natürlich $f^*(\varphi)(v) = \varphi(f(v)) = 0$ also liegt $f^*(\varphi)$ im Annihilator von $\operatorname{Ker}(f)$. Somit erhalten wir auch $\operatorname{Im}(f^*) \subset \operatorname{Ker}(f)^\circ$.

Betrachten wir insbesondere einen Teilraum $W \subset V$ und die natürliche Surjektion $\pi : V \to V/W$, dann ist die duale Abbildung eine lineare Abbildung $\pi^* : (V/W)^* \to V^*$. Da π surjektiv ist, folgt aus $\varphi \circ \pi = 0$ natürlich $\varphi = 0$, also ist π^* injektiv. Da $\operatorname{Ker}(\pi) = W$ gilt, folgt von oben, dass $\operatorname{Im}(\pi^*) \subset W^\circ$ gilt, also ist π^* eine injektive lineare Abbildung $(V/W)^* \to W^\circ \subset V^*$. Ist aber $\psi : V \to \mathbb{K}$ ein lineares Funktional, das in W° liegt, dann gilt $W \subset \operatorname{Ker}(\psi)$. Damit gibt es nach Proposition 5.11 eine lineare Abbildung $\underline{\psi} : V/W \to \mathbb{K}$, sodass $\psi = \underline{\psi} \circ \pi = \pi^*(\underline{\psi})$. Also ist $\pi^* : (V/W)^* \to W^\circ$ surjektiv, und somit ein linearer Isomorphismus, und (3) folgt.

Kehren wir zu einer allgemeinen linearen Abbildung $f : V \to Z$ mit dualer Abbildung $f^* : Z^* \to V^*$ zurück, für die wir $\operatorname{Ker}(f^*) = \operatorname{Im}(f)^\circ$ und $\operatorname{Im}(f^*) \subset \operatorname{Ker}(f)^\circ$ bereits bewiesen haben. Nun wissen wir aber aus (3), dass

$$\dim(\operatorname{Ker}(f)^\circ) = \dim(V) - \dim(\operatorname{Ker}(f)) = \dim(\operatorname{Im}(f))$$

gilt, wobei die zeite Gleichheit des Dimensionssatz 4.11 benutzt. Nochmals nach (3) ist $\dim(\operatorname{Ker}(f^*)) = \dim(\operatorname{Im}(f)^\circ) = \dim(Z) - \dim(\operatorname{Im}(f))$, und wir erhalten $\dim(\operatorname{Im}(f^*)) = \dim(Z) - \dim(\operatorname{Ker}(f^*)) = \dim(\operatorname{Im}(f))$. Damit folgt aber $\dim(\operatorname{Ker}(f)^\circ) = \dim(\operatorname{Im}(f^*))$ und damit $\operatorname{Im}(f^*) = \operatorname{Ker}(f)^\circ$, was den Beweis von (5) vervollständigt.

(2) Aus Teil (1) und $A \subset \langle A \rangle$ erhalten wir sofort $\langle A \rangle^\circ \subset A^\circ$. Nun gibt es für ein Element $v \in \langle A \rangle$ aber nach Proposition 4.1 endlich viele Elemente $a_i \in A$ und $r_i \in \mathbb{K}$, sodass $v = \sum_i r_i a_i$ gilt. Für $\varphi \in A^\circ$ folgt dann aber $\varphi(v) = \sum_i r_i \varphi(a_i) = 0$. Damit gilt $A^\circ \subset \langle A \rangle^\circ$, also stimmen die beiden Teilräume überein. In der Identifikation von V mit $(V^*)^*$ bildet das Element $v \in V$ ein Funktional $\psi \in V^*$ auf $\psi(v) \in \mathbb{K}$ ab. Damit zeigt die letzte Beobachtung aber auch, dass $\langle A \rangle \subset (A^\circ)^\circ$ gilt. Nun ist aber $(A^\circ)^\circ = (\langle A \rangle^\circ)^\circ$ ein Teilraum, dessen Dimension nach Teil (3) gleich

$$\dim(V^*) - \dim(\langle A \rangle^\circ) = \dim(V^*) - (\dim(V) - \dim(\langle A \rangle)) = \dim(\langle A \rangle)$$

ist, was sofort die Behauptung impliziert.

(4) Aus der Definition des Annihilators folgt sofort, dass $W_1^\circ \cap W_2^\circ = (W_1 \cup W_2)^\circ$ gilt. Nach (1) ist aber $(W_1 \cup W_2)^\circ = \langle W_1 \cup W_2 \rangle^\circ$ und nach Definition ist das Erzeugnis $W_1 + W_2$. Somit erhalten wir die zweite der behaupteten Gleichungen.

Wenden wir das auf die Teilräume $W_1^\circ, W_2^\circ \subset V^*$ an und identifizieren wir den Bidualraum mit V, dann erhalten wir nach (2) sofort $(W_1^\circ + W_2^\circ)^\circ = W_1 \cap W_2$. Daraus folgt die erste behauptete Gleichung sofort durch Bilden des Annihilators unter Verwendung von (2). □

Wir können die Abbildung $W \mapsto W^\circ$ sowohl als Abbildung von der Menge der Teilräume von V in die Menge der Teilräume des Dualraumes V^* betrachten, als auch als Abbildung in die umgekehrte Richtung. Ist $n = \dim(V) = \dim(V^*)$, dann werden jeweils k–dimensionale Teilräume auf Teilräume der komplementären Dimension $n-k$ abgebildet. Nach Teil (2) des Satzes gilt insbesondere $(W^\circ)^\circ = W$ für jeden Teilraum W. Das bedeutet aber gerade, dass die beiden Abbildungen $W \mapsto W^\circ$ invers zueinander sind. Insbesondere ist $W \mapsto W^\circ$ eine Bijektion zwischen der Menge der k–dimensionalen Teilräume von V und der Menge der $(n-k)$–dimensionalen Teilräume von V^* und umgekehrt. Insbesondere hat also ein n–dimensionaler Vektorraum "gleich viele" k–dimensionale wie $(n-k)$–dimensionale Teilräume.

5.14. Der Zeilenrang. In 4.13 konnten wir den Rang einer Matrix $A \in M_{m,n}(\mathbb{K})$ schön in Termen der Spaltenvektoren von A interpretieren. Nach Definition ist der Rang von A ja die Dimension des Bildes der linearen Abbildung $f : \mathbb{K}^n \to \mathbb{K}^m$, die durch $f(x) = Ax$ gegeben ist. Da die Spaltenvektoren von A ein Erzeugendensystem für $\mathrm{Im}(f)$ bilden, enthalten sie eine Basis für $\mathrm{Im}(f)$. Damit folgt sofort, dass der Rang von A genau die maximale Anzahl von Elementen einer linear unabhängigen Menge von Spaltenvektoren von A ist, siehe Proposition 4.13. Natürlich macht ein analoger Begriff auch in Termen der Zeilenvektoren von A Sinn.

Definition 5.14. Man sagt, eine Matrix $A \in M_{m,n}(\mathbb{K})$ hat *Zeilenrang* r, wenn es r Zeilenvektoren von A gibt, die lineare unabhängig sind, aber je $r+1$ Zeilenvektoren von A linear abhängig sind.

Will man den Unterschied zum Zeilenrang herausstreichen, dann wird der übliche Rang von A manchmal auch als *Spaltenrang* bezeichnet. Da beim Transponieren einer Matrix die Rollen von Zeilen und Spalten vertauscht werden, folgt sofort, dass der Zeilenrang einer Matrix A genau der Rang der transponierten Matrix A^t ist. Wir haben bereits in unseren einleitenden Überlegungen in Kapitel 1 beobachtet, dass Existenz von Lösungen eines linearen Gleichungssystems nicht nur in Termen der Spalten der entsprechenden Matrix sondern auch in Termen ihrer Zeilen schön interpretiert werden kann. Die Interpretation in Termen der Spalten ist wegen des Zusammenhangs mit dem Bild der Abbildung $x \mapsto Ax$ konzeptuell einsichtig. Jetzt können wir auch die Interpretation in Termen der Zeilen konzeptuell gut verstehen.

Korollar 5.14 (zu Satz 5.13). *Für eine Matrix* $A \in M_{m,n}(\mathbb{K})$ *der Zeilenrang mit dem Spaltenrang überein, also gilt* $\mathrm{rg}(A) = \mathrm{rg}(A^t)$.

Beweis. Nach Teil (4) von Satz 5.12, kann man A und A^t als Matrixdarstellungen dualer Abbildungen interpretieren. Wir haben aber in Teil (5) von Satz 5.13 gesehen, dass duale Abbildungen den gleichen Rang haben. $\square$

KAPITEL 6

Determinanten

In Kapitel 4 haben wir gesehen, wie man lineare Abbildungen zwischen zwei $\mathbb{K}$–Vektorräumen durch Matrizen beschreiben kann. Dazu muss man Basen für die beiden Räume wählen und die Abhängigkeit der Matrixdarstellung von diesen Wahlen motiviert den Begriff der Ähnlichkeit von Matrizen. Mit den Sätzen über Ähnlichkeit in Kapitel 4, insbesondere mit Satz 4.18, haben wir solche Abbildungen mehr oder minder vollständig verstanden.

Die Situation ändert sich aber, wenn man lineare Abbildungen von einem $\mathbb{K}$–Vektorraum V auf sich selbst studiert. In diesem Fall wird man nämlich nur Matrixdarstellungen bezüglich *einer* Basis von V betrachten und nicht solche bezüglich zweier verschiedener Basen von V. Damit muss der Begriff von Ähnlichkeit geändert werden und es wird schwieriger "schöne" Matrixdarstellungen für eine gegebene lineare Abbildung zu finden.

Ein zentrales Werkzeug zum Verständnis linearer Abbildungen von einem Vektorraum auf sich selbst ist die Determinante, die wir in diesem Kapitel studieren werden. Im Hinblick auf spätere Anwendungen werden wir die Theorie der Determinante in einem etwas allgemeineren Rahmen enwickeln, nämlich für kommutative Ringe mit Einselement. Das bereitet kaum zusätliche Schwierigkeiten, man erhält aber einerseits Resultate über lineare Gleichungssyteme über kommutativen Ringen mit Einselement, also zum Beispiel über die Frage ganzzahliger Lösungen von linearen Gleichungssystemen mit ganzzahligen Koeffizienten. Andererseits wird es später wichtig sein, Resultate über Determinanten auf Polynomringe anwenden zu können.

6.1. Ähnlichkeit für quadratische Matrizen. Erinneren wir uns zunächst an den Begriff von Ähnlichkeit aus Kapitel 4. Seien V und W endlichdimensionale Vektorräume über einem Körper $\mathbb{K}$ und sei $f : V \to W$ eine lineare Abbildung. Wählt man eine Basis $\mathcal{B} = \{v_1, \dots, v_n\}$ für V, dann kann man jedem Vektor $v \in V$, den *Koordinatenvektor* $[v]_{\mathcal{B}} \in \mathbb{K}^n$ bezüglich $\mathcal{B}$ zuordnen. Ist $v = a_1v_1 + \dots + a_nv_n$ die (eindeutige) Darstellung von v als Linearkombination der Elemente von $\mathcal{B}$, dann ist $[v]_{\mathcal{B}} = (a_1, \dots, a_n)$. Wählt man nun noch eine Basis $\mathcal{C} = \{w_1, \dots, w_m\}$ für W, dann kann man f durch die Matrixdarstellung $[f]^{\mathcal{B}}_{\mathcal{C}} \in M_{m,n}(\mathbb{K})$ beschreiben, die durch $[f(v)]_{\mathcal{C}} = [f]^{\mathcal{B}}_{\mathcal{C}}[v]_{\mathcal{B}}$ charakterisiert ist.

Es ist auch leicht zu beschreiben, wie die Matrixdarstellungen bezüglich verschiedener Basen zusammenhängen. Ist $\tilde{\mathcal{B}}$ eine weitere Basis für V, dann kann man die Matrizen zum Basiswechsel bilden, die man in der Notation von oben als $[\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ bzw $[\mathrm{id}]^{\tilde{\mathcal{B}}}_{\mathcal{B}} = ([\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}})^{-1}$ schreiben kann. Analog gibt es zu einer weiteren Basis $\tilde{\mathcal{C}}$ für W die entsprechenden Basiswechselmatrizen, und es gilt

$$[f]^{\tilde{\mathcal{B}}}_{\tilde{\mathcal{C}}} = [\mathrm{id}]^{\mathcal{C}}_{\tilde{\mathcal{C}}}[f]^{\mathcal{B}}_{\mathcal{C}}[\mathrm{id}]^{\tilde{\mathcal{B}}}_{\mathcal{B}} = [\mathrm{id}]^{\mathcal{C}}_{\tilde{\mathcal{C}}}[f]^{\mathcal{B}}_{\mathcal{C}}([\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}})^{-1}.$$

(Wir werden in Kürze sehen, warum die zweite Schreibweise die natürlichere ist.) Schließlich ist noch wichtig zu bemerken, dass für eine gegebene Basis $\mathcal{B}$ für V *jede* invertierbare $n \times n$–Matrix als $[\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ für eine geeignete Basis $\tilde{\mathcal{B}}$ von V geschrieben werden kann.

Damit sieht man, dass für eine gegeben Matrixdarstellung $A \in M_{m,n}(\mathbb{K})$ einer linearen Abbildung f, die Matrixdarstellungen von f bezüglich beliebiger Basen genau die Matrizen der Form SAT^{-1} für invertierbare Matrizen $S \in M_m(\mathbb{K})$ und $T \in M_n(\mathbb{K})$ sind. Entsprechend nennt man zwei rechteckige Matrizen $A, B \in M_{m,n}(\mathbb{K})$ *ähnlich*, wenn es invertierbare Matrizen S und T wie oben gibt, sodass $B = SAT^{-1}$ gilt, siehe 4.18. Das zeigt, wie man lineare Abbildungen über ihre Matrixdarstellungen studieren kann. Jede Eigenschaft einer Matrix, die sich automatisch auf alle ähnlichen Matrizen überträgt kann man als Eigenschaft von linearen Abbildungen betrachten. Diese Eigenschaft kann man dann einfach für eine beliebige Matrixdarstellung von f überprüfen.

Andererseits kann man natürlich fragen, wie eine möglist einfache Matrixdarstellung für eine gegeben lineare Abbildung f aussieht. Wir wissen aus Abschnitt 4.18, dass Ähnlichkeit von Matrizen eine Äquivalenzrelation ist und die suche nach einer möglich einfachen Matrixdarstellung bedeutet, dass man in jeder Äquivalenzklasse einen möglichst einfachen Repräsentanten angeben möchte.

Tatsächlich haben wir beide Probleme in Satz 4.18 vollständig gelöst. Einerseits sind zwei Matrizen genau dann ähnlich, wenn sie den gleichen Rang haben. Somit ist der Rang die einzige wesentliche Invariante einer linearen Abbildung zwischen zwei verschiedenen endlichdimensionalen Vektorräumen. Andererseits kann man leicht einen einfachen Repräsentanten für die Matrizen mit Rang r angeben, zum Beispiel die Matrix $\mathbb{J}_r = (a_{ij})$ für die $a_{11} = a_{22} = \dots = a_{rr} = 1$ gilt, während alle anderen Eintragungen Null sind.

Betrachten wir als nächstes den Fall einer lineare Abbildungen $f : V \to V$, also von einem Vektorraum auf sich selbst. Dann können Matrixdarstellungen bezüglich *einer* Basis von von V einschränken, anstatt zwei verschiedene Basen für V zu verwenden. Für eine Basis $\mathcal{B} = \{v_1, \dots, v_n\}$ von V schreiben wir ab sofort $[f]_{\mathcal{B}}$ statt $[f]^{\mathcal{B}}_{\mathcal{B}}$, und sofern nicht explizit anderes gesagt wird, werden wir uns ab sofort nur noch für solche Matrixdarstellungen interessieren.

Von oben können wir sofort ablesen, wie sich diese Matrixdarstellung ändert, wenn wir statt $\mathcal{B}$ eine andere Basis $\tilde{\mathcal{B}}$ für V betrachten. Ist $T := [\mathrm{id}]^{\mathcal{B}}_{\tilde{\mathcal{B}}}$ die invertierbare Matrix des Basiswechsels, dann gilt

$$[f]_{\tilde{\mathcal{B}}} = T[f]_{\mathcal{B}}T^{-1}.$$

Im Gegensatz zu vorher muss man also von links mit einer beliebigen invertierbaren Matrix und von rechts mit ihrer Inversen multiplizieren und nicht mit zwei beliebigen invertierbaren Matrizen. Entsprechend adaptiert man den Begriff von Ähnlichkeit für quadratische Matrizen:

Definition 6.1. Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}$. Zwei quadratische Matrizen $A, B \in M_n(\mathbb{K})$ heißen *ähnlich*, wenn es eine invertierbare Matrix $T \in M_n(\mathbb{K})$ gibt, sodass $B = TAT^{-1}$ gilt. In diesem Fall schreibt man $A \sim B$.

Sofern nicht explizit anderes gesagt wird, werden wir ab sofort nur noch diesen Begriff von Ähnlichkeit und nicht den “alten” Begriff, der auch für rechteckige Matrizen Sinn macht, verwenden.

Die grundlegenden Eigenschaften folgen nun sofort aus bereits bekannten Resultaten bzw. werden ganz analog zu diesen bewiesen:

Proposition 6.1. *(1) Sei V ein endlichdimensionaler Vektorraum über $\mathbb{K}$, $f : V \to V$ eine lineare Abbildung, und $A \in M_n(\mathbb{K})$ eine Matrixdarstellung für f, d.h. $A = [f]_{\mathcal{B}}$ für eine Basis $\mathcal{B}$ von V.*

Dann ist eine Matrix $B \in M_n(\mathbb{K})$ genau dann eine Matrixdarstellung von f, wenn sie ähnlich zu A im Sinne von Definition 6.1 ist.

(2) Ähnlichkeit quadratischer Matrizen ist eine Äquivalenzrelation. Ähnliche Matrizen haben den gleichen Rang, es gibt aber Matrizen mit gleichem Rang, die nicht zueinander ähnlich sind.

Beweis. (1) folgt sofort aus den obigen Überlegungen und Lemma 4.18, und den Anfang von (2) beweist man völlig analog zu Bemerkung 4.18 (2). Die Details auszuführen ist eine gute Übung. Dass ähnliche Matrizen gleichen Rang haben folgt aus Satz 4.18. Umgekehrt betrachten wir $\mathbb{K} = \mathbb{R}$ und für $t \in \mathbb{R}$ die Matrix $t\mathbb{I}_n$, wobei $\mathbb{I}_n$ die Einheitsmatrix bezeichnet. Ist $T \in M_n(\mathbb{R})$ invertierbar, dann ist $T(t\mathbb{I}_n)T^{-1} = tT\mathbb{I}_nT^{-1} = tTT^{-1} = t\mathbb{I}_n$. Somit sind für $t \neq s$ die Matrizen $t\mathbb{I}_n$ und $s\mathbb{I}_n$ nicht ähnlich, obwohl sie für $t, s \neq 0$ beide Rang n haben. □

Bemerkung 6.1. (1) Analog zu den Überlegungen von oben können wir nun lineare Abbildungen $f : V \to V$ über ihre Matrixdarstellungen studieren. Man kann einfach jede Eigenschaft quadratischer Matrizen, die sich automatisch auf ähnliche Matrizen überträgt, als Eigenschaft von linearen Abbildungen auffassen. Umgekehrt kann man versuchen in jeder Äquivalenzklasse ähnlicher Matrizen einen möglichst einfachen Repräsentanten zu finden, eine sogenannte *Normalform*. Diese stellt dann eine ausgezeichnete "einfache" Matrixdarstellung für eine Klasse von linearen Abbildungen dar. Wie schon bemerkt, sind diese Fragen hier wesentlich interessanter und schwieriger als im Fall von Abbildungen zwischen verschiedenen Vektorräumen und sie werden uns auf längere Zeit beschäftigen.

(2) In den Kapiteln 3 und 4 haben elementare Zeilenoperationen eine zentrale Rolle gespielt, insbesondere beim Studium linearer Gleichungssysteme. Dabei war es wichtig, dass jede Folge elementarer Zeilenoperationen durch Multiplikation von links mit einer invertierbaren Matrix dargestellt werden kann. Damit führt jede Folge elementarer Zeilenoperationen von einer Matrix A zu einer Matrix, die im "alten" Sinn ähnlich zu A ist. Das gilt aber für den neuen Begriff von Ähnlichkeit aus Definition 6.1 *nicht*! Daher werden elementare Zeilenoperationen ab nun keine so große Rolle mehr spielen.

6.2. Kommutative Ringe mit Einselement. Wie schon erwähnt wird es aus technischen Gründen günstig sein, die Resultate über Determinanten in einem allgemeineren Rahmen zu entwickeln als für quadratische Matrizen über Körpern. Dazu werden wir hier kurz die Grundlagen besprechen, wobei der begriffliche Schwerpunkt im Bereich der linearen Gleichungssysteme liegt, wo die Verallgemeinerung unproblematisch ist. Es gibe auch Analoga der Begriffe des Vektorraumes und der linearen Abbildung in diesem Setting (sogenannte Moduln und Modulhomomorphismen), diese werden aber für unsere Zwecke nicht notwendig sein. Der Begriff des kommutativen Rings mit Einselement ist schon aus der EMA bekannt:

Definition 6.2. (1) Ein *kommutativer Ring mit Einselement* ist eine Menge R zusammen mit zwei Operationen $+ : R \times R \to R$ und $\cdot : R \times R \to R$, die beide assoziativ und kommutativ sind und neutrale Elemente besitzen (die wir mit 0 und 1 bezeichnen). Weiters verlangt man, dass jedes Element von R ein additiv inverses Element besitzt, und dass das Distributivgesetz $a \cdot (b + c) = a \cdot b + a \cdot c$ erfüllt ist.

(2) Ein *Ringhomomorphismus* zwischen zwei kommutativen Ringen $(R, +, \cdot)$ und $(S, +, \cdot)$ mit Einselement ist eine Funktion $\varphi : R \to S$, die $\varphi(a + b) = \varphi(a) + \varphi(b)$ und $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ für alle $a, b \in R$, sowie $\varphi(1) = 1$ erfüllt.

Ein kommutativer Ring erfüllt also alle Köperaxiome aus 2.1 mit Ausnahme von (K8), das die Existenz multiplikativ inverser Elemente verlangt. Für eine Ringhomomorphismus φ gilt automatisch $\varphi(0) = 0$ und $\varphi(-a) = -\varphi(a)$ für alle $a \in R$, aber $\varphi(1) = 1$ folgt nicht aus den anderen Eigenschaften. Wie schon im Fall von Körperen werden wir das Symbol für die Multiplikation oft einfach weglassen.

Ist $(R, +, \cdot)$ ein kommutativer Ring mit Einselement, dann können wir natürlich die Menge R^n aller Spaltenvektoren mit n Eintragungen aus R bilden. Dann definiert man $+ : R^n \times R^n \to R^n$ und $\cdot : R \times R^n \to R^n$ komponentenweise wie für Körper, und natürlich erfüllen diese Operationen die Eigenschaften (V1)–(V8) aus 2.2. Wie für Körper werden wir auch hier das Symbol für die Skalarmultiplikation oft weglassen. Wie für Körper definiert man den Begriff einer linearen Abbildung $f : R^n \to R^m$, als Funktion, die mit beiden Operationen verträglich ist, und es folgt leicht, dass eine Abbildung $f : R^n \to R$ genau dann linear ist, wenn sie die Form $f(v_1, \dots, v_n) = a_1 v_1 + \dots + a_n v_n$ für Elemente $a_1, \dots, a_n \in R$ hat (man hat ja wieder die Einheitsvektoren $e_i \in R^n$).

Es macht auch kein Problem, die Menge $M_{m,n}(R)$ are $(m \times n)$–Matrizen mit Eintragungen aus R zu bilden und diese mit R^{mn} zu identifizieren. Einer Matrix $A \in M_{m,n}(R)$ kann man dann wieder die Funktion $f : R^n \to R^m$, $f(v) = Av$ zuzuordnen, wobei die Wirkung einer Matrix auf einen Spaltenvektor genau wie für Körper definiert ist. Wie im Fall von Körpern sieht man, dass dies alle linearen Abbildungen $R^n \to R^m$ liefert. Auch die Matrizenmultiplikation überträgt sich problemlos auf Matrizen mit Eintragungen aus R. Da R ein Einselement besitzt, gibt es auch hier die Einheitsmatrix $\mathbb{I}_n \in M_n(R)$. Damit macht auch die Definition einer invertierbaren Matrix mit Eintragungen in R keine Probleme. Schließlich überträgt sich noch der Begriff des linearen Gleichungssystems über eine Matrix und einen Spaltenvektor problemlos in das Setting eines kommutativen Rings mit Einselement.

Vorsicht: Dass sich die Begriffe so problemlos in das allgemeinere Setting übertragen lassen, bedeutet natürlich nicht, dass auch die Resultate, die wir bewiesen haben, in diesem allgemeineren Setting gelten. Die Lösbarkeit linearer Gleichungssysteme zu charakterisieren ist über kommutativen Ringen mit Einselement viel schwieriger als über Köpern. Es ist eher überraschend, dass man mit Hilfe der Determinante zumindest in Spezialfällen relative starke Resultate im Setting von kommutativen Ringen mit Einselement beweisen kann.

Beispiel 6.2. (1) Schon aus der EMA ist bekannt, dass $\mathbb{Z}$ mit den üblichen Operationen ein kommutativer Ring mit Einselement ist. In diesem Beispiel ist die Relvanz von Fragen wie Lösbarkeit von linearen Gleichungssystemen unmittelbar einsichtig, man fragt nach Existenz von ganzzahligen Lösungen zu Gleichungssystemen mit ganzzahligen Koeffizienten.

(2) Auch das zweite Beispiel kennen wir bereits, nämlich die Menge $\mathbb{K}[x]$ aller Polynome in einer Variablen x mit Koeffizienten in einem Körper $\mathbb{K}$. In Abschnitt 2.6 haben wir $\mathbb{K}[x]$ zu einem $\mathbb{K}$–Vektorraum gemacht, insbesondere macht die dort definierte Addition $\mathbb{K}[x]$ zu einer kommutativen Gruppe. Die Multiplikation von Polynomen haben wir in 4.12 definiert und festgestellt, dass sie assoziativ, kommutativ und distributiv bezüglich der Addition ist. Da wir auch gezeigt haben, dass das konstante Polynom 1 ein multiplikativ neutrales Element ist, ist $(\mathbb{K}[x], +, \cdot)$ ein kommutativer Ring mit Einselement.

Die Resultate aus 2.6 und 4.12 liefern uns auch ein wichtiges Beispiel für einen Homomorphismus zwischen zwei kommutativen Ringen mit Einselement. In 2.6 haben wir festgestellet, dass jedes Polynom $p \in \mathbb{K}[x]$ eine zugehörige Polynomfunktion $\mathbb{K} \to \mathbb{K}$

definiert, indem man für die Variable x Elemente von $\mathbb{K}$ einsetzt. Betrachten wir nur ein fixes Element $\lambda \in \mathbb{K}$ und die Funktion $\varphi : \mathbb{K}[x] \to \mathbb{K}$, die gegeben ist durch $\varphi(p) := p(\lambda)$. Nach Proposition 2.6 ist diese Abbildung mit den Additionen verträglich und nach Proposition 4.12 mit den Multiplikationen. Für das konstante Polynom 1 ist natürlich $1(\lambda) = 1$ für jedes λ, also ist φ ein Homomorphismus von kommutativen Ringen mit Einselement.

Ein Aspekt den wir bisher noch nicht besprochen haben, muss hier noch erwähnt werden. Ist $\varphi : R \to S$ ein Homomorphismus zwischen zwei kommutativen Ringen mit Einselement, dann kann man φ natürlich komponentenweise Anwenden um $\tilde{\varphi} : R^k \to S^k$ für jedes k und $\hat{\varphi} : M_{m,n}(R) \to M_{m,n}(S)$ für beliebiges m und n zu definieren. Für $v = (v_1, \dots, v_n) \in R^n$ ist also $\tilde{\varphi}(v) = (\varphi(v_1), \dots, \varphi(v_n)) \in S^n$ und für eine Matrix $A = (a_{ij}) \in M_{m,n}(R)$ ist $\hat{\varphi}(A) = (\varphi(a_{ij})) \in M_{m,n}(S)$. Diese Operationen sind in schöner Weise mit der Wirkung von Matrizen auf Vektoren und mit der Matrizenmultiplikation verträglich.

Proposition 6.2. *Sei $\varphi : R \to S$ ein Homomorphismus von kommutativen Ringen mit Einselement. Dann erfüllen die induzierten Funktionen*

$$\tilde{\varphi}(Av) = \hat{\varphi}(A)\tilde{\varphi}(v) \qquad \hat{\varphi}(AB) = \hat{\varphi}(A)\hat{\varphi}(B).$$

Beweis. Nach Definition ist die $i$te Komponente von Av durch $\sum_j a_{ij}v_j$ gegeben, also hat $\tilde{\varphi}(Av)$ als $i$te Komponente $\varphi(\sum_j a_{ij}v_j)$. Da φ als Ringhomomorphismus sowohl mit der Addition als auch mit der Multiplikation in R verträglich ist, ist $\varphi(\sum_j a_{ij}v_j) = \sum_j \varphi(a_{ij})\varphi(v_j)$. Nun ist aber $\varphi(a_{ij})$ genau die (i,j)–Komponente von $\hat{\varphi}(A)$ und $\varphi(v_j)$ die $j$te Komponente von $\tilde{\varphi}(v)$. Das beweist die erste Gleichheit, die Verträglichkeit mit der Matrizenmultiplikation folgt völlig analog. □

Existenz der Determinante

Der erste Schritt zum Verständnis quadratischer Matrizen ist der Begriff der Determinante einer Matrix. Um den allgemeinen Begriff zu motivieren besprechen wir zunächst den Fall der Determinante von 2×2–Matrizen. In diesem Fall ist der Begriff schon aus der Schule bekannt und man kann alle wesentlichen Eigenschaften ganz einfach nachrechnen.

6.3. Die Determinante von 2×2–Matrizen. Sei R ein kommutativer Ring mit Einselement $M_2(R)$ die Menge der 2×2–Matrizen über R. Dann definieren wir die *Determinantenfunktion* $\det : M_2(R) \to R$ durch $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc$. Wie schon zuvor ist es auch hier oft günstig, die Matrix als Familie von Spaltenvektoren zu betrachten. Wir werden daher det auch als Funktion $R^2 \times R^2 \to R$ betrachten und als $\det(v_1, v_2)$ für $v_1, v_2 \in R^2$ schreiben. Die grundlegenden Eigenschaften der Determinantenfunktion können wir nun sofort ablesen:

(1) Für $r \in R$ gilt $\det(rv_1, v_2) = \det(v_1, rv_2) = r\det(v_1, v_2)$ und für $v_1', v_2' \in R^2$ gelten $\det(v_1 + v_1', v_2) = \det(v_1, v_2) + \det(v_1', v_2)$ sowie $\det(v_1, v_2 + v_2') = \det(v_1, v_2) + \det(v_1, v_2')$. Man sagt, die Determinante ist *linear in jeder Spalte*. Die Bedingungen besagen ja gerade, dass $v_1 \mapsto \det(v_1, v_2)$ für fixes v_2 und $v_2 \mapsto \det(v_1, v_2)$ für fixes v_1 lineare Abbildungen $R^2 \to R$ sind.

(2) Für beliebiges $v \in R^2$ ist $\det(v, v) = 0$. Man sagt, die Determinante ist *alternierend*.

(3) Für die Einheitsmatrix (bzw. für die Einheitsvektoren) gilt $\det(\mathbb{I}_2) = \det(e_1, e_2) = 1$. Man sagt, die Determinante ist *normiert*.

Die erste zentrale Eigenschaft der Determinante ist ihr Bezug zur Invertierbarkeit, der im Fall der hier ganz einfach explizit nachgerechnet werden kann. Es gilt nämlich

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = (ad-bc)\mathbb{I}_2.$$

Nehmen wir an, dass für $A \in M_2(R)$ die Determinante $\det(A) \in R$ ein multiplikativ inverses Element besitzt, es also ein Element $s \in R$ gibt, sodass $s(ab - cd) = 1$ gilt. (Im Fall eines Körpers bedeutet das nur, dass $\det(A) \neq 0$ gilt.) Dann zeigt die obige Rechnung, dass die Matrix $\left(\begin{smallmatrix} sd & -sb \\ -sc & sa \end{smallmatrix}\right)$ invers zu A ist. Umgekehrt zeigt eine direkte Rechnung, dass für $A, B \in M_2(R)$ die Gleichung $\det(AB) = \det(A)\det(B)$ gilt. Ist A invertierbar, dann gibt es eine Matrix B, sodass $AB = BA = \mathbb{I}$ gilt. Damit ist aber $\det(A)\det(B) = \det(\mathbb{I}) = 1$, also besitzt $\det(A) \in R$ ein multiplikativ inverses Element.

Die Multiplikativität der Determinante hat eine wichtige direkte Konsequenz. Ist $A \in M_2(R)$ beliebig und $T \in M_2(R)$ invertierbar, dann betrachten wir $B = TAT^{-1}$. Dann ist $\det(B) = \det(T)\det(A)\det(T^{-1})$ und in R können wir die Reihenfolge des Produkts vertauschen und es gilt $\det(T)\det(T^{-1}) = 1$. Damit ist aber $\det(B) = \det(A)$, insbesondere haben ähnliche Matrizen ueber einem Körper die gleiche Determinante.

Nehmen wir nun an, dass $A \in M_2(R)$ invertierbar ist, also $\det(A) \in R$ invertierbar ist. Dann sieht man leicht, dass das multiplikativ inverse Element zu $\det(A)$ eindeutig bestimmt ist, also schrieben wir dafür wieder $\det(A)^{-1}$. Wie im Fall von Körpern folgt sofort, dass das lineare Gleichungssystem $Ax = y$ für jedes $y \in R^2$ die eindeutige Lösung $x = A^{-1}y$ besitzt. Ist $y = \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)$, dann erhalten wir mit der Formel für A^{-1} von oben gerade $x_1 = \det(A)^{-1}(dy_1 - by_2)$ und $x_2 = \det(A)^{-1}(-cy_1 + ay_2)$. Bezeichnen wir die Spaltenvektoren von A mit v_1 und v_2, dann kann man das als $x_1 = \det(A)^{-1}\det(y, v_2)$ und $x_2 = \det(A)^{-1}\det(v_1, y)$ schreiben. Das ist der einfachste Fall der sogenannten *Cramer'schen Regel*, die explizite Lösungsformeln für lineare Gleichungssysteme liefert.

6.4. Determinantenfunktionen. Sei R ein kommutativer Ring mit Einselement. Wie in 6.3 werden wir weiterhin eine Matrix auch als Familie ihrer Spaltenvektoren betrachten, also den Raum $M_n(R)$ der $n \times n$–Matrizen über R auch als $R^n \times \ldots \times R^n$ (n Faktoren) betrachten.

Definition 6.4. (1) Eine Funktion $F : M_n(R) = R^n \times \ldots \times R^n \to R$ heißt eine *Determinantenfunktion*, wenn sie folgende Bedingungen erfüllt:

- F ist linear in jeder Spalte, d.h. für jedes $i = 1, \ldots, n$ und beliebige fixe Elemente v_j für $j \neq i$ ist $v_i \mapsto F(v_1, \ldots, v_i, \ldots, v_n)$ eine lineare Abbildung $R^n \to R$.
- Ist $v_i = v_j$ für ein $i \neq j$, dann ist $F(v_1, \ldots, v_n) = 0$.

(2) Eine Determinantenfunktion $F : M_n(R) \to R$ heißt normiert, wenn $F(\mathbb{I}_n) = F(e_1, \ldots, e_n) = 1$ gilt.

Bemerkung 6.4. (1) Die Funktion $\det : M_2(R) \to R$ aus 6.3 ist eine normierte Determinantenfunktion. Für $n > 2$ müssen wir erst beweisen, dass Determinantenfunktionen überhaupt existieren.

(2) Zu den elementaren Zeilenoperationen aus 3.6 gibt es natürlich analoge *elementare Spaltenoperationen*. Diese sind gegeben, indem man zwei Spalten vertauscht, eine Spalte mit einem Skalar ungleich Null multipliziert, oder zu einer Spalte ein beliebiges Vielfaches einer anderen Spalte addiert. Aus der Definition ergibt sich sofort, wie sich eine Determinantenfunktion F unter diesen Operationen verhält:

Addiert man zur $i$ten Spalte das r–fache der $j$ten Spalte, dann erhält man wegen der Linearität in der $i$ten Spalte

$$F(v_1,\dots,v_i+rv_j,\dots,v_j,\dots,v_n)=F(v_1,\dots,v_n)+rF(v_1,\dots,v_j,\dots,v_j,\dots,v_n)$$

und der zweite Summand verschwindet, weil zwei Spalten gleich sind. Bei dieser Art von Spaltenoperationen bleibt also der Wert jeder Determinantenfunktion unverändert. Multipliziert man eine Spalte mit $r\in R$, dann wird wegen der Linearität in dieser Spalte auch der Wert von F mit r multipliziert.

Schließlich behaupten wir, dass F bei Vertauschung von zwei Spalten das Vorzeichen wechselt. Dazu schreiben wir in die $i$te und die $j$te Spalte v_i+v_j, und erhalten nach Definition $0=F(v_1,\dots,v_i+v_j,\dots,v_i+v_j,\dots,v_n)$. Nach der Linearität in der $i$ten Spalte können wir die rechte Seite als $F(v_1,\dots,v_i,\dots,v_i+v_j,\dots,v_n)+F(v_1,\dots,v_j,\dots,v_i+v_j,\dots,v_n)$ schreiben und dann die Linearität in der $j$ten Spalte benutzen. Da F verschwindet, wenn zwei Spalten gleich sind, bleibt vom ersten Summanden $F(v_1,\dots,v_i,\dots,v_j,\dots,v_n)$ und vom zweiten $F(v_1,\dots,v_j,\dots,v_i,\dots,v_n)$ übrig, also erhalten wir $F(v_1,\dots,v_j,\dots,v_i,\dots,v_n)=-F(v_1,\dots,v_i,\dots,v_j,\dots,v_n)$. Wir werden später sehen, dass für jede Determinantenfunktion F und jede Matrix $A\in M_n(R)$ die Gleichung $F(A^t)=F(A)$ gilt, wobei A^t die transponierte Matrix bezeichnet. Daraus folgt dann, dass ein analoges Verhalten unter elementaren Zeilenoperationen gilt.

Wir können nun direkt beweisen, dass Determinantenfunktionen für Matrizen beliebiger Größe tatsächlich existieren:

Satz 6.4 (Existenzsatz). *Für jeden kommutativen Ring R mit Einselement und jedes $n\in\mathbb{N}$ gibt es eine normierte Determinantenfunktion* $\det: M_n(R)\to R$.

Beweis. Wir beweisen den Satz durch Induktion nach n und definieren die Determinantenfunktion induktiv. Für $n=1$ ist die Identität $\mathrm{id}_R: M_1(R)=R\to R$ eine normierte Determinantenfunktion. Für $n=2$ ist die Funktion det aus 6.3 eine normierte Determinantenfunktion. Nehmen wir induktiv an, dass wir bereits eine normierte Determinantenfunktion $\det: M_{n-1}(R)\to R$ definiert haben. Für $A=(a_{ij})\in M_n(R)$ sei $A_j\in M_{n-1}(R)$, die $(n-1)\times(n-1)$–Matrix die entsteht, wenn man in A die erste Zeile und die $j$te Spalte weglässt. Anders gesagt, ist $A=(v_1,\dots,v_n)$, dann ist $A_j=(\tilde v_1,\dots,\tilde v_{j-1},\tilde v_{j+1},\dots,\tilde v_n)$, wobei $\tilde v_i$ aus v_i durch weglassen der ersten Komponente entsteht. Nun definieren wir $\det(A):=\sum_{j=1}^n(-1)^{j+1}a_{1j}\det(A_j)$.

Wir müssen nun die definierenden Eigenschaften einer normierten Determinantenfunktion verifizieren. Für $A=(v_1,\dots,v_n)$ betrachte $B=(v_1,\dots,rv_i,\dots v_n)$. Dann ist $b_{1i}=ra_{1i}$ und $B_i=A_i$, also wird der $i$te Summand in der Definition von det mit r multipliziert. Für $j\neq i$ ist $b_{1j}=a_{1j}$, aber eine Spalte von B_j ist gerade das r–fache der entsprechenden Spalte von A_j, also $\det(B_j)=r\det(A_j)$ nach Induktionsvoraussetzung. Somit werden auch alle anderen Summanden mit r multipliziert, und wir erhalten $\det(B)=r\det(A)$.

Betrachte zu $A=(v_1,\dots,v_n)$ die Matrizen $B=(v_1,\dots,v_i',\dots,v_n)$ und $C=(v_1,\dots,v_i+v_i',\dots,v_n)$, dann ist $c_{1i}=a_{1i}+b_{1i}$ und $C_i=A_i=B_i$ und damit $c_{1i}\det(C_i)=a_{1i}\det(A_i)+b_{1i}\det(B_i)$. Für $j\neq i$ erhalten wir $c_{1j}=a_{1j}=b_{1j}$, aber in der Matrix C_j ist gerade eine Spalte gleich der Summe der entsprechenden Spalten von A_j und B_j, während alle anderen Spalten in allen drei Matrizen gleich sind. Nach Induktionsvoraussetzung impliziert das $\det(C_j)=\det(A_j)+\det(B_j)$ und damit auch $c_{1j}\det(C_j)=a_{1j}\det(A_j)+b_{1j}\det(B_j)$. Damit erhalten wir aber $\det(C)=\det(A)+\det(B)$, und somit ist det linear in jeder Spalte.

Nehmen wir nun an, dass in $A = (v_1, \dots, v_n)$ zwei Spalten gleich sind, etwa $v_i = v_j = v$ mit $i < j$. Für $k \neq i, j$ sind dann zwei Spalten von A_k gleich $\tilde{v}$, und damit reduziert sich die definierende Formel für $\det(A)$ zu $(-1)^{i+1} a_{1i} \det(A_i) + (-1)^{j+1} a_{1j} \det(A_j)$. Nach Definition gilt

$$\begin{aligned} A_i &= (\tilde{v}_1, \dots, \tilde{v}_{i-1}, \tilde{v}_{i+1}, \dots, \tilde{v}_{j-1}, \tilde{v}, \tilde{v}_{j+1}, \dots, \tilde{v}_n) \\ A_j &= (\tilde{v}_1, \dots, \tilde{v}_{i-1}, \tilde{v}, \tilde{v}_{i+1}, \dots, \tilde{v}_{j-1}, \tilde{v}_{j+1}, \dots, \tilde{v}_n). \end{aligned}$$

Nach Induktionsvoraussetzung ist det auf $(n-1) \times (n-1)$–Matrizen eine Determinantenfunktion, also wissen wir aus Bemerkung 6.4 (2), dass bei Vertauschen von zwei Spalten der Wert von det das Vorzeichen wechselt. Nun erhält man aber A_i aus A_j indem man die Spalte $\tilde{v}$ der Reihe nach mit den $j-i-1$ Spalten $\tilde{v}_{i+1}, \dots, \tilde{v}_{j-1}$ vertauscht. Damit ist aber $\det(A_i) = (-1)^{j-i-1} \det(A_j)$, also $(-1)^{i+1} \det(A_i) = -(-1)^{j+1} \det(A_j)$ und somit $\det(A) = 0$.

Die Tatsache, dass det normiert ist, ist ganz einfach zu zeigen: Ist $A = \mathbb{I}_n$ die $n \times n$–Einheitsmatrix, dann ist $a_{11} = 1$ und $A_1 = \mathbb{I}_{n-1}$ und $a_{1i} = 0$ für alle $i > 1$. Damit ist aber nach Definition $\det(\mathbb{I}_n) = 1 \det(\mathbb{I}_{n-1})$ und nach Induktionsvoraussetzung ist $\det(\mathbb{I}_{n-1}) = 1$. □

Bemerkung 6.4. Wir hätten im Beweis den Fall $n = 2$ gar nicht extra behandeln müssen. Die Funktion det aus 6.3 entsteht nämlich aus der Identität (die normierte Determinantenfunktion für $n = 1$) durch die im Beweis angegebene Konstruktion. Ist nämlich $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ dann ist $A_1 = d$ und $A_2 = c$ (jeweils als 1×1–Matrizen betrachtet). Die Formel aus dem Beweis vereinfacht sich also zu $a \det(d) - b \det(c) = ad - bc$.

Beispiel 6.4. Wir wollen die Formel für die Determinante einer 3×3–Matrix $A = (a_{ij})$ explizit berechnen. Aus der Definition von det erhalten wir

$$a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

und setzt man die 2×2–Determinanten ein, dann erhält man

$$a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$

Diese Formel kann man sich einfach mit der sogenannten *Regel von Sarrus* merken. Dazu bildet man eine 5×3–Matrix, indem man die ersten beiden Spalten der Matrix A nochmals hinten anhängt: $\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix}$. Dann addiert man die Produkte über die drei "Hauptdiagonalen" (von links oben nach rechts unten) mit positivem Vorzeichen und die Produkte über die drei "Nebendiagonalen" (von rechts oben nach links unten) mit negativem Vorzeichen.

Vorsicht: Für größere Matrizen (ab 4×4) gibt es kein Analogon der Regel von Sarrus! Versuche, Analoga so einer Regel zu verwenden sind ein Zeichen für schlimmes Unverständnis.

Aus der Konstruktion der Determinantenfunktion det im Beweis von Satz 6.4 erhalten wir leicht eine Verträglichkeit mit Homomorphismen.

Korollar 6.4. *Sei $\varphi : R \to S$ ein Homomorphismus zwischen kommutativen Ringen mit Einselement und für ein $n \in \mathbb{N}$ sei $\hat{\varphi} : M_n(R) \to M_n(S)$ die induzierte Funktion wie in 6.3. Dann gilt für jede Matrix $A \in M_n(R)$ die Gleichung* $\det(\hat{\varphi}(A)) = \varphi(\det(A))$.

Beweis. Wir beweisen die Behauptung durch Induktion nach n. Für $n = 1$ ist $A = (r)$ für ein Element $r \in R$ und $\hat{\varphi}(A) = (\varphi(r))$ nach Definition. Aus dem Beweis von Satz 6.3 wissen wir, dass $\det(A) = r$ gilt, also folgt die Behauptung.

Nehmen wir induktiv an, dass wir die Behauptung für Matrizen in $M_{n-1}(R)$ bewiesen haben. Für $A = (a_{ij}) \in M_n(R)$ und $B := \hat{\varphi}(A) = (b_{ij})$ gilt dann $b_{ij} = \varphi(a_{ij})$. Für jedes $j = 1, \dots, n$ entsteht $A_j \in M_{n-1}(R)$ durch Weglassen gewisser Eintragungen von A, woraus man sofort $B_j = \hat{\varphi}(A_j)$ schließt. Damit ist aber

$$\det(B) = \textstyle\sum_j (-1)^j b_{1j} \det(B_j) = \sum_j (-1)^j \varphi(a_{1j}) \det(\hat{\varphi}(A_j)).$$

Nach Induktionsvoraussetzung ist $\det(\hat{\varphi}(A_j)) = \varphi(\det(A_j))$. Da φ ein Ringhomomorphismus ist und $\varphi(1) = 1$ erfüllt, folgt $\det(B) = \varphi(\det(A))$. □

Zusammen mit Teil (2) von Beispiel 6.2 können wir jetzt die Motivation dafür sehen, das wir Determinanten über kommutativen Ringen mit Eins studieren. Betrachten wir einen Körper $\mathbb{K}$ und für $i, j = 1, \dots, n$ sei $p_{ij} \in \mathbb{K}[x]$ ein Polynom. Dann können wir die Matrix $A := (p_{ij}) \in M_n(\mathbb{K}[x])$ betrachten und ihre Determinante $\det(A)$ bilden, die in $\mathbb{K}[x]$ liegt, also ein Polynom ist. Wählt man nun ein fixes Element $\lambda \in \mathbb{K}$ dann können wir die Matrix $(p_{ij}(\lambda)) \in M_n(\mathbb{K})$ bilden. Natürlich ist diese Matrix gerade $\hat{\varphi}(A)$, wobei $\varphi(p) := p(\lambda)$ der Homomorphismus aus Teil (2) von Beispiel 6.2 ist. Damit folgt aus dem Korollar sofort, dass man die Determinante der Matrix $(p_{ij}(\lambda))$ erhält, wenn man λ in das Polynom $\det(A)$ einsetzt.

6.5. Die Cramer'sche Regel. Aus den elementaren Eigenschaften der Determinantenfunktion det können wir sofort eine Anwendung auf lineare Gleichungssysteme folgern:

Satz 6.5. *Ist R ein kommutativer Ring mit Einselement, $A \in M_n(R)$ eine $n \times n$–Matrix mit Spaltenvektoren $v_1, \dots, v_n$ und $b \in R^n$, dann gilt: Ist $x = (x_1, \dots, x_n) \in R^n$ eine Lösung des linearen Gleichungssystems $Ax = b$, dann ist*

$$x_j \det(v_1, \dots, v_n) = \det(v_1, \dots, v_{j-1}, b, v_{j+1}, \dots, v_n)$$

Ist insbesondere $\det(A) \in R$ *kein Nullteiler, dann gibt es für jedes $b \in R^n$ höchstens eine Lösung des linearen Gleichungssystems $Ax = b$.*

Beweis. In Termen der Spaltenvektoren $v_1, \dots, v_n$ von A bedeutet $Ax = b$ gerade $x_1v_1 + \dots + x_nv_n = b$, vergleiche mit 1.3. Betrachten wir nun $\det(v_1, \dots, b, \dots, v_n)$, wobei wir b anstelle von v_j eingesetzt haben, dann erhalten wir

$$\det(v_1, \dots, b, \dots, v_n) = \det(v_1, \dots, x_1v_1 + \dots + x_nv_n, \dots, v_n).$$

Nach der Linearität in der $j$ten Spalte ist das gleich $\sum_{i=1}^n x_i \det(v_1, \dots, v_i, \dots, v_n)$, wobei v_i an der $j$ten Stelle steht. Für $i \neq j$ kommt aber in der Determinante zwei mal der selbe Spaltenvektor vor (nämlich v_i an der $i$ten und der $j$ten Stelle), also bleibt nur der $j$te Summand übrig, indem in der $j$ten Spalte v_j steht. Damit folgt die erste Behauptung des Satzes sofort.

Sind $x_1, \dots, x_n$ und $\tilde{x}_1, \dots, \tilde{x}_n$ zwei Lösungen von $Ax = b$ dann folgt natürlich sofort, dass $x_i \det(A) = \tilde{x}_i \det(A)$ und damit $(x_i - \tilde{x}_i)\det(A) = 0$ für alle $i = 1, \dots, n$ gilt. Ist $\det(A) \in R$ kein Nullteiler, dann ist das nur für $x_i - \tilde{x}_i = 0$ also $x_i = \tilde{x}_i$ möglich. □

Im Fall eines Körpers erhalten wir leicht eine Folgerung, die wir später auch auf Ringe verallgemeinern werden.

Korollar 6.5. *Sei $\mathbb{K}$ ein Körper und $A \in M_n(\mathbb{K})$ eine Matrix mit* $\det(A) \neq 0$. *Dann sind die Spaltenvektoren v_i von A linear unabhängig, A ist invertierbar, und für jedes $b \in \mathbb{K}^n$ ist die eindeutige Lösung $x = (x_1, \dots, x_n)$ des linearen Gleichungssystems $Ax = b$ gegeben durch*

$$x_j = \frac{\det(v_1, \dots, v_{j-1}, b, v_{j+1}, \dots, v_n)}{\det(A)}$$

Beweis. Nach dem letzten Teil des Satzes ist die Lösung von $Ax = b$ für jedes $b \in \mathbb{K}^n$ eindeutig bestimmt. Nach Satz 4.14 ist damit A invertierbar und es gibt für jedes $b \in \mathbb{K}^n$ eine Lösung von $Ax = b$. Da $0 \neq \det(A)$ in $\mathbb{K}$ ein inverses Element besitzt, folgt die explizite Form der Lösung sofort aus dem Satz. □

Bemerkung 6.5. Für die praktische Lösung von linearen Gleichungssystemen ist die Cramer'sche Regel nicht besonders gut geeignet, weil die Berechnung der vielen Determinanten sehr aufwändig ist. Daher lösen Computerprogramme lineare Gleichungssysteme nicht mit dieser Methode. Um das zu sehen, bestimmen wir, wie viele Rechenoperationen zur Lösung eines linearen Gleichungssystems mit n Gleichungen in n Unbekannten nötig sind. Wir werden uns dabei auf die Multiplikationen und Divisionen beschränken (die mehr Rechenaufwand benötigen) und Additionen, Subtraktionen und Vertauschungen von Zeilen oder Spalten außer Acht lassen. Die Formel für eine 2×2–Determinante besteht aus zwei Summanden, von denen jeder ein Produkt von zwei Zahlen ist. Nach Definition muss man n mal ein Element mit einer $(n-1) \times (n-1)$–Determinante multiplizieren, um eine $n \times n$–Determinante zu berechnen. Eine 3×3–Determinante hat also 6 Summanden, von denen jeder ein Produkt von 3 Zahlen ist, man benötigt also 12 Multiplikationen für so eine Determinante. Induktiv sieht man, dass eine $n \times n$–Determinante aus $n! = n(n-1) \cdots 3 \cdot 2$ Summanden besteht, von denen jeder ein Produkt von n Zahlen ist, also $n!(n-1)$ Multiplikationen benötigt.

Um ein $n \times n$–Gleichungssystem nach der Cramer'schen Regel zu berechnen, muss man $n+1$ Determinanten von $n \times n$–Matrizen berechnen, benötigt als $(n+1)n!(n-1) = (n+1)!(n-1)$ Multiplikationen. Anschließend braucht man noch n Divisionen, was aber kaum mehr ins Gewicht fällt. Berechnet man das explizit, dann braucht man für ein 3×3–System 48 Multiplikationen, ein 4×4–System benötigt 360 Multiplikationen, für 5×5–Systeme sind 2880 Multiplikationen nötig und bei einem 10×10–System sind es schon 359251200 Multiplikationen.

Versuchen wir auf ähnliche Weise den Gauß'schen Algorithmus zu analysieren, dann müssen wir nur bedenken, dass man nach Satz 3.9 jede invertierbare $n \times n$–Matrix (der einzige Fall in dem die Cramer'sche Regel funktioniert) durch elementare Zeilenoperationen in die Einheitsmatrix umgewandelt werden kann, was genau äquivalent zur Lösung des Systems ist. Betrachten wir also die erweiterte Matrix (A, b) des Systems. Da wir Vertauschungen von Zeilen ignorieren, dürfen wir annehmen, dass das die erste Zeile von A mit einem Element ungleich Null beginnt. Mit n Divisionen erreichten wir, dass die erste Zeile mit 1 beginnt. Um in den weiteren Zeilen das Element in der ersten Spalte zu Null zu machen, benötigen wir pro Zeile n Multiplikationen. Somit sind wir mit n Divisionen und $(n-1)n$ Multiplikationen so weit, dass in der ersten Spalte der erweiterten Matrix der erste Einheitsvektor e_1 steht. Damit das zweite Element der zweiten Zeile gleich Eins ist, brauchen wir $(n-1)$ Divisionen, und mit $(n-1)(n-1)$ Multiplikationen erreichen wir, dass in der zweiten Spalte der zweite Einheitsvektor e_2 steht. Induktiv sehen wir, dass wir $n + (n-1) + \cdots + 2 + 1 = \frac{1}{2}n(n+1)$ Divisionen und $(n-1)(n + (n-1) + \cdots + 2 + 1) = \frac{1}{2}(n-1)n(n+1)$ Multiplikationen benötigen, und ein $n \times n$–Gleichungssystem mit dem Gauß'schen Algorithmus zu lösen. Für ein

3×3–System sind das 6 Divisionen und 12 Multiplikationen, bei einem 4×4–System 10 Divisionen und 30 Multiplikationen und bei einem 5×5–System 15 Divisionen und 60 Multiplikationen. Bei einem 10×10–System erhält man die moderate Zahl von 55 Divisionen und 495 Multiplikationen.

Die Cramer'sche Regel ist aber von einem theoretischen Standpunkt aus interessant, weil sie in einfacher Weise beschreibt, wie die eindeutige Lösung des linearen Gleichungssystems $Ax = b$ (für invertierbares A) von der Matrix A und vom Vektor b abhängt, siehe Bemerkung 6.12

Eindeutigkeit der Determinante und die Leibniz–Formel

Unser nächstes Hauptziel ist zu zeigen, dass es nur eine normierte Determinantenfunktion gibt. Dazu müssen wir einige Grundtatsachen über die sogenannten Permutationsgruppen, also die Gruppen von Bijektionen einer endlichen Mengen beweisen.

6.6. Permutationen. Für $n \in \mathbb{N}$ bezeichnen wir mit $\mathfrak{S}_n$ die Menge aller Bijektionen der Menge $\{1, \dots, n\}$. Für zwei bijektive Funktionen $\sigma, \tau \in \mathfrak{S}_n$ ist natürlich auch die Komposition $\sigma \circ \tau$ eine Bijektion, liegt also wieder in $\mathfrak{S}_n$. Wir werden im weiteren meist einfach $\sigma\tau$ statt $\sigma \circ \tau$ schreiben. Zu $\sigma \in \mathfrak{S}_n$ ist auch die inverse Funktion σ^{-1} bijektiv und liegt damit ebenfalls in $\mathfrak{S}_n$. Nach Definition gilt $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \mathrm{id}$ und $\sigma \circ \mathrm{id} = \mathrm{id} \circ \sigma = \sigma$ und da die Komposition von Funktionen immer assoziativ ist, ist $\mathfrak{S}_n$ eine Gruppe mit neutralem Element id. Diese Gruppe ist allerdings *nicht* kommutativ.

Ist $\sigma \in \mathfrak{S}_n$, also $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ eine bijektive Funktion, dann können wir σ einfach beschreiben, indem wir das n–Tupel $(\sigma(1), \dots, \sigma(n))$ angeben. Klarerweise ist eine Funktion $f : \{1, \dots, n\} \to \{1, \dots, n\}$ genau dann injektiv, wenn sie surjektiv ist. Die möglichen n–Tupel zu Bijektionen $\sigma \in \mathfrak{S}_n$ sind also genau die, in denen jede der Zahlen $\{1, \dots, n\}$ einmal vorkommt. Somit hat man aber n Möglichkeiten $\sigma(1)$ zu wählen, es bleiben $(n - 1)$ Möglichkeiten für $\sigma(2)$, und so weiter bis zu $\sigma(n)$, das dann schon eindeutig festgelegt ist. Daraus folgt aber, dass $\mathfrak{S}_n$ genau $n!$ Elemente besitzt.

Schließlich bemerken wir noch, dass wir $\mathfrak{S}_{n-1}$ in natürlicher Weise als Teilmenge von $\mathfrak{S}_n$ betrachten können. Betrachten wir nämlich eine Permutation $\sigma \in \mathfrak{S}_n$, für die $\sigma(n) = n$ gilt, dann ist $\sigma(i) \in \{1, \dots, n - 1\}$ für alle $i \in \{1, \dots, n - 1\}$, also liefert σ eine Funktion $\{1, \dots, n - 1\} \to \{1, \dots, n - 1\}$, und diese ist offensichtlich injektiv, also bijektiv, also ein Element von $\mathfrak{S}_{n-1}$. Umgekehrt kann man aus jedem $\sigma \in \mathfrak{S}_{n-1}$ eine Bijektion von $\{1, \dots, n\}$ machen, indem man n auf sich selbst abbildet. Natürlich ist der Wechsel zwischen den beiden Bildern mit der Komposition verträglich, also ist $\{\sigma \in \mathfrak{S}_n : \sigma(n) = n\}$ eine *Untergruppe* von $\mathfrak{S}_n$, die wir mit $\mathfrak{S}_{n-1}$ identifizieren können.

Für $n = 1$ gibt es natürlich nur eine Bijektion der Menge $\{1\}$, nämlich die Identität. Für $n = 2$ gibt es neben der Identität nur noch eine Bijektion, die dem geordneten Paar $(2, 1)$ entspricht und gerade die beiden Elemente vertauscht. Für $n = 3$ entsprechen die 6 möglichen Bijektionen gerade den Tripeln $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$ und $(3, 2, 1)$, und so weiter. Das erste und das dritte Tripel entsprechen gerade den beiden Elementen der Untergruppe $\mathfrak{S}_2 \subset \mathfrak{S}_3$.

Besonders einfache Permutationen sind die sogenannten *Transpositionen* oder Vertauschungen. Für $n \in \mathbb{N}$ und $i, j \in \{1, \dots, n\}$ mit $i \neq j$ definieren wir die Transposition $(i, j) \in \mathfrak{S}_n$ als die bijektive Abbildung die i auf j, j auf i und alle anderen Elemente von $\{1, \dots, n\}$ auf sich selbst abbildet. (i, j) vertauscht also gerade die Elemente i und j. Offensichtlich ist $(i, j)^{-1} = (i, j)$. Was wir nun benötigen ist, dass man jeder Permutation σ ein sogenanntes *Signum* $\operatorname{sgn}(\sigma) \in \{1, -1\}$ zuordnen kann, sodass $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$ gilt und so, dass das Signum einer Transposition -1 ist. Zunächst beweisen wir:

Satz 6.6. *Sei $n \in \mathbb{N}$, $n \geq 2$. Dann gilt:*

(1) Man kann jede Permutation $\sigma \in \mathfrak{S}_n$ als Produkt von höchstens n Transpositionen schreiben.

(2) Sind τ_i für $i = 1, \dots, k$ und τ'_j für $j = 1, \dots, \ell$ Transpositionen und ist $\tau_1 \cdots \tau_k = \tau'_1 \cdots \tau'_\ell$, dann sind die Zahlen k und ℓ entweder beide gerade oder beide ungerade.

Beweis. (1) Wir verwenden Induktion nach n. Für $n = 2$ haben wir nur die Transposition $(1, 2)$ und die Identität. Wie wir aber bereits bemerkt haben, ist $(1, 2)^{-1} = (1, 2)$, also $(1, 2)(1, 2) = \mathrm{id}$. Nehmen wir induktiv an, dass $n \geq 3$ ist, und wir den Satz für $\mathfrak{S}_{n-1}$ bereits bewiesen haben. Zu $\sigma \in \mathfrak{S}_n$ betrachte $\sigma(n) \in \{1, \dots, n\}$. Ist $\sigma(n) = n$, dann ist σ ein Element der Untergruppe $\mathfrak{S}_{n-1} \subset \mathfrak{S}_n$, kann also nach Induktionsvoraussetzung als Produkt von höchstens $n - 1$ Transpositionen geschrieben werden. Ist andererseits $\sigma(n) = k \neq n$, dann betrachten wir $(k, n)\sigma$. Das bildet n auf n ab, also gibt nach Induktionsvoraussetzung eine Zahl $s \leq n-1$ und Transpositionen $\tau_1, \dots, \tau_s$, sodass $(k, n)\sigma = \tau_1 \cdots \tau_s$ gilt. Damit ist aber $\sigma = (k, n)(k, n)\sigma = (k, n)\tau_1 \cdots \tau_s$ ein Produkt von höchstens n Transpositionen.

(2) Sei $\{e_1, \dots, e_n\}$ die Standardbasis für den $\mathbb{R}$–Vektorraum $\mathbb{R}^n$, sei $\sigma = \tau_1 \cdots \tau_k = \tau'_1 \cdots \tau'_\ell$ und betrachte $\det(e_{\sigma(1)}, \dots, e_{\sigma(n)})$, wobei det die normierte Determinantenfunktion aus Satz 6.4 bezeichnet. Nach Konstruktion ist $\det(e_1, \dots, e_n) = 1$, und nach Bemerkung (2) von 6.4 wechselt die Determinante das Vorzeichen, wenn man zwei Spalten vertauscht. Damit folgt aber aus $\sigma = \tau_1 \cdots \tau_k$ sofort, dass $\det(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = (-1)^k$ gelten muss. Analog folgt aus $\sigma = \tau'_1 \cdots \tau'_\ell$, dass diese Determinante gleich $(-1)^\ell$ sein muss, also folgt die Behauptung. □

Bemerkung 6.6. (1) Die Darstellung als Produkt von Transpositionen ist nicht eindeutig. So gilt etwa in $\mathfrak{S}_3$ zum Beispiel trivialerweise $(1, 3) = (1, 2)(1, 2)(1, 3)$ aber auch die nichttriviale Gleichung $(1, 2)(2, 3)(1, 2) = (1, 3)$.

(2) Die induktive Methode aus dem Beweis von Teil (1) des Satzes kann man verwenden, um eine explizite Darstellung einer gegebenen Permutation als Produkt von Transpositionen zu erhalten. Betrachten wir etwa die Permutation $\sigma \in \mathfrak{S}_4$, die gegeben ist durch $\sigma(1) = 3$, $\sigma(2) = 4$, $\sigma(3) = 2$ und $\sigma(4) = 1$. Betrachten wir nun $(1, 4)\sigma$, dann bildet das 1 auf 3, 2 auf 1, 3 auf 2 und 4 auf 4 ab. Damit vertauscht $(2, 3)(1, 4)\sigma$ die Elemente 2 und 1 und lässt 3 und 4 fix, also ist $(2, 3)(1, 4)\sigma = (1, 2)$ und somit $\sigma = (1, 4)(2, 3)(1, 2)$.

(3) Bei Rechnungen mit Permutationsgruppen ist immer Vorsicht geboten, weil diese Gruppen nicht kommutativ sind. Betrachten wir etwa $\mathfrak{S}_3$, dann bildet $(1, 2)(2, 3)$ das Element 3 auf 1 ab, während $(2, 3)(1, 2)$ klarerweise 3 auf 2 abbildet.

Definition 6.6. Sei $\sigma \in \mathfrak{S}_n$ eine Permutation. Dann definieren wir das *Signum* $\operatorname{sgn}(\sigma) \in \{-1, 1\}$ von σ als $(-1)^k$ falls es eine Darstellung von σ als Produkt von k Transpositionen gibt. Nach Teil (1) des Satzes kann man σ immer als Produkt von Transpositionen schreiben und nach Teil (2) des Satzes ist $\operatorname{sgn}(\sigma)$ wohldefiniert, weil sich für verschiedene Darstellungen der selben Permutation immer der selbe Wert für $\operatorname{sgn}(\sigma)$ ergibt.

Wir erhalten nun leicht die Eigenschaften der Signumfunktion:

PROPOSITION 6.6. *Für $n \geq 2$ hat die Signumfunktion* $\operatorname{sgn} : \mathfrak{S}_n \to \{-1, 1\}$ *folgende Eigenschaften:*

(1) sgn *ist ein Gruppenhomomorphismus, d.h.* $\operatorname{sgn}(\sigma\sigma') = \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma')$.

(2) Ist τ eine Transposition, dann ist $\operatorname{sgn}(\tau) = -1$.

(3) $\operatorname{sgn}(\mathrm{id}) = 1$ *und für jedes $\sigma \in \mathfrak{S}_n$ ist* $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$.

BEWEIS. (1) Sind $\sigma = \tau_1 \cdots \tau_k$ und $\sigma' = \tau'_1 \cdots \tau'_\ell$ Darstellungen von σ und σ' als Produkte von Transpositionen, dann ist $\tau_1 \cdots \tau_k \cdot \tau'_1 \cdots \tau'_\ell$ eine Darstellung von $\sigma\sigma'$ als Produkt von Transpositionen. Damit ist aber nach Definition $\operatorname{sgn}(\sigma\sigma') = (-1)^{k+\ell} = (-1)^k(-1)^\ell = \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma')$.

(2) folgt direkt aus der Definition.

(3) Weil $\mathrm{id} = (1,2)(1,2)$ gilt ist $\operatorname{sgn}(\mathrm{id}) = 1$, also gilt $\operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\mathrm{id}) = 1$ nach Teil (1), also ist $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$. □

6.7. Eindeutigkeit der Determinante. Mit den obigen Resultaten über Permutation können wir beweisen, dass es nur eine normierte Determinantenfunktion gibt. Gleichzeitig erhalten wir eine allgemeine Formel für die Determinante, die sogenannte Leibniz–Formel, die zwar für die praktische Berechnung von Determinanten zu kompliziert ist, aber theoretisch in vielen Fällen nützlich ist. Aus diesen beiden Ergebnissen werden wir rasch alle weiteren Resultate über Determinanten erhalten.

Wir können nun nämlich sofort das Verhalten von Determinantenfunktionen unter Permutation der Spaltenvektoren beschreiben. Ist $F : M_n(R) \to R$ eine Determinantenfunktion und $\sigma \in \mathfrak{S}_n$ eine Permutation, dann kann man σ als Produkt von Transpositionen schreiben, und bei Vertauschung von zwei Spalten wechselt F nach Bemerkung (2) von 6.4 das Vorzeichen. Somit ist aber $F(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \operatorname{sgn}(\sigma)F(v_1, \dots, v_n)$ für beliebige Elemente $v_1, \dots, v_n \in R^n$.

SATZ 6.7. *Sei R ein kommutativer Ring mit Einselement, $F : M_n(R) \to R$ eine Determinantenfunktion und $\mathbb{I} \in M_n(R)$ die $n \times n$–Einheitsmatrix. Dann gilt für $A = (a_{ij}) \in M_n(R)$ die Formel*

$$F(A) = F(\mathbb{I}) \textstyle\sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}.$$

Insbesondere ist $\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}$ *und* $F(A) = F(\mathbb{I})\det(A)$.

BEWEIS. Seien $e_1, \dots, e_n$ die Einheitsvektoren in R^n. Dann kann man den $i$ten Spaltenvektor von A als $a_{1i}e_1 + \dots + a_{ni}e_n$ schreiben. Somit ist

$$F(A) = F(a_{11}e_1 + \dots + a_{n1}e_n, \dots, a_{1n}e_1 + \dots + a_{nn}e_n).$$

Wegen der Linearität von F in jeder Spalte zerfällt das in eine riesige Summe wobei in jedem Summanden in jeder Spalte nur mehr ein e_j steht. Ist $f : \{1, \dots, n\} \to \{1, \dots, n\}$ definiert dadurch, dass in der $j$ten Spalte gerade $e_{f(j)}$ steht, dann erhalten wir

$$F(A) = \textstyle\sum_f a_{f(1)1} \dots a_{f(n)n} F(e_{f(1)}, \dots, e_{f(n)}),$$

wobei die Summe über alle Funktionen $f : \{1, \dots, n\} \to \{1, \dots, n\}$ geht. Ist nun f nicht injektiv, dann kommen im entsprechenden Summanden zwei gleiche Vektoren in zwei verschiedene Spalten von F vor, also verschwindet der entsprechende Summand. Daher müssen wir nur über die bijektiven Funktionen, also über $\mathfrak{S}_n$, summieren. Für $\sigma \in \mathfrak{S}_n$ haben wir aber bereits festgestellt, dass $F(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = \operatorname{sgn}(\sigma)F(e_1, \dots, e_n)$ gilt. Offensichtlich ist $F(e_1, \dots, e_n) = F(\mathbb{I})$ also folgt die erste Formel für $F(A)$. Wegen $\det(\mathbb{I}) = 1$ erhalten wir daraus sofort die Formel für $\det(A)$ und daraus folgt sofort die zweite Formel für $F(A)$. □

Dieser Satz liefert nun eine vollständige Beschreibung aller Determinantenfunktionen $F : M_n(R) \to R$. Ist nämlich $r \in R$ beliebig, dann ist offensichtlich $A \mapsto r\det(A)$ eine Determinantenfunktion, und nach dem Satz ist jede Determinantenfunktion von dieser Form, wobei $r = F(\mathbb{I})$ gilt.

6.8. Determinante und Produkte. Aus der Eindeutigkeit der Determinante können wir leicht schließen, dass die Determinante mit Produkten verträglich ist.

Satz 6.8. *Sei R ein kommutativer Ring mit Einselement und $F : M_n(R) \to R$ eine Determinantenfunktion. Dann gilt $F(Av_1, \dots, Av_n) = \det(A)F(v_1, \dots, v_n)$ für alle $A \in M_n(R)$ und $v_1, \dots, v_n \in R^n$. Insbesondere gilt $\det(AB) = \det(A)\det(B)$ für alle $A, B \in M_n(R)$.*

Beweis. Betrachte die Funktion $F_A(v_1, \dots, v_n) := F(Av_1, \dots, Av_n)$. Da $v \mapsto Av$ linear und F linear in jeder Spalte ist, ist auch F_A linear in jeder Spalte. Ist $v_i = v_j$ für $i \neq j$, dann ist $Av_i = Av_j$, also ist $F_A(v_1, \dots, v_n) = 0$, weil F eine Determinantenfunktion ist. Nach Satz 6.7 ist damit $F_A(v_1, \dots, v_n) = F_A(e_1, \dots, e_n)\det(v_1, \dots, v_n)$. Nun ist aber nach Definition $F_A(e_1, \dots, e_n) = F(Ae_1, \dots, Ae_n)$, und $Ae_1, \dots, Ae_n$ sind gerade die Spaltenvektoren von A, also ist $F_A(e_1, \dots, e_n) = F(A)$. Nochmals nach Satz 6.7 ist $F(A) = F(\mathbb{I})\det(A)$ und $F(\mathbb{I})\det(v_1, \dots, v_n) = F(v_1, \dots, v_n)$ und wir erhalten $F_A(v_1, \dots, v_n) = \det(A)F(v_1, \dots, v_n)$.

Sind $v_1, \dots, v_n$ die Spaltenvektoren einer Matrix B, dann sind $Av_1, \dots, Av_n$ die Spaltenvektoren von AB und wir erhalten

$$\det(AB) = \det(Av_1, \dots, Av_n) = \det(A)\det(v_1, \dots, v_n) = \det(A)\det(B).$$

□

Diese Produktformel hat einige wichtige Konsequenzen:

Korollar 6.8. *(1) Ist $A \in M_n(R)$ invertierbar, dann besitzt $\det(A) \in R$ ein multiplkativ inverses Element (in R).*

(2) Sind $A, T \in M_n(R)$ Matrizen, sodass T invertierbar ist, dann ist $\det(TAT^{-1}) = \det(A)$.

(3) Im Fall eines Körpers $\mathbb{K}$ haben ähnliche Matrizen $A, B \in M_n(\mathbb{K})$ die gleiche Determinante. Damit kann man für einen endlichdimensionalen $\mathbb{K}$–Vektorraum V und eine lineare Abbildung $f : V \to V$ die Determinante $\det(f) \in \mathbb{K}$ von f als die Determinante einer beliebigen Matrixdarstellung von f definieren.

Beweis. (1) Nach dem Satz ist $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(\mathbb{I}_n) = 1$.

(2) Es gilt $\det(TAT^{-1}) = \det(T)\det(A)\det(T^{-1})$. Das ist aber ein Produkt in R, also kann man die Reihenfolge vertauschen und nach Teil (1) ist $\det(T)\det(T^{-1}) = 1$.

(3) Das folgt sofort aus (2) und den Überlegungen in 6.1. □

Bemerkung 6.8. Aus Bemerkung 3.4 wissen wir, dass die Menge $GL(n, \mathbb{K})$ aller invertierbaren $n \times n$–Matrizen über einem Körper $\mathbb{K}$ eine Gruppe unter der Matrizenmultiplikation bildet. Nach Korollar 6.5 und Korollar 6.8 liegt eine Matrix A genau dann in $GL(n, \mathbb{K})$, wenn $\det(A) \neq 0$ gilt. Betrachten wir nun die Teilmenge $SL(n, \mathbb{K}) := \{A \in GL(n, \mathbb{K}) : \det(A) = 1\}$. Dann ist natürlich $\mathbb{I} \in SL(n, \mathbb{K})$ und für $A, B \in SL(n, \mathbb{K})$ liegen nach der Produktformel auch AB und A^{-1} in $SL(n, \mathbb{K})$. Das bedeutet aber gerade, dass $SL(n, \mathbb{K})$ eine Untergruppe der Gruppe $GL(n, \mathbb{K})$ ist, sie sogenannte *spezielle lineare Gruppe*. Man kann das auch elegant dadurch sehen, dass die Produktformel bedeutet, dass det ein Homomorphismus von $GL(n, \mathbb{K})$ in die Gruppe $(\mathbb{K} \setminus \{0\}, \cdot)$ ist. Nach Definition besteht $SL(n, \mathbb{K})$ genau aus jenen Elementen, die durch diesen Homomorphismus auf das neutrale Element 1 abgebildet werden, ist also der Kern dieses Homomorphismus. Damit ist $SL(n,\mathbb{K})$ sogar eine *normale Untergruppe* in $GL(n,\mathbb{K})$, d.h. für $A \in SL(n,\mathbb{K})$ und $B \in GL(n,\mathbb{K})$ gilt $BAB^{-1} \in SL(n,\mathbb{K})$.

Für spezielle Körper liefert die Produktformel weitere Untergruppen von $GL(n,\mathbb{K})$. So erhält man etwa für $\mathbb{K} = \mathbb{R}$ die Gruppe $\{A \in GL(n,\mathbb{R}) : \det(A) > 0\}$ und für $\mathbb{K} = \mathbb{C}$ die Gruppe $\{A \in GL(n,\mathbb{C}) : |\det(A)| = 1\}$.

6.9. Determinante der Transponierten, Entwicklungssätze. Betrachten wir die Leibnizformel $\det(A) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}$ aus Satz 6.7. In jedem Summanden kommt nach Definition genau ein Eintrag aus jeder Spalte vor. Weil aber σ eine Bijektion ist, kommt auch aus jeder Zeile genau ein Eintrag vor, und das lässt eine Symmetrie zwischen Zeilen und Spalten vermuten. Aus Definition 5.12 kennen wir (für Körper) die Operation des Transponierens einer Matrix, die Zeilen und Spalten vertauscht. Nach Definition ist für $A = (a_{ij}) \in M_n(R)$ die transponierte Matrix $A^t = (b_{ij}) \in M_n(R)$ gegeben durch $b_{ij} := a_{ji}$. Dies bedeutet gerade, dass der $k$te Spaltenvektor von A^t gerade mit dem $k$ten Zeilenvektor von A übereinstimmt, wenn man beide einfach als Elemente von R^n betrachtet, und umgekehrt (wegen $(A^t)^t = A$).

Satz 6.9. *Sei R ein kommutativer Ring mit Einselement. Dann gilt* $\det(A) = \det(A^t)$ *für alle $A \in M_n(R)$.*

Beweis. Nach Satz 6.7 ist $\det(A^t) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)}$. Weil σ bijektiv ist, kann man das Produkt $a_{1\sigma(1)} \dots a_{n\sigma(n)}$ als $a_{\sigma^{-1}(1)1} \dots a_{\sigma^{-1}(n)n}$ schreiben. Nach Proposition 6.6 ist $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$, also folgt $\det(A^t) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma^{-1}) a_{\sigma^{-1}(1)1} \dots a_{\sigma^{-1}(n)n}$. Wenn σ durch alle Permutationen in $\mathfrak{S}_n$ läuft, dann kommt als σ^{-1} jedes Element von $\mathfrak{S}_n$ genau ein mal vor, also können wir die Summe auch als $\sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} = \det(A)$ schreiben. □

In der Definition von det haben wir die Berechnung von $n \times n$–Determinanten auf die Berechnung von $(n-1) \times (n-1)$–Determinanten zurückgeführt, wobei wir durch die Elemente der ersten Zeile gelaufen sind und diese Zeile, sowie jeweils eine Spalte der verbleibenden $(n-1) \times n$–Matrix weggelassen haben ("Entwicklung nach der ersten Zeile"). Mit Hilfe unserer jetzigen Resultate können wir leicht zeigen, dass man nach jeder beliebigen Zeile und auch nach jeder beliebigen Spalte entwickeln kann. Außerdem können wir den Zusammenhang der Determinante mit elementaren Zeilenoperationen klären, was uns auch eine effektive Methode zur Berechnung von Determinanten liefert. Dazu benötigen wir noch eine Notation. Für $A \in M_n(R)$ sei $A_{ij} \in M_{n-1}(R)$ die Matrix, die entsteht, wenn man in A die $i$te Zeile und die $j$te Spalte weglässt. Schließlich finden wir noch eine Klasse von Matrizen, für die wir die Determinante sofort ablesen können:

Definition 6.9. Eine Matrix $A = (a_{ij}) \in M_n(R)$ heißt *obere* (bzw. untere) *Dreiecksmatrix* falls $a_{ij} = 0$ für alle $i < j$ (bzw. für alle $i > j$) gilt.

Korollar 6.9. *(1) Die Funktion* $\det : M_n(R) \to R$ *ist linear in jeder Zeile der Matrix und verschwindet, falls zwei Zeilen gleich sind. Addiert man zu einer Zeile ein Vielfaches einer anderen Zeile, dann bleibt der Wert der Determinante unverändert. Vertauscht man zwei Zeilen, dann wechselt die Determinante ihr Vorzeichen.*

(2) Für $A = (a_{ij}) \in M_n(R)$ gelten die Formeln

$$\det(A) = \textstyle\sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}) \qquad \text{“\textit{Entwicklung nach der $i$ten Zeile}”}$$

$$\det(A) = \textstyle\sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}) \qquad \text{“\textit{Entwicklung nach der $j$ten Spalte}”}$$

(3) Ist $A = (a_{ij})$ eine obere oder untere Dreiecksmatrix, dann ist $\det(A) = \prod_{i=1}^n a_{ii}$.

Beweis. (1) Da $\det(A) = \det(A^t)$ gilt, folgt die Linearität in jeder Zeile aus der Linearität in jeder Spalte. Sind in A zwei Zeilen gleich, dann sind in A^t zwei Spalten gleich, also gilt $\det(A^t) = 0$, also $\det(A) = 0$. Daraus folgt sofort, dass sich die Determinante nicht ändert, wenn man zu einer Zeile ein Vielfaches einer anderen Zeile addiert. Vertauschen von zwei Zeilen in A entspricht genau Vertauschen der entsprechenden Spalten in A^t, also wechselt die Determinante ihr Vorzeichen.

(2) Im Beweis von Satz 6.4 haben wir det durch $\det(A) = \sum_{j=1}^n (-1)^{j+1} a_{1j} \det(A_{1j})$ definiert. Sei $\tilde{A} = (\tilde{a}_{ij})$ die Matrix, die aus A entsteht, indem man die $i$te Zeile ganz nach oben tauscht. Das kann man realisieren, indem man die $i$te Zeile mit der $(i-1)$-ten, dann mit der $(i-2)$ten und so weiter bis zur ersten Zeile vertauscht, also ist $\det(\tilde{A}) = (-1)^{i-1} \det(A)$ nach Teil (1). Andererseits ist $\tilde{A}_{1j} = A_{ij}$ für alle $j = 1, \dots, n$. Damit erhalten wir

$$\det(\tilde{A}) = \sum_{j=1}^n (-1)^{j+1} \tilde{a}_{1j} \det(\tilde{A}_{1j}) = \sum_{j=1}^n (-1)^{j+1} a_{ij} \det(A_{ij})$$

und somit $\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})$.

Ist andererseits $B = (b_{ij}) = A^t$, dann ist nach der zuletzt bewiesenen Formel $\det(B) = \sum_{j=1}^n (-1)^{i+j} b_{ij} \det(B_{ij})$. Nun gilt aber $\det(B) = \det(A)$ und $b_{ij} = a_{ji}$, sowie $B_{ij} = A^t_{ji}$, also erhalten wir $\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ji} \det(A^t_{ji})$, und wegen $\det(A^t_{ji}) = \det(A_{ji})$ liefert das die letzte behauptete Formel.

(3) Eine obere Dreiecksmatrix hat nach Definition unterhalb der Hauptdiagonale nur Nullen als Eintragungen. Entwickeln wir die Determinante nach der ersten Spalte, dann folgt $\det(A) = a_{11} \det(A_{11})$. Offensichtlich ist aber A_{11} wieder eine obere Dreiecksmatrix mit den Elementen $a_{22}, \dots, a_{nn}$ auf der Hauptdiagonale. Damit folgt die Behauptung sofort mit Induktion. Für eine untere Dreiecksmatrix kann man entweder nach der ersten Zeile entwickeln oder benutzen, dass die transponierte Matrix eine obere Dreiecksmatrix mit den gleichen Eintragungen auf der Hauptdiagonale ist. □

Beispiel 6.9. (1) In manchen Fällen führt die Entwicklung nach einer Zeile oder einer Spalte schnell zur Berechnung einer Determinante. Betrachten wir etwa die Matrix $A = \begin{pmatrix} 5 & 0 & 3 & 2 \\ 3 & 2 & 4 & 5 \\ 4 & 0 & 0 & 1 \\ 6 & 0 & 1 & 3 \end{pmatrix} \in M_4(\mathbb{R})$. Entwickeln nach der zweiten Spalte liefert $\det(A) = 2 \det \begin{pmatrix} 5 & 3 & 2 \\ 4 & 0 & 1 \\ 6 & 1 & 3 \end{pmatrix}$. Entwickelt man diese Determinante nach der zweiten Zeile, dann erhält man $\det(A) = -8 \det \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix} - 2 \det \begin{pmatrix} 5 & 3 \\ 6 & 1 \end{pmatrix} = -8 \cdot 7 + 2 \cdot 13 = -30$.

(2) In den meisten Fällen ist der beste Weg zur Berechnung einer Determinante die Anwendung von elementaren Zeilen– oder Spaltenoperationen. Betrachten wir als Beispiel die Matrix $A = \begin{pmatrix} 1 & 1 & -2 & 4 \\ 0 & 1 & 1 & 3 \\ 2 & -1 & 1 & 0 \\ 3 & 1 & 2 & 5 \end{pmatrix}$. Subtrahieren wir von der dritten Zeile das doppelte der ersten, und von der vierten Zeile das dreifache der ersten, dann ändert das die

Determinante nicht, also ist $\det(A)$ gleich der Determinante von $\begin{pmatrix} 1 & 1 & -2 & 4 \\ 0 & 1 & 1 & 3 \\ 0 & -3 & 5 & -8 \\ 0 & -2 & 8 & -7 \end{pmatrix}$. Addieren wir nun zur dritten Zeile das dreifache der zweiten und zur vierten Zeile das doppelte der zweiten, dann ändert das die Determinante wieder nicht, und wir erhalten die Matrix $\begin{pmatrix} 1 & 1 & -2 & 4 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 8 & 1 \\ 0 & 0 & 10 & -1 \end{pmatrix}$. Subtrahieren wir nun noch von der vierten Zeile $\frac{10}{8}$ mal die dritte Zeile, dann wird diese Zeile zu $(0, 0, 0, -\frac{18}{8})$, und die entstehende Matrix hat immer noch die selbe Determinante wie A. Nun habe wir aber eine obere Dreiecksmatrix vor uns, und nach Teil (3) des Korollars erhalten wir $\det(A) = -18$.

(3) Zum Abschluss wollen wir noch einen wichtigen Spezialfall einer Determinante besprechen, nämlich die sogenannte *Vandermonde–Determinante*. Dazu betrachten wir Elemente $x_1, \dots, x_n \in R$ und bilden die $n \times n$–Matrix

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{pmatrix}$$

Wir behaupten, dass die Determinante dieser Matrix gegeben ist durch $\prod_{i<j}(x_j - x_i)$. Wir beweisen das durch Induktion nach n. Für $n = 2$ ist $\det \begin{pmatrix} 1 & 1 \\ x_1 & x_2 \end{pmatrix} = x_2 - x_1$, also stimmt die Behauptung. Nehmen wir also an, dass wir die Formel für $(n-1) \times (n-1)$–Matrizen bereits bewiesen haben. Nun subtrahieren wir von der letzten ($n$ten) Zeile das x_1–fache der vorletzten ($n - 1$-ten) Zeile. Wir erhalten als neue letzte Zeile $(0, (x_2 - x_1)x_2^{n-2}, \dots, (x_n - x_1)x_n^{n-2})$ und die Determinante bleibt unverändert. Subtrahieren wir nun von der $(n-1)$ten Zeile das x_1–fache der $(n-2)$ten, dann bleibt die Determinante unverändert und die vorletzte Zeile wird zu $(0, (x_2-x_1)x_2^{n-3}, \dots, (x_n-x_1)x_n^{n-3})$. Das setzten wir nun so lange fort, bis wir von der zweiten Zeile das x_1–fache der ersten Zeile abgezogen haben. Dann sehen wir, das unsere ursprüngliche Determinante übereinstimmt mit der Determinante der Matrix

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & (x_2 - x_1) & \dots & (x_n - x_1) \\ 0 & (x_2 - x_1)x_2 & \dots & (x_n - x_1)x_n \\ \vdots & \vdots & & \vdots \\ 0 & (x_2 - x_1)x_2^{n-2} & \dots & (x_n - x_1)x_n^{n-2} \end{pmatrix}.$$

Entwickeln nach der ersten Spalte zeigt nun, dass wir in der Berechnung der Determinante einfach die erste Zeile und die erste Spalte weglassen können. Dann können wir aber aus der (neuen) ersten Spalte $(x_2 - x_1)$ aus der nächsten Spalte $(x_3 - x_1)$ und so weiter bis $(x_n - x_1)$ herausheben und sehen, dass unsere Determinante gegeben ist durch

$$(x_n - x_1) \dots (x_2 - x_1) \det \begin{pmatrix} 1 & \dots & 1 \\ x_2 & \dots & x_n \\ \vdots & & \vdots \\ x_2^{n-2} & \dots & x_n^{n-2} \end{pmatrix}.$$

Nach Induktionsvoraussetzung liefert die verbleibende Determinante $\prod_{2\leq i<j\leq n}(x_j - x_i)$ und somit folgt die Behauptung.

6.10. Determinante und die inverse Matrix. Die Entwicklungssätze für Determinanten erlauben uns im Fall einer Matrix $A \in M_n(R)$, für die $\det(A) \in R$ invertierbar ist, explizit eine inverse Matrix anzugeben. Diese Formel für die Inverse wird oft auch als Cramer'sche Regel bezeichnet.

Satz 6.10. *Sei R ein kommutativer Ring mit Einselement. Für eine Matrix $A = (a_{ij}) \in M_n(R)$ sei $A_{ij} \in M_{n-1}(R)$ die Matrix, die entsteht, wenn man in A die $i$te Zeile und die $j$te Spalte streicht. Definiert man $B = (b_{ij}) \in M_n(R)$ durch $b_{ij} = (-1)^{i+j}\det(A_{ji})$, dann ist $AB = BA = \det(A)\mathbb{I}_n$.*

Beweis. Für das Produkt $AB = (c_{ij})$ gilt nach Definition der Matrizenmultiplikation

$$c_{ij} = \textstyle\sum_k a_{ik}b_{kj} = \sum_k (-1)^{k+j} a_{ik}\det(A_{jk}).$$

Für $i = j$ ist $c_{ii} = \det(A)$ nach der Formel für die Entwicklung der Determinante nach der $i$ten Zeile aus Korollar 6.9 (2). Nehmen wir andererseits an, dass $j \neq i$ gilt. Sei $\tilde{A} = (\tilde{a}_{ij})$ die Matrix, die aus A entsteht, indem man statt der $j$ten Zeile nochmals die $i$te Zeile einsetzt. Dann ist natürlich $\det(\tilde{A}) = 0$, weil $\tilde{A}$ zwei gleiche Zeilen hat. Entwickelt man nun diese Determinante nach der $i$ten Zeile, dann erhält man $0 = \sum_k (-1)^{i+k}\tilde{a}_{ik}\det(\tilde{A}_{ik})$. Nun ist aber $\tilde{a}_{ik} = a_{ik}$. Andererseits unterscheidet sich $\tilde{A}_{ik}$ von A_{jk} nur durch eine Vertauschung von Zeilen, also $\det(\tilde{A}_{ik})$ von $\det(A_{jk})$ nur durch ein Vorzeichen, das außerdem unabhängig von k ist. Damit gilt aber $c_{ij} = 0$ für $i \neq j$, also $AB = \det(A)\mathbb{I}$.

Betrachten wir andererseits das Produkt $BA = (d_{ij})$. Wiederum nach Definition ist $d_{ij} = \sum_k a_{kj}(-1)^{i+k}\det(A_{ki})$. Für $i = j$ gilt wieder $d_{ii} = \det(A)$ nach der Formel für die Entwicklung nach der $i$ten Spalte aus Korollar 6.9 (2). Analog zu den obigen Überlegungen beweist man, dass $d_{ij} = 0$ für $i \neq j$ gilt, indem man die Matrix betrachtet, die aus A entsteht, indem man die $i$te Spalte durch eine Kopie der $j$ten Spalte ersetzt. □

Korollar 6.10. *Sei R ein kommutativer Ring mit Einselement. Dann ist eine Matrix $A \in M_n(R)$ genau dann invertierbar, wenn $\det(A) \in R$ invertierbar ist. Ist das der Fall, dann ist für jedes $b \in R^n$ das lineare Gleichungssystem $Ax = b$ eindeutig lösbar und die Lösung ist wie im Fall eines Körpers durch*

$$x_j = \det(A)^{-1}\det(v_1, \dots, v_{j-1}, b, v_{j+1}, \dots, v_n)$$

gegeben, wobei die v_i die Spaltenvektoren von A sind.

Beweis. Aus Korollar 6.8 wissen wir schon, dass für eine invertierbare Matrix A auch $\det(A) \in R$ invertierbar ist. Hat umgekehrt $\det(A) \in R$ ein multiplikativ inverses Element $\det(A)^{-1}$ dann folgt aus dem Satz sofort, dass die Matrix (c_{ij}) mit $c_{ij} = \det(A)^{-1}(-1)^{i+j}\det(A_{ji})$ invers zu A ist.

Ist A invertierbar, dann hat $Ax = b$ natürlich für jedes $b \in R^n$ die eindeutige Lösung $x = A^{-1}b$. Mit der Formel für A^{-1} von oben folgt damit

$$x_j = \textstyle\sum_k \det(A)^{-1}(-1)^{i+k}\det(A_{kj})b_k.$$

Berechnen wir andererseits $\det(v_1, \dots, v_{j-1}, b, v_{j+1}, \dots, v_n)$ durch Entwickeln nach der $j$ten Spalte. Das liefert eine Summe von Termen, in denen man $(-1)^{j+k}b_k$ mit der Determinante jener Matrix multipliziert, die durch Streichen der $k$ten Zeile und der

$j$ten Spalte entsteht. Aber diese Matrix stimmt offensichtlich mit A_{kj} überein und das Ergebnis folgt. □

6.11. Der Determinantenrang. Man kann den Rang einer Matrix nicht nur über die Maximalzahl linear unabhängiger Spalten ("Spaltentenrang") oder Zeilen ("Zeilenrang") sondern auch mit Hilfe von Determinanten charakterisieren. Bei der Berechnung von Determinanten durch Entwicklung nach einer Zeile oder einer Spalte haben wir schon die Idee angetroffen, aus einer Matrix eine Zeile und eine Spalte zu streichen um eine kleinere Matrix zu konstruieren. Das kann man natürlich auch allgemeiner machen.

Definition 6.11. Sei $A \in M_{m,n}(\mathbb{K})$ eine beliebige rechteckige Matrix mit Eintragungen aus einem Körper $\mathbb{K}$. Ein *$k \times \ell$–Minor* von A ist eine Matrix $B \in M_{k,\ell}(\mathbb{K})$, die man erhält, indem man in A $(m-k)$ Zeilen und $(n-\ell)$ Spalten weglässt.

Damit können wir nun den Rang durch Determinanten charakterisieren.

Proposition 6.11. *Eine Matrix $A \in M_{m,n}(\mathbb{K})$ hat genau dann Rang r, wenn es einen $r \times r$–Minor B von A gibt, der $\det(B) \neq 0$ erfüllt, aber für einen $(r+1)\times(r+1)$–Minor C von A immer $\det(C) = 0$ gilt.*

Beweis. Nehmen wir zunächst an, dass A Rang $\geq r$ hat. Dann finden wir r linear unabhängige Spaltenvektoren von A. Lassen wir die restlichen Spalten weg, dann habe wir einen $m \times r$–Minor von A gefunden, der linear unabhängige Spaltenvektoren besitzt, also ebenfalls Rang r hat. Nach Satz 5.14 wissen wir, dass diese Matrix r linear unabhängige Zeilen hat, und wir bezeichnen mit B den $r \times r$–Minor von A, der durch Weglassen der anderen Zeilen erhalten wird. Dieser hat aber immer noch r linear unabhängige Zeilen, also Rang r, also gilt $\det(B) \neq 0$.

Nemhmen wir umgekehrt an, dass die Matrix A einen $r \times r$–Minor B besitzt, der $\det(B) \neq 0$ erfüllt. Die Wahl des Minors B bestimmt r Spaltenvektoren und r Zeilen von A (die "übrigbleiben"). Versucht man 0 als Linearkombination dieser Spaltenvektoren zu schreiben, so bilden die r ausgewählten Zeilen gerade das Gleichungssystem $Bx = 0$. Damit müssen aber alle Koeffizienten gleich Null sein und somit sind die gegebenen Spaletenvektoren von A linear unabhängig und $\operatorname{rg}(A) \geq r$.

Damit folgt die Behauptung weil der Rang von A genau dann r ist, wenn er $\geq r$ aber nicht $\geq r+1$ ist. □

6.12. Bemerkung. Zum Abschluss möchte ich noch kurz einige Beobachtungen zu den in diesem Kapitel entwickelten Begriffen im Fall $\mathbb{K} = \mathbb{R}$ machen, die insbesondere in den Anwendungen dieser Begriffe in der Analysis wichtig sind (wo lineare Abbildungen insbesondere als Ableitungen auftreten). Wie schon in Abschnitt 3.1 kann man den Raum $M_n(\mathbb{R})$ einfach als $\mathbb{R}^{n^2}$ betrachten, und in diesem Bild sind die Koordinaten einfach durch die Eintragungen der Matrix gegeben. Die Leibnizformel aus 6.7 drückt nun die Determinante $\det(A)$ einfach als eine Summe von Produkten von Matrixeintragungen aus. Damit folgt aber aus elementaren Resultaten der Analysis, dass $\det : \mathbb{R}^{n^2} \to \mathbb{R}$ stetig und sogar beliebig oft differenzierbar ist. Das bedeutet aber, dass für eine Matrix A mit $\det(A) \neq 0$, Matrizen die sehr nahe bei A liegen, ebenfalls Determinante ungleich Null haben müssen. In der Nähe einer invertierbaren Matrix befinden sich also nur invertierbare Matrizen. (Technisch gesagt ist die Menge der invertierbaren Matrizen offen.) Diese Tatsache spielt eine wichtige Rolle beim Beweis des inversen Funktionensatzes in der Analysis.

Für eine invertierbare Matrix $A \in M_n(\mathbb{R})$ und einen Vektor $b \in \mathbb{R}^n$ kann man nach der Cramer'schen Regel aus 6.5 die eindeutige Lösung des Gleichungssystems $Ax = b$

durch Determinanten beschreiben. Das impliziert aber, dass diese eindeutige Lösung sowohl von der Matrix A als auch von dem Vektor b stetig abhängt, was wiederum in vielen Anwendungen wichtig ist. Analog zeigt die Formel für die inverse Matrix aus 6.10, dass die Funktion $A \mapsto A^{-1}$ auf der (offenen) Teilmenge der invertierbaren Matrizen stetig und beliebig oft differenzierbar ist.

Schließlich liefert der Determinantenrang auch noch eine Beschreibung der Teilmengen von Matrizen mit fixem Rang in Termen von Polynomen. Betrachten wir etwa $M_{m,n}(\mathbb{R})$ mit $m \geq n$. Für eine Matrix A dieser Größe ist natürlich $\operatorname{rg}(A) \leq n$ und nach Proposition 6.11 gilt $\operatorname{rg}(A) = n$ genau dann, wenn es einen $n \times n$–Minor B von A gibt, der $\det(B) \neq 0$ erfüllt. Betrachtet man die Determinanten aller $n \times n$–Minoren als Komponenten einer Funktion mit Werten in $\mathbb{R}^k$ (für passendes k), dann ist diese Funktion wieder stetig. Damit bilden die Matrizen vom Rang n eine offene Teilmenge und die Matrizen von kleinerem Rang eine abgeschlossene Teilmenge von $\mathbb{R}^{mn}$.

KAPITEL 7

Charakteristisches Polynom, Eigenwerte und Diagonalisierbarkeit

Nachdem wir nun die Determinante als Hilfsmittel zur Verfügung haben, beginnen wir lineare Abbildungen von einem Vektorraum auf sich selbst genauer zu studieren. Insbesondere versuchen wir, zu einer gegebenen linearen Abbildung eine Basis zu finden, bezüglich der wir eine besonders einfach Matrixdarstellung erhalten. Ein wesentlicher Schritt dahin wird sein, jeder linearen Abbildung (bzw. jeder quadratischen Matrix) das sogenannte charakteristische Polynom zuzuordnen. Damit können dann algebraische Resultate über Polynome auf das Studium von linearen Abbildungen bzw. von Matrizen angewandt werden, was im weiteren eine zentrale Rolle spielen wird.

Eigenwerte und Eigenräume – Diagonalisierbarkeit

7.1. Grundlegende Definitionen. Sie $\mathbb{K}$ ein Körper, V ein endlichdimensionaler $\mathbb{K}$–Vektorraum und $f : V \to V$ eine lineare Abbildung. Wir wollen versuchen, eine Basis für V zu finden, in der f besonders einfach aussieht. Äquivalent kann man das natürlich so formulieren, dass wir zu einer gegebenen Matrix $A \in M_n(\mathbb{K})$ eine möglichst einfache ähnliche Matrix suchen. Wir werden im weiteren öfters zwischen den Sichtweisen der Matrizen und der linearen Abbildungen hin und her schalten.

Als Anfang sollten wir vielleicht überlegen, was man eigentlich mit "besonders einfach aussehen" meinen könnte, bzw. Beispiele für besonders einfach lineare Abbildungen und Matrizen geben. Eine Klasse von besonders einfachen Matrizen sind die sogenannten *Diagonalmatrizen*, also Matrizen $A = (a_{ij})$, für die $a_{ij} = 0$ für $i \neq j$ gilt. Eine Diagonalmatrix hat also nur auf der Hauptdiagonale nichttriviale Eintragungen. Geometrisch bedeutet das, dass die Elemente der Standardbasis durch die Matrix A nur in ihrer Länge, aber nicht in ihrer Richtung geändert werden. Man kann nun natürlich für eine allgemeine lineare Abbildung Vektoren suchen, bei denen nur die Länge und nicht die Richtung verändert wird. Das führt sofort zu den zentralen Begriffen dieses Kapitels:

Definition 7.1. Sei $\mathbb{K}$ ein Körper, V ein $\mathbb{K}$–Vektorraum und $f : V \to V$ eine lineare Abbildung.

(1) Ein Element $\lambda \in \mathbb{K}$ heißt ein *Eigenwert* von f, falls es einen Vektor $v \in V$ mit $v \neq 0$ gibt, sodass $f(v) = \lambda v$ gilt. Ist das der Fall, so heißt v ein *Eigenvektor für f zum Eigenwert λ*.

(2) Ist λ ein Eigenwert von f, dann heißt $V_\lambda^f := \{v \in V : f(v) = \lambda v\}$ der *Eigenraum von f zum Eigenwert λ*.

(3) Die lineare Abbildung f heißt *diagonalisierbar* falls es eine Basis $\{v_i\}$ für V gibt, sodass jedes v_i ein Eigenvektor für f ist.

(4) Analog definieren wir die Begriffe Eigenwert, Eigenvektor, Eigenraum und Diagonalisierbarkeit für eine Matrix $A \in M_n(\mathbb{K})$, indem wir die lineare Abbildung $\mathbb{K}^n \to \mathbb{K}^n$ benutzen, die durch $x \mapsto Ax$ gegeben ist.

Bemerkung 7.1. Sei V ein endlichdimensionaler $\mathbb{K}$–Vektorraum, $f : V \to V$ eine diagonalisierbare lineare Abbildung und $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis für V, die aus Eigenvektoren für f besteht. Sind $\lambda_1, \dots, \lambda_n$ die zugehörigen Eigenwerte, dann gilt nach Definition $f(v_i) = \lambda_i v_i$. Nach Definition bedeutet das aber für die Matrixdarstellung $[f]_{\mathcal{B}} = (a_{ij})$ von f bezüglich der Basis $\mathcal{B}$ gerade $a_{ii} = \lambda_i$ und $a_{ij} = 0$ für $i \neq j$, also ist $[f]_{\mathcal{B}}$ eine Diagonalmatrix.

Ist umgekehrt $f : V \to V$ linear, und $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis von V, sodass $[f]_{\mathcal{B}}$ eine Diagonalmatrix ist, dann ist nach Definition $f(v_i) = a_{ii}v_i$, also v_i ein Eigenvektor zum Eigenwert a_{ii}. Damit ist f diagonalisierbar.

Somit sehen wir, dass eine lineare Abbildung genau dann diagonalisierbar ist, wenn sie bezüglich einer geeigneten Basis durch eine Diagonalmatrix dargestellt werden kann. Analog ist eine $n \times n$–Matrix A genau dann diagonalisierbar, wenn sie ähnlich zu einer Diagonalmatrix ist, d.h. wenn es eine invertierbare Matrix $T \in M_n(\mathbb{K})$ gibt, sodass TAT^{-1} eine Diagonalmatrix ist.

Beispiel 7.1. Zur Illustration des Begriffes betrachten wir zunächst drei Beispiele im $\mathbb{R}$–Vektorraum $\mathbb{R}^2$:

(1) Betrachte die Drehung um den Winkel $\varphi \in [0, 2\pi)$, d.h.

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Ist $\varphi \neq 0, \pi$, dann hat f offensichtlich keine Eigenwerte und kann daher auch nicht diagonalisierbar sein. Für $\varphi = 0$ ist $f = \mathrm{id}$, also jedes $(x, y) \neq (0, 0)$ ein Eigenvektor zum Eigenwert 1, und für $\varphi = \pi$ ist $f = -\mathrm{id}$, also jeder Vektor (außer Null) ein Eigenvektor zum Eigenwert -1.

(2) Betrachte die lineare Abbildung $f(x, y) = (2x, x+y)$. Offensichtlich ist $f(0, y) = (0, y)$, also jeder Vektor dieser Form ein Eigenvektor zum Eigenwert 1. Andererseits ist $f(x, x) = (2x, 2x)$ also ist jeder Vektor mit gleichen Komponenten ein Eigenvektor zum Eigenwert 2. Damit ist f diagonalisierbar, weil etwa $\mathcal{B} = \{(1, 1), (0, 1)\}$ eine Basis aus Eigenvektoren ist. Offensichtlich gilt dann $[f]_{\mathcal{B}} = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

(3) Betrachte die lineare Abbildung $f(x, y) = (x, x + y)$. Offensichtlich ist wieder jeder Vektor der Form $(0, y)$ ein Eigenvektor zum Eigenwert Eins. Es kann aber keine weiteren Eigenvektoren geben. Wäre nämlich $f(x, y) = (x, x+y) = (\lambda x, \lambda y)$ und $x \neq 0$, dann müsste natürlich $\lambda = 1$ gelten. Die zweite Komponenten liefert aber dann $y = x+y$, also $x = 0$, ein Widerspruch. Insbesondere kann f nicht diagonalisierbar sein, weil je zwei Eigenvektoren für f linear abhängig sind.

(4) Das Konzept von Eigenvektoren ist auch für unendlichdimensionale Vektorräume sehr wichtig. Betrachte den Raum $C^\infty(\mathbb{R}, \mathbb{R})$ der beliebig oft differenzierbaren Funktionen von $\mathbb{R}$ nach $\mathbb{R}$. In der Analysis lernt man, dass $(f + g)' = f' + g'$ und $(\lambda f)' = \lambda f'$ für alle $f, g \in C^\infty(\mathbb{R}, \mathbb{R})$ und $\lambda \in \mathbb{R}$ gilt, also definiert die Ableitung $D(f) = f'$ eine lineare Abbildung $D : C^\infty(\mathbb{R}, \mathbb{R}) \to C^\infty(\mathbb{R}, \mathbb{R})$. Wie ebenfalls aus der Analysis bekannt, ist für jedes $\lambda \in \mathbb{R}$ die Funktion $x \mapsto e^{\lambda x}$ beliebig oft differenzierbar, und $(e^{\lambda x})' = \lambda e^{\lambda x}$. Somit ist für jedes $\lambda \in \mathbb{R}$ die Funktion $e^{\lambda x}$ ein Eigenvektor von D zum Eigenwert λ.

7.2. Geometrische Vielfachheit. Sei V ein endlichdimensionaler Vektorraum über $\mathbb{K}$, und $f : V \to V$ eine lineare Abbildung. Ist λ ein Eigenwert von f, dann betrachten wir den zugehörigen Eigenraum $V_\lambda := V_\lambda^f$. Ist $v \in V_\lambda$ und $\mu \in \mathbb{K}$, dann ist $f(\mu v) = \mu f(v) = \mu\lambda v = \lambda\mu v$, also $\mu v \in V_\lambda$. Sind andererseits $v, w \in V_\lambda$, dann ist $f(v + w) = f(v) + f(w) = \lambda v + \lambda w = \lambda(v + w)$, also ist V_λ ein Teilraum vom V.

Definition 7.2. Die *geometrische Vielfachheit* eines Eigenwerts λ einer linearen Abbildung $f : V \to V$ ist die Dimension des zugehörigen Eigenraumes V_λ^f.

Damit können wir nun relativ einfach zu einer (allerdings in der Praxis noch nicht sehr nützlichen) Charakterisierung diagonalisierbarer Abbildungen kommen. Zunächst zeigen wir, dass Eigenvektoren zu verschiedenen Eigenwerten automatisch linear unabhängig sind.

Lemma 7.2. *Sei V ein Vektorraum über $\mathbb{K}$, $f : V \to V$ eine lineare Abbildung und seien $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ verschiedene Eigenwerte von f. Ist $v_i \in V$ ein Eigenvektor zum Eigenwert λ_i für jedes $i = 1, \dots, k$, dann sind die Vektoren $v_1, \dots, v_k$ linear unabhängig.*

Beweis. Durch Induktion nach k. Ist $k = 1$, dann ist v_1 ein Eigenvektor, also $v_1 \neq 0$, also ist $\{v_1\}$ linear unabhängig. Nehmen wir also an, dass $k \geq 2$ gilt, und dass wir die Behauptung für $k - 1$ Eigenvektoren zu verschiedenen Eigenwerten bereits bewiesen haben. Seien $a_1, \dots, a_k \in \mathbb{K}$ so, dass $a_1v_1 + \cdots + a_kv_k = 0$ gilt. Wenden wir auf diese Gleichung f an, dann erhalten wir wegen der Linearität $a_1f(v_1)+\cdots+a_kf(v_k) = 0$, und nach Definition der v_i erhalten wir $a_1\lambda_1v_1+\cdots+a_k\lambda_kv_k = 0$. Subtrahieren wir von dieser Gleichung das λ_k–fache der ursprünglichen Gleichung $a_1v_1+\cdots+a_kv_k = 0$, dann erhalten wir $a_1(\lambda_1-\lambda_k)v_1+\cdots+a_{k-1}(\lambda_{k-1}-\lambda_k)v_{k-1}+0 = 0$. Nach Induktionsvoraussetzung sind die Vektoren $v_1, \dots, v_{k-1}$ linear unabhängig, also ist $a_i(\lambda_i - \lambda_k) = 0$ für $i = 1, \dots, k-1$. Da die λ_i alle verschieden sind, ist $\lambda_i - \lambda_k \neq 0$, also $a_i = 0$ für alle $i = 1, \dots, k - 1$. Damit ist $a_kv_k = 0$, also $a_k = 0$ weil v_k ein Eigenvektor und damit nicht der Nullvektor ist, und die Behauptung folgt. □

Satz 7.2. *Sei V ein n–dimensionaler Vektorraum über einem Körper $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. Dann gilt:*

(1) f hat höchstens n verschiedene Eigenwerte.

(2) Sind $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von f und ist n_i die geometrische Vielfachheit von λ_i für $i = 1, \dots, k$, dann ist $\sum_{j=1}^k n_j \leq n$ und es gilt genau dann Gleichheit, wenn f diagonalisierbar ist.

Beweis. (1) Zu jedem Eigenwert gibt es nach Definition mindestens einen Eigenvektor, und nach Lemma 7.2 sind Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig. Da eine linear unabhängige Teilmenge von V nach Korollar 4.5 höchstens $\dim(V) = n$ Elemente haben kann, folgt die Behauptung.

(2) Für $i = 1, \dots, k$ sei $\{v_1^{(i)}, \dots, v_{n_i}^{(i)}\}$ eine Basis für den Eigenraum $V_{\lambda_i}^f$. Dann behaupten wir, dass die Vektoren $v_1^{(1)}, \dots, v_{n_1}^{(1)}, \dots, v_1^{(k)}, \dots, v_{n_k}^{(k)}$ linear unabhängig sind. Seien also $a_i^j \in \mathbb{K}$ für $j = 1, \dots, k$ und $i = 1, \dots, n_k$ so, dass $\sum_{i,j} a_i^jv_i^{(j)} = 0$ gilt. Dann können wir diese endliche Summe als $\sum_{j=1}^k v^{(j)} = 0$ schreiben, wobei $v^{(j)} := \sum_{i=1}^{n_j} a_i^jv_i^{(j)}$. Nach Konstruktion ist $v^{(j)} \in V_{\lambda_j}^f$. Nach Lemma 7.2 ist aber $\sum v^{(j)} = 0$ nur möglich, wenn $v^{(j)} = 0$ für alle j gilt.

Damit ist aber $\sum_{i=1}^{n_j} a_i^jv_i^{(j)} = 0$ für alle j, was $a_i^j = 0$ für alle i und j impliziert, weil nach Konstruktion $v_1^{(j)}, \dots, v_{n_j}^{(j)}$ für alle j linear unabhängig sind. Damit ist die $\sum n_j$–elementige Teilmenge $\{v_1^{(1)}, \dots, v_{n_k}^{(k)}\}$ von V linear unabhängig, also folgt $\sum n_j \leq n$ wieder aus Korollar 4.5.

Falls $\sum n_j = n$ gilt, dann ist diese Menge nach Teil (6) von Korollar 4.5 eine Basis von V, also ist f diagonalisierbar. Ist umgekehrt f diagonalisierbar, dann finden wir eine Basis $\{v_1, \dots, v_n\}$ aus Eigenvektoren. Elemente dieser Basis, die Eigenvektoren zu einem fixen Eigenwert λ sind, bilden eine linear unabhängige Teilmenge von V_λ^f, also ist ihre Anzahl höchstens gleich der geometrischen Vielfachheit von λ. Damit folgt aber sofort $\sum n_j \geq n$. □

Insbesondere impliziert der Satz dass eine lineare Abbildung $f : V \to V$, die $\dim(V)$ viele verschieden Eigenwerte hat, automatisch diagonalisierbar ist.

Das charakteristische Polynom

Wir wissen an dieser Stelle noch nicht, wie wir zeigen können, dass eine gegebene lineare Abbildung Eigenwerte hat, bzw. wie man Eigenwerte bestimmen könnte. Aus Beispiel (1) von 7.1 wissen wir sogar, dass es im Allgemeinen keine Eigenwerte geben muss. Um diese Fragen zu studieren kommt uns die Theorie der Polynome zur Hilfe.

7.3. Eigenwerte und Determinanten. Zunächst können wir die Eigenräume einer linearen Abbildung leicht als Kern interpretieren. Damit kommen wir der Frage der Bestimmung von Eigenwerten einen großen Schritt näher:

Proposition 7.3. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. Dann ist eine Zahl $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von f, wenn* $\det(f - \lambda\,\mathrm{id}) = 0$ *gilt. Ist λ ein Eigenwert, dann ist der Eigenraum V_λ^f genau der Kern von* $f - \lambda\,\mathrm{id}$.

Beweis. Aus 2.5 wissen wir, dass die Vektorraumoperationen auf $L(V, V)$ punktweise definiert sind, also ist $(f - \lambda\,\mathrm{id})(v) = f(v) - \lambda v$ für alle $v \in V$. Daher gilt $f(v) = \lambda v$ genau dann, wenn $(f - \lambda\,\mathrm{id})(v) = 0$ gilt. Somit ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert, wenn $\mathrm{Ker}(f - \lambda\,\mathrm{id}) \neq \{0\}$ ist. Nach dem Dimensionssatz für lineare Abbildungen (Satz 4.11) ist das äquivalent dazu, dass $f - \lambda\,\mathrm{id}$ kein linearer Isomorphismus ist. Aus 6.8 wissen wir, dass $f - \lambda\,\mathrm{id}$ genau dann kein linearer Isomorphismus ist, wenn $\det(f - \lambda\,\mathrm{id}) = 0$ ist. Für einen Eigenwert λ gilt dann offensichtlich $V_\lambda^f = \mathrm{Ker}(f - \lambda\,\mathrm{id})$. □

Betrachten wir diese Charakterisierung von Eigenwerten im Bild von Matrizen, dann ist also für eine Matrix $A = (a_{ij}) \in M_n(\mathbb{K})$ eine Zahl $\lambda \in \mathbb{K}$ genau dann ein Eigenwert, wenn die Matrix $B = A - \lambda\mathbb{I}$, d.h. $b_{ii} = a_{ii} - \lambda$ und $b_{ij} = a_{ij}$ für $i \neq j$ Determinante Null hat.

An dieser Stelle können wir, wie in Abschnitt 6.2 skizziert, Polynome ins Spiel bringen. Dort haben wir festgestellt, dass die Menge $\mathbb{K}[x]$ aller Polynome in einer Variablen x mit Koeffizienten im Körper $\mathbb{K}$ einen kommutativen Ring mit Einselement bilden. Ausserdem haben wir gesehen, dass für ein Element $\lambda \in \mathbb{K}$ die Abbildung $\varphi = \varphi_\lambda : \mathbb{K}[x] \to \mathbb{K}$, die jedem Polynom p den Wert der zugehörigen Polynomfunktion auf $\mathbb{K}$ im Punkt λ zuordnet, eine Homomorphismus von kommutativen Ringen mit Einselement ist. Wir werden diese Abbildung weiterhin einfach als $\varphi_\lambda(p) = p(\lambda)$ schreiben.

Für die gegebene Matrix $A = (a_{ij}) \in M_n(\mathbb{K})$ betrachten wir nun die Matrix $C \in M_n(\mathbb{K}[x])$ die definiert ist durch $c_{ii} = a_{ii} - x$ und $c_{ij} = a_{ij}$ für $i \neq j$. Wir haben also auf der Hauptdiagonalen Polynome ersten Grades und außerhalb der Hauptdiagonale nur konstante Polynome in C. Wir werden die Matrix C auch als $A - x\mathbb{I}$ bezeichnen. Aus Kapitel 6 wissen wir nun, dass wir die Determinate $\det(C)$ von C als Element von $\mathbb{K}[x]$ bilden können.

Definition 7.3. Für $A \in M_n(\mathbb{K})$ ist das *charakteristische Polynom* $p_A \in \mathbb{K}[x]$ definiert als die Determinante der Matrix $A - x\mathbb{I} \in M_n(\mathbb{K}[x])$.

Das charakteristische Polynom ist eines der zentralen Konzepte der linearen Algebra. Insbesondere sind seine Nullstellen (wie in 4.12 definiert)genau die Eigenwerte der Matrix:

Korollar 7.3. *Sei $A \in M_n(\mathbb{K})$ eine $n \times n$–Matrix. Dann sind die Eigenwerte der linearen Abbildung $x \mapsto Ax$ genau die Nullstellen des charakteristischen Polynoms p_A, d.h. $\lambda \in \mathbb{K}$ ist genau dann ein Eigenwert von A, wenn $p_A(\lambda) = 0$ gilt.*

Beweis. In 6.2 haben wir dem Homomorphismus $\varphi_\lambda : \mathbb{K}[x] \to \mathbb{K}$ auf die einzelnen Eintragungen von Matrizen angewandt um eine Funktion $\widehat{\varphi_\lambda} : M_n(\mathbb{K}[x]) \to M_n(\mathbb{K})$ zu definieren. Offensichtlich bildet diese Funktion $A - x\mathbb{I} \in M_n(\mathbb{K}[x])$ auf $A - \lambda\mathbb{I} \in M_n(\mathbb{K})$ ab. In Korollar 6.4 haben wir bewiesen, dass für jede Matrix $B \in M_n(\mathbb{K}[x])$ die Gleichung $\det(\widehat{\varphi_\lambda}(B)) = \varphi(\det(B))$ gilt. Angewandt auf $A - x\mathbb{I}$ erhalten wir $\det(A - \lambda\mathbb{I}) = p_A(\lambda)$ und die Behauptung folgt aus Proposition 7.3. □

Beispiel 7.3. Betrachten wir die 2×2–Matrizen zu den Beispielen aus 7.1. Für die Drehung um den Winkel φ erhalten wir die Matrixdarstellung $A = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$. Daraus ergibt sich $p_A = (\cos(\varphi) - x)^2 + \sin^2(\varphi) = x^2 - 2\cos(\varphi)x + 1$. Nach der üblichen Formel für die Lösung von quadratischen Gleichungen ergibt sich als Nullstellen $\lambda_{1,2} = \cos(\varphi) \pm \sqrt{\cos^2(\varphi) - 1} = \cos(\varphi) \pm i\sin(\varphi)$. Somit existieren Nullstellen in $\mathbb{R}$ nur für $\sin(\varphi) = 0$, also $\varphi = k\pi$ mit $k \in \mathbb{Z}$, und das liefert nur die Fälle $A = \pm\mathbb{I}$. Wir werden später sehen, wie man die komplexen Nullstellen benutzen kann.

Für die zweite lineare Abbildung erhalten wir $A = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$ und damit das charakteristische Polynom $p_A = (2-x)(1-x)$ woraus schon offensichtlich ist, dass die Nullstellen gerade 2 und 1 sind.

Im Fall der Scherung erhalten wir schließlich $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ und damit das charakteristische Polynom $p_A = (1 - x)^2$. Hier ist also 1 eine "doppelte" Nullstelle des charakteristischen Polynoms, obwohl es nur einen Eigenvektor zum Eigenwert 1 gibt.

7.4. Charakteristisches Polynom und Ähnlichkeit. Bevor wir uns allgemein mit Polynomen beschäftigen, zeigen wir, dass man das charakteristische Polynom für lineare Abbildungen definieren kann. Wie in 6.1 besprochen muss man dazu die charakteristischen Polynome von ähnlichen Matrizen betrachten.

Satz 7.4. *Seien $A, B \in M_n(\mathbb{K})$ ähnliche $n \times n$–Matrizen über einem Körper $\mathbb{K}$. Dann haben A und B das gleiche charakteristische Polynom.*

Beweis. Nach Definition gibt es eine invertierbare Matrix $T \in M_n(\mathbb{K})$, sodass $B = TAT^{-1}$ gilt. Betracten wir die Eintragungen t_{ij} von T als konstante Polynome, dann können wir T auch als Element von $M_n(\mathbb{K}[x])$ betrachten. Das geht natürlich analog auch für T^{-1} und klarerweise sind T und T^{-1} auch als Elemente von $M_n(\mathbb{K}[x])$ invers zueinander.

Nun gilt $T(A - x\mathbb{I})T^{-1} = TAT^{-1} - Tx\mathbb{I}T^{-1}$. Nach Defintion sind die Koeffizienten von $Tx\mathbb{I}$ gerade xt_{ij} und daraus folgt sofort, dass $Tx\mathbb{I}T^{-1} = x\mathbb{I}$ gilt. Damit ist aber $B - x\mathbb{I} = T(A - x\mathbb{I})T^{-1}$, also ist $p_B = \det(T)p_A\det(T^{-1})$ und nach Korollar 6.8 is $\det(T)\det(T^{-1}) = 1$. □

Mit Hilfe dieses Satzes können wir nun für eine lineare Abbildungen $f : V \to V$ auf einem endlichdimensionalen $\mathbb{K}$–Vektorraum V das charakteristische Polynom definieren.

Dazu definieren wir einfach $p_f \in \mathbb{K}[x]$ als das charakteristische Polynom der Matrixdarstellung $[f]_{\mathcal{B}}$ bezüglich einer beliebigen Basis $\mathcal{B}$ von V. Wählen wir eine andere Basis, dann erhalten wir nach 6.1 als Matrixdarstellung eine Matrix, die ähnlich zu $[f]_{\mathcal{B}}$ ist, und damit nach dem obigen Satz das gleiche charakteristische Polynom. Natürlich sind die Nullstellen von p_f genau die Eigenwerte von f.

7.5. Die Spur. Nachdem wir nun Matrizen und linearen Abbildungen ein charakteristisches Polynom zugeordnet haben, kann man die Koeffizienten dieses Polynoms betrachten, die sogenannten *charakteristischen Koeffizienten*. Nach Satz 7.4 müssen diese Koeffizienten für ähnliche Matrizen immer gleich sein und für lineare Abbildungen kann man sie aus einer beliebigen Matrixdarstellung bestimmen. Wir werden später sehen, dass die Determinante einer dieser Koeffizienten ist. Hier betrachten wir einen anderen Koeffizienten, der noch wesentlich einfacher ist als die Determinante und damit eine einfache Möglichkeit liefert, um Ähnlichkeit von Matrizen auszuschließen.

Sei $A \in M_n(\mathbb{K})$ eine $n \times n$–Matrix mit charakteristischem Polynom $p_A \in \mathbb{K}[x]$. Dann ist es ziemlich einfach, den Koeffizienten von x^{n-1} in p_A berechnen. Setzen wir nämlich $B = A - x\mathbb{I} = (b_{ij})$ und betrachten die Leibniz Formel

$$p_A = \det(B) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) b_{\sigma(1)1} \cdots b_{\sigma(n)n}$$

als Produkt von Polynomen. Nach Konstruktion ist für $i \neq j$ das Polynom b_{ij} konstant, während $b_{ii} = a_{ii} - x$ für $i = 1, \dots, n$ gilt. Ist aber $\sigma \neq \mathrm{id}$ eine Permutation, dann gilt $\sigma(i) \neq i$ für mindestens zwei Indizes i, also kann der entsprechende Summand als höchste Potenz maximal x^{n-2} enthalten und trägt daher nicht zum Koeffizienten von x^{n-1} bei. Damit kommen aber alle Beiträge zum Koeffizienten von x^{n-1} aus dem Summanden

$$(a_{11} - x) \cdots (a_{nn} - x),$$

der $\sigma = \mathrm{id}$ entspricht. Der Koeffizient von x^{n-1} in diesem Produkt ist aber offensichtlich $(-1)^{n-1}(a_{11} + \cdots + a_{nn})$. Das motiviert die folgende Definition.

Definition 7.5. Sei $A = (a_{ij}) \in M_n(\mathbb{K})$. Dann definieren wir die *Spur* (englisch “trace”) $\operatorname{tr}(A)$ von A durch $\operatorname{tr}(A) := a_{11} + \cdots + a_{nn}$, also die Summe der Elemente auf der Hauptdiagonale.

Proposition 7.5. *(1) Die Spurabbildung* $\operatorname{tr} : M_n(\mathbb{K}) \to \mathbb{K}$ *ist linear.*

(2) Ähnliche Matrizen haben die gleiche Spur. Allgemeiner gilt für $A, B \in M_n(\mathbb{K})$ *die Gleichung* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Beweis. (1) ist aus der Definition offensichtlich und der erste Teil von (2) ist klar, weil die Spur ein Koeffizient des charakteristischen Polynoms ist.

Also müssen wir nur die letzte Behauptung beweisen. Ist $A = (a_{ij})$ und $B = (b_{ij})$, dann sind die Hauptdiagonalelemente von AB gegeben durch $\sum_j a_{ij}b_{ji}$. Damit ist aber $\operatorname{tr}(AB) = \sum_i \sum_j a_{ij}b_{ji}$, und dieser Ausdruck ist offensichtlich unabhängig von der Reihenfolge von A und B. □

Bemerkung 7.5. (1) Die Aussage, dass $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ für alle $A, B \in M_n(\mathbb{K})$ gilt impliziert die Aussage, dass ähnliche Matrizen gleiche Spur haben. Man rechnet einfach $\operatorname{tr}(TAT^{-1}) = \operatorname{tr}(T^{-1}TA) = \operatorname{tr}(A)$. Der Zusammenhang mit dem charakteristischen Polynom ist also für den Beweis von Proposition 7.5 nicht notwendig, er ist aber der tiefere Grund für die schönen Eigenschaften der Spurabbildung.

(2) Die Proposition impliziert natürlich, dass man die Spur einer linearen Abbildung als die Spur einer beliebigen Matrixdarstellung definieren kann.

Polynome und ihre Nullstellen

Um das charakteristische Polynom nutzen zu können, entwickeln wir als nächsten Schritt etwas allgemeine Theorie über Polynome.

7.6. Der Euklidische Algorithmus. Wie wir uns in 7.3 schon erinnert haben ist der Raum $\mathbb{K}[x]$ aller Polynome mit Koeffizienten in $\mathbb{K}$ nicht nur ein Vektorraum über $\mathbb{K}$ sondern auch ein kommutativer Ring mit Einselement. Das andere offensichtliche Beispiel eines kommutativen Ringes mit Einselement bilden die ganzen Zahlen $\mathbb{Z}$. Viele der Anwendungen von Polynomen, die wir im weiteren kennen lernen werden, beruhen darauf, dass sich $\mathbb{K}[x]$ und $\mathbb{Z}$ überraschend ähnlich studieren lassen. Insbesondere gibt es in $\mathbb{K}[x]$ Analoga von Primzahlen und der eindeutigen Primfaktorzerlegung, die später noch eine große Rolle spielen werden. Die Basis für die enge Analogie zwischen $\mathbb{K}[x]$ und $\mathbb{Z}$ ist, dass es in beiden Fällen möglich ist, mit Rest zu dividieren (was für Polynome auch schon aus der Schule bekannt ist). Dazu benötigt man (um auszudrücken, dass der Rest "klein" ist) einen Ersatz für die Ordnung auf $\mathbb{Z}$, den der Grad von Polynomen darstellt.

Definition 7.6. Sei $p = \sum a_k x^k \in \mathbb{K}[x]$ ein Polynom mit $p \neq 0$. Dann gibt es nach Definition einen maximalen Index $n \in \mathbb{N}$, sodass $a_n \neq 0$ gilt. Dieses n heißt der *Grad* $\deg(p)$ von p und der Koeffizient a_n heißt der *führende Koeffizient* von p. Äquivalent ist $\deg(p) = n$ genau dann, wenn $p \in \mathbb{K}_n[x]$ aber $p \notin \mathbb{K}_{n-1}[x]$ gilt. Man nennt ein Polynom $p \in \mathbb{K}[x]$ *monisch*, wenn sein führender Koeffizient 1 ist.

Die Verträglichkeit des Grades von Polynomen mit den algebraischen Operationen ist leicht zu analysieren: Für Polynome $p = \sum a_k x^k$ und $q = \sum b_\ell x^\ell$ ist das Polynome $p + q = \sum c_i x^i$ bestimmt durch $c_i = a_i + b_i$. Damit folgt sofort, dass $\deg(p + q) \leq \max\{\deg(p), \deg(q)\}$, wobei Gleichheit gelten muss, wenn die beiden Grade verschieden sind. Analog ist für $r \in \mathbb{K}$ mit $r \neq 0$ natürlich $\deg(rp) = \deg(p)$. Für das Produkt $pq = \sum d_i x^i$ gilt ja $d_k = \sum_{i=0}^{k} a_i b_{k-i}$. Ist $\deg(p) = n$ und $\deg(q) = m$, dann ist für $k > n + m$ in dieser Summe natürlich entweder $i > n$ also $a_i = 0$ oder $k - i > m$ also $b_{k-i} = 0$ und es folgt $d_k = 0$. In der Summe, die d_{n+m} definiert sind offensichtlich alle Summanden gleich Null, außer $a_n b_m \neq 0$. Damit gilt für $p, q \neq 0$ offensichtlich $\deg(pq) = \deg(p) + \deg(q)$ und der führende Koeffizient eines Produkts ist das Produkt der führenden Koeffizienten der beiden Faktoren. Damit können wir nun das Resultat über Division von Polynomen mit Rest ("Euklidischer Algorithmus") formulieren, das eigentlich schon aus der Schule bekannt ist:

Lemma 7.6. *Seien $p_1, p_2 \in \mathbb{K}[x]$ Polynome, $p_2 \neq 0$. Dann gibt es eindeutig bestimmte Polynome $q, r \in \mathbb{K}[x]$, sodass $p_1 = qp_2 + r$, sowie $r = 0$ oder $\deg(r) < \deg(p_2)$ gilt.*

Beweis. Wir beweisen zunächst die Existenz der Zerlegung. Falls $p_1 = 0$ ist oder $\deg(p_1) < \deg(p_2)$ gilt, dann setzen wir $q = 0$ und $r = p_1$ und beide Eigenschaften sind erfüllt.

Sei also $n := \deg(p_1) \geq \deg(p_2) =: m$. Wir führen wir den Beweis durch Induktion nach n. Seien a_n und b_m die führenden Koeffizienten der beiden Polynome. Dann betrachten wir $\tilde{p}_1 = p_1 - \frac{a_n}{b_m} x^{n-m} p_2$. Nach Konstruktion ist $\deg(\tilde{p}_1) \leq n$. Die Terme in $\tilde{p}_1$, die x^n enthalten sind aber gerade $a_n x^n - \frac{a_n}{b_m} x^{n-m} b_m x^m = 0$, also ist $\deg(\tilde{p}_1) < n$. Nach Induktionsvoraussetzung finden wir Polynome $\tilde{q}$ und $\tilde{r}$, sodass $\tilde{p}_1 = \tilde{q} p_2 + \tilde{r}$ und $\tilde{r} = 0$ oder $\deg(\tilde{r}) < \deg(p_2)$ gelten. Damit ist aber $p_1 - \frac{a_n}{b_m} x^{n-m} p_2 = \tilde{q} p_2 + \tilde{r}$, und man setzt $q = \tilde{q} + \frac{a_n}{b_m} x^{n-m}$ und $r = \tilde{r}$. Dann erhält man $p_1 = qp_2 + r$ mit den gewünschten Eigenschaften.

Zur Eindeutigkeit: Ist $qp_2+r = \tilde{q}p_2+\tilde{r}$, wobei r und $\tilde{r}$ Null sind oder kleineren Grad als p_2 haben, dann erhalten wir $r - \tilde{r} = (\tilde{q} - q)p_2$. Die linke Seite dieser Gleichung ist entweder Null oder hat Grad $< \deg(p_2)$. Wäre $\tilde{q} - q \neq 0$, dann hätte die rechte Seite Grad $\deg(p_2) + \deg(\tilde{q} - q) \geq \deg(p_2)$ ein Widerspruch. Somit muss $\tilde{q} - q = 0$, also $q = \tilde{q}$ gelten. Daraus folgt dann sofort $\tilde{r} = r$. □

Beispiel 7.6. Der Induktionsbeweis des Satzes entspricht genau dem aus der Schule bekannten Algorithmus zur Division von Polynomen mit Rest. Betrachten wir zum Beispiel die Division $(x^3 + 2x^2 - x + 1) : (x + 3)$ Im ersten Schritt erhält man x^2 für den Quotienten, und zieht man $x^2(x + 3)$ vom Dividenden ab, dann erhält man $-x^2 - x + 1$. Damit erhält man $-x$ für den Quotienten und Abziehen von $-x(x + 3)$ liefert $2x + 1$. Somit erhält man noch 2 für den Quotienten und $2x + 1 - 2(x + 3) = -5$ für den Rest. Also gilt $x^3 + 2x^2 - x + 1 = (x^2 - x + 2)(x + 3) - 5$.

Mit diesem Resultat können wir aber nun eine schöne Charakterisierung der Nullstellen von Polynomen finden. Für $\lambda \in \mathbb{K}$ gibt es natürlich ein einfaches Polynom, das λ als Nullstelle hat, nämlich $x - \lambda \in \mathbb{K}_1[x]$. Da die Multiplikation von Polynomen der punktweisen Multiplikation der zugehörigen Polynomfunktionen entspricht folgt sofort, dass für jedes Polynom $q \in \mathbb{K}[x]$ auch $(x - \lambda)q$ eine Nullstelle in λ hat. Das sind aber die einzigen Polynome, die λ als Nullstelle haben:

Satz 7.6. *Sei $p \in \mathbb{K}[x]$ ein Polynom, $p \neq 0$. Dann ist eine Zahl $\lambda \in \mathbb{K}$ genau dann eine Nullstelle von p, wenn das Polynom $x - \lambda$ das Polynom p teilt, d.h. wenn es ein Polynom $q \in \mathbb{K}[x]$ gibt, sodass $p = (x - \lambda)q$ gilt.*

Beweis. Wenden wir den Euklidischen Algorithmus auf p und $(x - \lambda)$ an, dann erhalten wir Polynome $q, r \in \mathbb{K}[x]$ mit $p = (x - \lambda)q + r$ und $r = 0$ oder $\deg(r) < \deg(x - \lambda)$. Damit muss aber r konstant sein, also $r \in \mathbb{K}$ gelten. Setzten wir nun in dieser Gleichung λ für x ein, dann erhalten wir $p(\lambda) = 0q(\lambda)+r = r$, und die Behauptung folgt. □

7.7. Vielfachheit von Nullstellen. Der letzte Satz liefert uns sofort einen alternativen Beweise für die Tatsache, dass ein Polynom $0 \neq p \in \mathbb{K}_n[x]$ höchsten n verschiedene Nullstellen haben kann. Seien nämlich $\lambda_1, \dots, \lambda_k$ verschiedene Nullstellen eines Polynoms $0 \neq p \in \mathbb{K}[x]$. Nach Satz 7.6 gibt es ein Polynom q_1, sodass $p = (x - \lambda_1)q_1$ gilt. Setzen wir nun λ_2 ein, dann erhalten wir $0 = p(\lambda_2) = (\lambda_2 - \lambda_1)q_1(\lambda_2)$. Da $\lambda_2 \neq \lambda_1$ gilt, ist $\lambda_2 - \lambda_1 \neq 0$, also muss $q_1(\lambda_2) = 0$ gelten. Nach Satz 7.6 gibt es somit ein Polynom $q_2 \in \mathbb{K}[x]$, sodass $q_1 = (x - \lambda_2)q_2$ und damit $p = (x - \lambda_1)(x - \lambda_2)q_2$ gilt. Induktiv sehen wir, dass es ein Polynom $q_k \in \mathbb{K}[x]$ geben muss, sodass $p = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)q_k$ gilt, was insbesondere $\deg(p) = k + \deg(q_k)$, also $\deg(p) \geq k$ impliziert.

Die algebraische Sichtweise erlaubt uns nun eine feinere Analyse von Nullstellen, indem wir den Begriff der *Vielfachheit* einer Nullstelle einführen. Ist $p \in \mathbb{K}[x]$, $p \neq 0$ und λ eine Nullstelle von p, dann sagt man λ ist eine *k–fache Nullstelle* von p, wenn es ein Polynom $q \in \mathbb{K}[x]$ mit $q(\lambda) \neq 0$ gibt, sodass $p = (x - \lambda)^k q$ gilt. Die Vielfachheit einer Nullstelle gibt also die höchste Potenz von $(x - \lambda)$ an, die das Polynom p noch teilt.

Proposition 7.7. *Sei $0 \neq p \in \mathbb{K}[x]$ ein Polynom, seien $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ verschiedene Nullstellen von p und sei m_j die Vielfachheit von λ_j für $j = 1, \dots, k$. Dann gibt es ein Polynom $q \in \mathbb{K}[x]$, sodass*

$$p = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k} q$$

gilt. Insbesondere ist $\sum_{j=1}^{k} m_j \leq \deg(p)$*, also kann ein Polynom vom Grad* n*, mit Vielfachheit gezählt, nur höchstens* n *Nullstellen haben.*

Beweis. Wir beweisen die Aussage durch Induktion nach k. Nach Definition gibt es ein Polynom $q_1 \in \mathbb{K}[x]$, sodass $p = (x - \lambda_1)^{m_1} q_1$ gilt. Nehmen wir induktiv an, dass $i \geq 2$ gilt, und wir ein Polynom q_{i-1} gefunden haben, sodass

$$p = (x - \lambda_1)^{m_1} \cdots (x - \lambda_{i-1})^{m_{i-1}} q_{i-1}$$

gilt. Nun ist aber $p(\lambda_i) = 0$, und wenn wir λ_i in die rechte Seite einsetzen, dann erhalten wir einen Faktor ungleich Null mal $q_{i-1}(\lambda_i)$. Also muss $q_{i-1}(\lambda_i) = 0$ und damit $q_{i-1} = (x - \lambda_i) q_{i,1}$ für ein Polynom $q_{i,1} \in \mathbb{K}[x]$ gelten. Ist $m_i = 1$, dann können wir einfach $q_i := q_{i,1}$ setzen und erhalten $q_{i-1} = (x - \lambda_i)^{m_i} q_i$.

Ist $m_i > 1$, dann wissen wir nach Definition der Vielfachheit, dass es ein Polynom $\tilde{q}$ gibt, sodass $p = (x - \lambda_i)^{m_i} \tilde{q}$ gilt. Setzen wir oben für q_{i-1} ein, dann erhalten wir

$$p = (x - \lambda_i)(x - \lambda_1)^{m_1} \cdots (x - \lambda_{i-1})^{m_{i-1}} q_{i,1} = (x - \lambda_i)(x - \lambda_i)^{m_i - 1} \tilde{q}.$$

Wegen der Eindeutigkeit in Lemma 7.6 folgt

$$(x - \lambda_1)^{m_1} \cdots (x - \lambda_{i-1})^{m_{i-1}} q_{i,1} = (x - \lambda_i)^{m_i - 1} \tilde{q}.$$

Die rechte Seite hat λ_i als Nullstelle, also erhalten wir wie oben $q_{i,1}(\lambda_i) = 0$, also $q_{i,1} = (x - \lambda_i) q_{i,2}$ und damit $q_{i-1} = (x - \lambda_i)^2 q_{i,2}$ für ein Polynom $q_{i,2} \in \mathbb{K}[x]$ folgern. Ist $m_i = 2$, dann setzen wir $q_i = q_{i,2}$ und erhalten $q_{i-1} = (x - \lambda_i)^{m_i} q_i$. Sonst können wir weiter Faktoren $(x - \lambda_i)$ abspalten, bis wir ein Polynom $q_i := q_{i,m_i}$ erreichen, für das $q_{i-1} = (x - \lambda_i)^{m_i} q_i$. Damit folgt die Darstellung von p mittels Induktion und die zweite Aussage ist eine unmittelbare Konsequenz. □

Die Charakterisierung von Diagonalisierbarkeit

Wir wollen nun unser allgemeinen Resultate über Polynome im Spezialfall von charakteristischen Polynomen anwenden. Wir werden die Resultate oft entweder nur für Matrizen oder nur für lineare Abbildungen formulieren, die nötige Umformulierung für den jeweils anderen Fall ist offensichtlich.

7.8. Zunächst können wir aus der Definition sofort den Grad des charakteristischen Polynoms, sowie seinen höchsten Koeffizienten und den konstanten Koeffizienten bestimmen.

Satz 7.8. *Sei* $A \in M_n(\mathbb{K})$ *eine* $n \times n$*–Matrix über einem Körper* $\mathbb{K}$ *mit charakteristischem Polynom* $p_A \in \mathbb{K}[x]$*. Dann ist* $\deg p_A = n$*, der führende Koeffizient von* p_A *ist* $(-1)^n$ *und der konstante Koeffizient ist* $\det(A)$.

Beweis. Für den ersten Teil betrachte $C = (c_{ij}) = (A - x\mathbb{I}) \in M_n(\mathbb{K}[x])$. Dann ist $\deg(c_{ij}) = 0$ für $i \neq j$ und $\deg(b_{ii}) = 1$ für alle i und der führende Koeffizient von b_{ii} ist -1. Nach Satz 6.7 ist

$$p_A = \det(B) = \textstyle\sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) b_{\sigma(1)1} \cdots b_{\sigma(n)n},$$

wobei $\mathfrak{S}_n$ die Menge aller Permutationen der Menge $\{1, \dots, n\}$ bezeichnet. Für $\sigma = \mathrm{id}$ erhalten wir ein Produkt von n Polynomen ersten Grades also ein Polynom vom Grad n, und nach Definition des Produktes von Polynomen ist der führende Koeffizient eines Produktes das Produkt der führenden Koeffizienten. Damit erhalten wir $(-1)^n$ als führenden Koeffizienten dieses Summanden. In allen anderen Summanden treten auch Faktoren auf, die Grad Null besitzen, also erhalten wir Polynome vom Grad $< n$, also hat die Summe Grad n und führenden Koeffizienten $(-1)^n$. Wegen $p_A(\lambda) = \det(A - \lambda\mathbb{I})$

erhalten wir natürlich sofort $p_A(0) = \det(A)$, und das ist der konstante Koeffizient von p_A. □

Bemerkung 7.8. Das charakteristische Polynom liefert überraschende weitere Informationen über die Struktur der Teilmenge $GL(n, \mathbb{R})$ aller invertierbaren $n \times n$–Matrizen des Vektorraumes $M_n(\mathbb{R}) = \mathbb{R}^{n^2}$. Aus 6.12 wissen wir ja schon, dass $GL(n, \mathbb{R})$ eine offene Teilmenge von $\mathbb{R}^{n^2}$ ist. Jetzt können wir beweisen, dass diese Menge auch *dicht* ist, also beliebig nahe bei einer gegeben Matrix A invertierbare Matrizen liegen.

Ist nämlich $A \in M_n(\mathbb{R})$ beliebig, dann betrachten wir das charakteristische Polynom p_A von A als Funktion $\mathbb{R} \to \mathbb{R}$. Falls $p_A(\lambda) \neq 0$ ist, dann ist $\det(A - \lambda\mathbb{I}) \neq 0$, also $A - \lambda\mathbb{I}$ eine invertierbare Matrix. Da aber p_A ein Polynom vom Grad $n \geq 1$ ist, hat es nach Proposition 7.7 höchstens n verschiedene Nullstellen. Insbesondere bedeutet das, dass es zu jedem $\varepsilon > 0$ eine Zahl λ mit $|\lambda| < \varepsilon$ gibt, sodass $p_A(\lambda) \neq 0$ gilt. Das bedeutet aber gerade, dass beliebig nahe bei A invertierbare Matrizen liegen. Insbesondere hat das die nützliche Konsequenz, dass eine stetige Funktion $F : M_n(\mathbb{R}) \to \mathbb{R}^m$, die $F(A) = 0$ für jede invertierbare Matrix A erfüllt, schon identisch Null sein muss.

An dieser Stelle stellt sich die Frage, ob die charakteristischen Polynome von Matrizen spezielle Eigenschaften haben (die etwa beim Finden von Nullstellen helfen könnten). Das ist aber (abgesehen vom führenden Koeffizienten, den wir in Satz 7.8 bestimmt haben) nicht der Fall:

Proposition 7.8. *Sei $\mathbb{K}$ ein Körper und sei $p \in \mathbb{K}_n[x]$ ein Polynom vom Grad n mit führendem Koeffizienten $(-1)^n$. Dann gibt es eine Matrix $A \in M_n(\mathbb{K})$ mit charakteristischem Polynom $p_A = p$.*

Beweis. Schreiben wir $p = (-1)^n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$ und betrachten die Matrix $A = (a_{ij})$, die gegeben ist durch $a_{i+1,i} = 1$ für $i = 1, \dots, n-1$, $a_{jn} = (-1)^{n-1} b_{j-1}$ für $j = 1, \dots, n$ und alle anderen $a_{ij} = 0$. Wir beweisen mit Induktion nach n, dass $p_A = (-1)^n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$ gilt.

Für den Induktionsanfang betrachten wir den Fall $n = 2$, also die Matrix $\left(\begin{smallmatrix} 0 & -b_0 \\ 1 & -b_1 \end{smallmatrix}\right)$. Das charakteristische Polynom dieser Matrix ist offensichtlich $(-x)(-b_1 - x) - (-b_0) = x^2 + b_1 x + b_0$.

Nehmen wir induktiv an, dass $n \geq 3$ gilt, und die Behauptung für $(n-1) \times (n-1)$–Matrizen schon bewiesen ist. Entwickeln wir nun die Determinante von $A - x\mathbb{I}$ nach der ersten Zeile, so erhalten wir nur zwei Summanden. Der erste dieser Summanden ist $(-x)$ mal die Determinante der Matrix die entsteht, wenn man in $A - x\mathbb{I}$ die erste Zeile und die erste Spalte streicht. Nach Induktionsvoraussetzung ist diese Determinante aber $(-1)^{n-1}x^{n-1} - b_{n-1}x^{n-2} - \cdots - b_2 x - b_1$, also liefert dieser Summand schon $p - b_0$. Der andere Summand ist aber gerade $(-1)^{n+1}(-1)^{n-1}b_0 = b_0$ mal der Determinante der Matrix die entsteht, wenn man in $A - x\mathbb{I}$ die erste Zeile und die letzte Spalte streicht. Die resultierende Matrix ist aber eine obere Dreiecksmatrix mit lauter Einsen auf der Hauptdiagonale, hat also nach Korollar 6.9 Determinante Eins. □

Dieses Resultat zeigt, dass Finden von Eigenwerten von beliebigen Matrizen über $\mathbb{K}$ äquivalent zum Finden von Nullstellen von beliebigen Polynomen über $\mathbb{K}$, und somit ein schwieriges Problem ist.

7.9. Die algebraische Vielfachheit. Da wir ja wissen, dass die Nullstellen des charakteristischen Polynoms genau die Eigenwerte einer linearen Abbildung (oder einer Matrix) sind, können wir nun natürlich das Konzept der Vielfachheit einer Nullstelle auf Eigenwerte übertragen.

Definition 7.9. Die *algebraische Vielfachheit* eines Eigenwertes λ einer Matrix $A \in M_n(\mathbb{K})$ bzw. einer linearen Abbildung f ist die Vielfachheit der Nullstelle λ des charakteristischen Polynoms p_A bzw. p_f.

Satz 7.9. *Sei V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $f : V \to V$ eine lineare Abbildung. Dann gilt:*

(1) Für jeden Eigenwert λ von f ist die algebraische Vielfachheit zumindest so groß wie die geometrischen Vielfachheit aus 7.2.

(2) Die Summe der algebraischen Vielfachheiten aller Eigenwerte von f ist $\leq n$ und falls das charakteristische Polynome von f in ein Produkt von Polynomen ersten Grades zerfällt, ist die Summe der algebraischen Vielfachheiten gleich n.

Beweis. (1) Sei V_λ^f der Eigenraum zum Eigenwert λ. Seine Dimension ist nach Definition die geometrische Vielfachheit von λ. Ist diese geometrische Vielfachheit gleich k, dann betrachten wir eine Basis $\{v_1, \dots, v_k\}$ für V_λ^f und erweitern diese zu einer Basis $\mathcal{B} := \{v_1, \dots, v_n\}$ von V. Nach Konstruktion ist dann $f(v_i) = \lambda v_i$ für $i = 1, \dots, k$, also hat die Matrixdarstellung $[f]_\mathcal{B}$ von f bezüglich dieser Basis die Blockform $\begin{pmatrix} \lambda\mathbb{I}_k & A \\ 0 & B \end{pmatrix}$ für $A \in M_{k,n-k}(\mathbb{K})$ und $B \in M_{n-k,n-k}(\mathbb{K})$. Berechnen wir nun das charakteristische Polynom dieser Matrix, dann können wir k–mal nach der ersten Spalte entwickeln, und erhalten $p_f = (\lambda - x)^k \det(B - x\mathbb{I})$. Somit sehen wir, dass die algebraische Vielfachheit von λ zumindest gleich k ist.

(2) folgt sofort aus Proposition 7.7. □

7.10. Charakterisierung von Diagonalisierbarkeit. Mit Hilfe unserer Resultate über charakteristische Polynome können wir nun eine vollständige Charakterisierung diagonalisierbarer linearer Abbildungen beweisen:

Satz 7.10. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$. Dann ist eine lineare Abbildung $f : V \to V$ genau dann diagonalisierbar über $\mathbb{K}$, wenn*

(a) *das charakteristische Polynom p_f von f über $\mathbb{K}$ in ein Produkt von Polynomen ersten Grades zerfällt und*
(b) *für jeden Eigenwert von f die algebraische Vielfachheit mit der geometrischen Vielfachheit übereinstimmt.*

Beweis. Ist f diagonalisierbar, dann sei $\mathcal{B}$ eine Basis aus Eigenvektoren. Die Matrixdarstellung $[f]_\mathcal{B}$ ist dann eine Diagonalmatrix. Sind $\lambda_1, \dots, \lambda_n$ die Eintragungen auf der Hauptdiagonale, dann ist das charakteristische Polynom dieser Matrix (also das charakteristische Polynom von f) gegeben durch $(\lambda_1 - x) \cdots (\lambda_n - x)$ und zerfällt somit in Produkt von Polynomen ersten Grades. Ist λ einer der Eigenwerte, dann ist die Anzahl der Basisvektoren, die diesem Eigenwert entsprechen, genau die algebraische Vielfachheit des Eigenwertes und diese Vektoren sind linear unabhängig. Somit muss die Dimension des Eigenraumes V_f^λ mindestens gleich der algebraischen Vielfachheit von λ sein, und damit folgt Bedingung (b) aus Satz 7.9(1).

Nehmen wir umgekehrt an, dass die Bedingungen (a) und (b) erfüllt sind. Seien $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von f und sie m_i die algebraische Vielfachheit von λ_i. Aus Bedingung (a) und Satz 7.9 (2) folgt nun, dass $\sum m_i = n$ gilt, und wegen Bedingung (b) ist m_i gleich der geometrischen Vielfachheit von λ_i. Damit ist aber f nach Satz 7.2 diagonalisierbar. □

Damit können wir nun ein vollständiges Schema angeben, wie man entscheidet, ob eine Matrix $A \in M_n(\mathbb{K})$ diagonalisierbar ist, und wie man sie explizit diagonalisiert, d.h. eine invertierbare Matrix T findet, sodass TAT^{-1} eine Diagonalmatrix ist.

(i) Berechne das charakteristische Polynom $p_A \in \mathbb{K}[x]$ und finde seine Nullstellen. Zerfällt dieses Polynom nicht in ein Produkt von Polynomen ersten Grades, dann kann A nicht diagonalisierbar sein. Ansonsten seien λ_i die verschiedenen Eigenwerte von A und m_i ihre algebraischen Vielfachheiten.

(ii) Sind alle $m_i = 1$, also alle λ_i verschieden, dann ist A diagonalisierbar, also ähnlich zu einer Diagonalmatrix mit den Eintragungen $\lambda_1, \dots, \lambda_n$ auf der Hauptdiagonale. Ist das nicht der Fall, dann bestimme für jene m_i, die größer als 1 sind, die geometrische Vielfachheit n_i von λ_i. Diese Vielfachheit ist die Dimension des Kerns der linearen Abbildung $A - \lambda_i\mathbb{I}$, also gilt $n_i = n - \operatorname{rg}(A - \lambda\mathbb{I})$ und dieser Rang kann mit Hilfe des Gauß'schen Algorithmus bestimmt werden. Die Matrix A ist nun genau dann diagonalisierbar, wenn $n_i = m_i$ für alle i mit $m_i > 1$ gilt.

(iii) Möchte man für diagonalisierbares A eine invertierbare Matrix $T \in M_n(\mathbb{K})$ finden, sodass TAT^{-1} diagonal ist, dann geht man wie folgt vor: Für jeden Eigenwert λ_i bestimme mit Hilfe des Gauß'schen Algorithmus eine Basis $\{v_1^i, \dots, v_{n_i}^i\}$ für den Raum der Lösungen des homogenen lineare Gleichungssystems $(A - \lambda_i\mathbb{I})x = 0$. Dann wissen wir aus unseren allgemeinen Resultaten, dass $\{v_1^1, \dots, v_{n_k}^k\}$ eine Basis $\mathcal{B}$ für $\mathbb{K}^n$ ist. Bezeichnen wir mit $\mathcal{S}$ die Standardbasis, dann ist die Matrix mit Spaltenvektoren $v_1^1, \dots, v_{n_k}^k$ gerade $[\mathrm{id}]_{\mathcal{S}}^{\mathcal{B}}$. Bezeichnen wir mit f die lineare Abbildung $x \mapsto Ax$, dann ist $A = [f]_{\mathcal{S}}$, und nach Korollar 4.17 ist damit $[\mathrm{id}]_{\mathcal{B}}^{\mathcal{S}} A [\mathrm{id}]_{\mathcal{S}}^{\mathcal{B}} = [f]_{\mathcal{B}}$, also eine Diagonalmatrix. Da $[\mathrm{id}]_{\mathcal{B}}^{\mathcal{S}} = ([\mathrm{id}]_{\mathcal{S}}^{\mathcal{B}})^{-1}$ gilt, erhält man die Matrix T indem man mit Hilfe des Algorithmus aus 3.9 die Matrix mit Spaltenvektoren $v_1^1, \dots, v_{n_k}^k$ invertiert.

Hat man es mit einer linearen Abbildung $f : V \to V$ auf einem beliebigen endlichdimensionalen Vektorraum zu tun, dann geht man analog vor. Zunächst wählt man eine beliebige Basis $\mathcal{B}$ von V und bestimmt die Matrixdarstellung $A := [f]_{\mathcal{B}}$. Wie oben kann man dann aus A das charakteristische Polynom $p_A = p_f$, sowie die Eigenwerte und ihre algebraischen und geometrischen Vielfachheiten bestimmen. Die Lösungen von $(A - \lambda_i\mathbb{I})x = 0$ sind dann gerade die Koordinatendarstellungen von Eigenvektoren bezüglich der Basis $\mathcal{B}$, und damit kann man dann eine Basis $\mathcal{B}'$ von V angeben, die aus Eigenvektoren für f besteht.

Beispiel 7.10. (1) Betrachten wir die Matrix $A = \begin{pmatrix} 3 & -1 & -1 \\ 2 & 1 & -2 \\ 0 & -1 & 2 \end{pmatrix} \in M_3(\mathbb{R})$. Nach der Regel von Sarrus erhält man für das charakteristische Polynom

$$p_A = (3-x)(1-x)(2-x) + 2 - 2(3-x) + 2(2-x) = (3-x)(2-x)(1-x).$$

Somit hat A drei verschiedene Eigenwerte, nämlich 1, 2 und 3 und muss daher diagonalisierbar sein. Eigenvektoren zu diesen Eigenwerten kann man leicht durch Lösen von linearen Gleichungssystemen bestimmen. Betrachten wir etwa den Eigenwert 1, dann müssen wir die Gleichung $Bx = 0$ für $B = \begin{pmatrix} 2 & -1 & -1 \\ 2 & 0 & -2 \\ 0 & -1 & 1 \end{pmatrix}$ lösen. Wenden wir darauf den Gauß'schen Algorithmus an, dann erhalten wir

$$\begin{pmatrix}2 & -1 & -1\\ 2 & 0 & -2\\ 0 & -1 & 1\end{pmatrix} \mapsto \begin{pmatrix}2 & -1 & -1\\ 0 & 1 & -1\\ 0 & -1 & 1\end{pmatrix} \mapsto \begin{pmatrix}2 & -1 & -1\\ 0 & 1 & -1\\ 0 & 0 & 0\end{pmatrix}.$$

Setzt man $x_3 = 1$, so erhält man $x_2 = 1$ und $2x_1 = 2$, also $x_1 = 1$, und damit ist $(1, 1, 1)$ ein Eigenvektor zum Eigenwert 1.

(2) Sei nun $A = \begin{pmatrix}2 & -2 & 1\\ 3 & -1 & 1\\ -1 & 2 & 0\end{pmatrix} \in M_3(\mathbb{R})$. Mit Hilfe der Regel von Sarrus berechnet man das charakteristische Polynom von A als

$$p_A = (2-x)(-1-x)(-x) + 6 + 2 + (-1-x) - 2(2-x) - 6x = -x^3 + x^2 - 3x + 3.$$

Offensichtlich ist 1 eine Nullstelle dieses Polynoms, und dividiert man p_A durch $(x - 1)$, dann erhält man $-(x^2 + 3)$. Das letztere Polynom hat über $\mathbb{R}$ offensichtlich keine Nullstellen, also kann A über $\mathbb{R}$ nicht diagonalisierbar sein. Über $\mathbb{C}$ hat dieses Polynom natürlich die Nullstellen $\pm i\sqrt{3}$, also ist A über $\mathbb{C}$ diagonalisierbar mit Eigenwerten 1, $i\sqrt{3}$ und $-i\sqrt{3}$. Eigenvektoren kann man wie zuvor durch Lösen von linearen Gleichungssystemen finden.

(3) Ist $A = \begin{pmatrix}2 & -1 & -1\\ -1 & 2 & -1\\ -1 & -1 & 2\end{pmatrix} \in M_3(\mathbb{R})$, dann erhalten wir als charakteristisches Polynom $(2-x)^3 - 1 - 1 - (2-x) - (2-x) - (2-x) = -x^3 + 6x^2 - 3x = -x(x-3)^2$. Damit hat A die Eigenwerte 0 und 3 mit algebraischen Vielfachheiten 1 bzw. 2. Offensichtlich ist $(1, 1, 1)$ ein Eigenvektor zum Eigenwert 0. Für den Eigenwert 3 müssen wir die Gleichung $Bx = 0$ lösen, wobei alle drei Zeilen von B gerade $(-1, -1, -1)$ sind. Das bedeutet aber sofort, dass B den Rang 1 hat, also der Eigenwert 3 auch geometrische Vielfachheit 2 besitzt. Somit ist A diagonalisierbar. Eine Basis des Eigenraumes zum Eigenwert 3 erhält man natürlich, indem man in obigen Gleichungssystem die Lösungen mit $x_2 = 1$ und $x_3 = 0$ bzw. die mit $x_2 = 0$, $x_3 = 1$ berechnet. Das liefert aber gerade $(-1, 1, 0)$ und $(-1, 0, 1)$. Betrachten wir also die Basis $\mathcal{B} = \{(1, 1, 1), (-1, 1, 0), (-1, 0, 1)\}$ und bezeichnen wir mit $\mathcal{S}$ die Standardbasis, dann ist $[\mathrm{id}]_{\mathcal{B},\mathcal{S}} = \begin{pmatrix}1 & -1 & -1\\ 1 & 1 & 0\\ 1 & 0 & 1\end{pmatrix}$. Berechnet man mit Hilfe des Algorithmus aus 3.9 die inverse dieser Matrix, dann erhält man die explizite Diagonalisierung

$$\begin{pmatrix}1/3 & 1/3 & 1/3\\ -1/3 & 2/3 & -1/3\\ -1/3 & -1/3 & 2/3\end{pmatrix}\begin{pmatrix}2 & -1 & -1\\ -1 & 2 & -1\\ -1 & -1 & 2\end{pmatrix}\begin{pmatrix}1 & -1 & -1\\ 1 & 1 & 0\\ 1 & 0 & 1\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 3\end{pmatrix}$$

(4) Betrachte den Raum $\mathbb{R}_n[x]$ der Polynome vom Grad $\leq n$ und den Ableitungsoperator $D : \mathbb{R}_n[x] \to \mathbb{R}_n[x]$. Man kann D einfach algebraisch durch $D(\sum a_j x^j) := \sum j a_j x^{j-1}$ definieren. Betrachten wir die Basis $\mathcal{B} = \{1, x, x^2, \dots, x^n\}$, dann ist $D(x^j) = jx^{j-1}$, also erhalten wir $A := [D]_{\mathcal{B}} = (a_{ij})$, wobei $a_{ij} = 0$ für $j \neq i + 1$ und $a_{i,i+1} = i$. Betrachten wir nun $A - y\mathbb{I} \in M_n(\mathbb{K}[y])$, dann ist das eine obere Dreiecksmatrix, und alle Eintragungen auf der Hauptdiagonale sind gleich $-y$. Nach Korollar 6.6 ist $p_D = \det(A - y\mathbb{I}) = (-1)^n y^n$. Das bedeutet aber, dass 0 der einzige Eigenwert von D ist. Wäre D diagonalisierbar, dann gäbe es eine Basis aus Eigenvektoren zum Eigenwert 0, also wäre D die Nullabbildung. Tatsächlich liegt hier ein extremer Fall vor: Die Matrix

A hat ja Zeilenstufenform, und daraus sieht man sofort, dass $\mathrm{rg}(A) = \mathrm{rg}(D) = n-1$ gilt. Damit ist aber die Dimension des Kerns von D, die ja nach Definition genau die geometrische Vielfachheit des Eigenwertes 0 ist, gleich Eins.

Existenz von Nullstellen und Triangulierbarkeit

Zum Abschluss des Kapitels wenden wir uns noch Resultaten über Existenz von Nullstellen von Polynomen und Konsequenzen solcher Resultate für die linare Algebra zu.

7.11. Existenz von Nullstellen. Es ist leicht zu sehen, dass die Frage der Existenz von Nullstellen eher schwierig ist. Insbesondere hängt die Existenz von Nullstellen stark von dem Körper ab, den wir betrachten. Zum Beispiel können wir ein Polynom $p \in \mathbb{Q}[x]$ auch als Polynom mit Koeffizienten in $\mathbb{R}$ oder $\mathbb{C}$ betrachten. Das Polynom $p = x^2 - 2$ zum Beispiel hat in $\mathbb{R}$ die Nullstellen $\sqrt{2}$ und $-\sqrt{2}$, aber es hat keine Nullstellen in $\mathbb{Q}$, weil $\sqrt{2}$ irrational ist. Natürlich sind diese beiden Nullstellen auch die einzigen Nullstellen in $\mathbb{C}$. Das Polynom $p = x^2 + 1 \in \mathbb{Q}[x]$ hat weder in $\mathbb{Q}$ noch in $\mathbb{R}$ Nullstellen, währen in $\mathbb{C}$ die Nullstellen i und $-i$ sind.

Überraschenderweise kommen Existenzresultate für Nullstellen eher aus der Analysis als aus der Algebra. So können wir zum Beispiel leicht sehen, dass ein Polynom $p \in \mathbb{R}[x]$ mit $\deg(p) = 3$ in $\mathbb{R}$ immer eine Nullstelle haben muss. Betrachten wir so ein Polynom, dann können wir natürlich durch den führenden Koeffizienten dividieren ohne die Nullstellen zu ändern, also genügt es, ein Polynom der Form $p = x^3 + ax^2 + bx + c$ für $a, b, c \in \mathbb{R}$ zu betrachten. Betrachten wir die zugehörige Polynomfunktion $p : \mathbb{R} \to \mathbb{R}$, dann ist $\lim_{\lambda \to \pm\infty} p(\lambda) = \pm\infty$. Für $\lambda \neq 0$ können wir nämlich $p(\lambda)$ als $\lambda^3(1 + \frac{a}{\lambda} + \frac{b}{\lambda^2} + \frac{c}{\lambda^3})$ schreiben und der Ausdruck in der Klammer geht für $\lambda \to \pm\infty$ offensichtlich egen Eins. Insbesondere bedeutet das, dass es Zahlen $\lambda_1, \lambda_2 \in \mathbb{R}$ mit $\lambda_1 < 0$ und $\lambda_2 > 0$ gibt, die außerdem $p(\lambda_1) < 0$ und $p(\lambda_2) > 0$ erfüllen. Nach dem Zwischenwertsatz der Analysis muss die stetige Funktion p auf dem Intervall (λ_1, λ_2) auch alle Werte annehmen, die zwischen $p(\lambda_1)$ und $p(\lambda_2)$ liegen, also muss insbesondere zwischen λ_1 und λ_2 eine Nullstelle von p liegen. Analog sieht man die Existenz mindestens einer Nullstelle für jedes Polynom ungeraden Grades über $\mathbb{R}$.

Der wichtigste Satz über Nullstellen von Polynomen ist der *Fundamentalsatz der Algebra*, der zeigt, dass über dem Körper $\mathbb{C}$ keine Probleme ohne Nullstellen auftreten. Über $\mathbb{C}$ kann jedes Polynom als Produkt von Polynomen ersten Grades geschrieben werden und damit hat ein Polynom $p \in \mathbb{C}[x]$ vom Grad n genau n Nullstellen (mit Vielfachheit gezählt). Auch bei diesem Satz ist der Beweis eher analytisch. Daher passt er nicht wirklich in die Vorlesung und wird hier nur der Vollständigkeit halber präsentiert. Wir setzen im Beweis einige Kenntnisse aus der Analysis voraus.

Satz 7.11 (Fundamentalsatz der Algebra). *Sei $p \in \mathbb{C}[x]$ ein Polynom mit $\deg(p) \geq 1$. Dann besitzt p mindestens eine Nullstelle in $\mathbb{C}$.*

Beweis. Die Vorschrift $f(z) := |p(z)|$ definiert eine stetige Funktion $f : \mathbb{C} \to \mathbb{R}$, und $f(z) \geq 0$ für alle $z \in \mathbb{C}$. Der erste Schritt des Beweises ist zu zeigen, dass diese Funktion ein globales Minimum besitzt, d.h. ein $z_0 \in \mathbb{C}$ existiert, sodass $f(z_0) \leq f(z)$ für alle $z \in \mathbb{C}$ gilt. Falls p eine Nullstelle hat ist das offensichtlich, also nehmen wir an, dass $f(z) > 0$ für alle z gilt.

Für eine reelle Zahl $R > 0$ betrachten wir die Teilmenge $B_R := \{z \in \mathbb{C} : |z| \leq R\} \subset \mathbb{C} = \mathbb{R}^2$. Diese Teilmenge ist beschränkt und abgeschlossen, also kompakt, also besitzt die Einschränkung der stetigen Funktion f auf B_R nach dem Satz vom Maximum ein

Maximum und ein Minimum. Schreiben wir andererseits $p = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, dann gilt für $z \in \mathbb{C} \setminus \{0\}$ die Gleichung

$$p(z) = a_n z^n \left(1 + \frac{a_{n-1}}{a_n z} + \cdots + \frac{a_1}{a_n z^{n-1}} + \frac{a_0}{a_n z^n}\right).$$

Nehmen wir den Betrag der rechten Seite, dann erhalten wir $|a_n||z|^n|(1+\dots)|$, und für $|z| \to \infty$ geht das gegen ∞. Somit gibt es für jedes $C > 0$ eine Zahl $R > 0$, sodass $f(z) > C$ für alle $z \in \mathbb{C}$ mit $|z| > R$ gilt. Sei nun $\tilde{z}_0$ der Punkt, an dem die Einschränkung von f auf B_1 ihr Minimum annimmt und sei $C := f(\tilde{z}_0) > 0$ dieses Minimum. Dann gibt es dazu einen entsprechenden Wert $\tilde{R}$. Sei nun $R := \max\{1, \tilde{R}\}$ und $z_0 \in B_R$ ein Punkt, an dem die Einschränkung von f auf B_R ihr Minimum annimmt. Weil $R \geq 1$ ist, ist $f(z_0) \leq C$. Nach Definition ist $f(z_0) \leq f(z)$ für alle $z \in B_R$. Andererseits ist nach Konstruktion $f(z) > C = f(z_0)$ für alle $z \in \mathbb{C}$ mit $|z| > R$. Damit ist aber $f(z_0) \leq f(z)$ für alle $z \in \mathbb{C}$, also nimmt f ein absolutes Minimum bei z_0 an.

Der zweite Schritt des Beweises besteht nun darin zu zeigen, dass ein positiver Wert niemals ein globales Minimum von f sein kann. Nehmen wir zunächst an, dass $z_0 = 0$ gilt. Natürlich ist $a_0 = p(0) \neq 0$. Weil der Grad von p mindestens Eins ist, gibt es einen kleinsten Index $m \in \{1, \dots, n\}$, sodass $a_m \neq 0$ ist. Dann ist $p = a_0 + a_m x^m + x^{m+1}q$, wobei $q \in \mathbb{C}[x]$ gegeben ist durch $a_{m+1} + a_{m+2}x + \cdots + a_n x^{n-m-1}$. Wir können $-\frac{a_0}{a_m} \in \mathbb{C}$ in Polarkoordinaten schreiben, also finden wir eine eindeutige reelle Zahl $r > 0$ und einen eindeutigen Winkel $\varphi \in [0, 2\pi)$, sodass $-\frac{a_0}{a_m} = re^{i\varphi}$ gilt. Setzen wir $z_1 = r^{1/m}e^{i(\varphi/m)}$, dann ist $z_1^m = -\frac{a_0}{a_m}$. Für $\lambda \in [0,1] \subset \mathbb{R}$ betrachten wir nun $p(\lambda z_1) = a_0 + a_m \lambda^m z_1^m + \lambda^{m+1} z_1^{m+1} q(\lambda z_1)$. Nach Konstruktion ist der zweite Summand gerade $-a_0\lambda^m$, und wir erhalten

$$p(\lambda z_1) = a_0 \left(1 - \lambda^m + a_0^{-1}\lambda^{m+1} z_1^{m+1} q(\lambda z_1)\right).$$

Aus Stetigkeitsgründen gibt es eine Zahl $C > 0$, sodass $|a_0^{-1} z_1^{m+1} q(\lambda z_1)| < C$ für alle $\lambda \in [0,1]$ gilt. Damit erhalten wir aber $|p(\lambda z_1)| \leq |a_0||(1 - \lambda^m + C\lambda^{m+1})|$. Für $0 < \lambda < 1/C$ ist aber $0 < 1 - C\lambda < 1$, und damit sicher auch $0 < 1 - \lambda^m(1 - C\lambda) < 1$. Damit ist aber $|p(\lambda z_1)| < |a_0| = p(0)$, also 0 kein globales Minimum.

Im Fall von beliebigem $z_0 \in \mathbb{C}$ bemerken wir, dass es Koeffizienten $b_0, \dots, b_n$ gibt, sodass $a_n z^n + \cdots + a_0 = b_n(z - z_0)^n + b_{n-1}(z - z_0)^{n-1} + \cdots + b_1(z - z_0) + b_0$ gilt. Dazu muss man nur $z = z_0 + (z - z_0)$ zu den Potenzen $2, \dots, n$ erheben, und dann einsetzen. Insbesondere sehen wir, dass $b_0 = p(z_0)$ gelten muss. Die Tatsache, dass $z \mapsto |a_n z^n + \cdots + a_1 z + a_0|$ ein globales Minimum bei z_0 hast, übersetzt sich nun natürlich direkt in die Tatsache, dass $z \mapsto |b_n z^n + \cdots + b_1 z + b_0|$ ein globales Minimum bei 0 hat. Nach dem letzten Schritt folgt daraus aber $0 = b_0 = p(z_0)$. □

Korollar 7.11. *Sei $p \in \mathbb{C}[x]$ ein Polynom mit* $\deg(p) = n > 0$. *Dann gibt es Elemente $a_n \in \mathbb{C} \setminus \{0\}$, $\lambda_1, \dots, \lambda_k \in \mathbb{C}$ und $m_1, \dots, m_k \in \mathbb{N}$ mit $m_1 + \cdots + m_k = n$, sodass $p = a_n(x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$. Insbesondere zerfällt jedes solche Polynom in ein Produkt von Polynomen ersten Grades und es hat genau* $\deg(p)$ *viele Nullstellen, wenn man die Nullstellen mit Vielfachheit zählt.*

Beweis. Durch Induktion nach n. Für $n = 1$ ist $p = a_1 x + a_0$ mit $a_1 \neq 0$, und das können wir als $a_1(x - \lambda_1)$ schreiben, wobei $\lambda_1 = -\frac{a_0}{a_1}$.

Nehmen wir also an, dass $n > 1$ gilt und der Satz für Polynome vom Grad $n - 1$ schon bewiesen wurde. Dann hat p nach dem Satz eine Nullstelle $\lambda \in \mathbb{C}$, also gibt es nach Satz 7.6 ein Polynom $q \in \mathbb{C}[x]$ vom Grad $n - 1$, sodass $p = (x - \lambda)q$ gilt. Wendet man auf q die Induktionsvoraussetzung an, dann folgt die Behauptung. □

Bemerkung 7.11. (1) Ein Körper $\mathbb{K}$ in dem jedes Polynom $p \in \mathbb{K}[x]$ in ein Produkt von Polynomen ersten Grades zerfällt heißt *algebraisch abgeschlossen*. Der Fundamentalsatz der Algebra sagt also gerade, dass $\mathbb{C}$ algebraisch abgeschlossen ist. Die Körper $\mathbb{Q}$ und $\mathbb{R}$ sind nicht algebraisch abgeschlossen, weil etwa das Polynom $x^2 + 1$ in diesen Körpern keine Nullstelle besitzt. Auch die Körper $\mathbb{Z}_p$ für p eine Primzahl sind nicht algebraisch abgeschlossen.

(2) Der Beweis des Fundamentalsatzes der Algebra ist nicht konstruktiv. Er gibt uns also keinen Hinweis, wie wir Nullstellen eines gegebenen Polynoms über $\mathbb{C}$ konkret finden können. Im Allgemeinen ist das auch ein schwieriges Problem. Für quadratische Polynome gibt es natürlich die schon aus der Schule bekannte Formel: Wir können durch den führenden Koeffizienten dividieren ohne die Nullstellen zu ändern, also dürfen wir ohne Beschränkung der Allgemeinheit ein Polynom der Form $x^2 + ax + b$ betrachten. Ist das gleich $(x - \lambda_1)(x - \lambda_2)$, dann folgt sofort $a = -(\lambda_1 + \lambda_2)$ und $b = \lambda_1\lambda_2$. Damit ist aber $(\lambda_1 - \lambda_2)^2 = (\lambda_1 + \lambda_2)^2 - 4\lambda_1\lambda_2 = a^2 - 4b$. Daraus erhalten wir aber sofort $\lambda_{1,2} = \frac{\lambda_1+\lambda_2}{2} \pm \frac{\lambda_1-\lambda_2}{2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}$.

Für Polynome dritten und vierten Grades gibt es noch allgemeine (aber deutlich kompliziertere) Lösungsformeln. Für Polynome vom Grad ≥ 5 wurde um 1820 von N.H. Abel und E. Galois bewiesen, das man Lösungen nicht durch endlich Anwendung der Grundrechnungsarten und des Wurzelziehens bilden kann. Insbesondere gibt es also keine allgemeinen Lösungsformeln, die nur diese Operationen benutzen. Diese Beweise beruhen auf der (heute so genannten) Galois–Theorie, die eine Verbindung zwischen der Theorie der Körper, der Theorie der Polynome und der Gruppentheorie darstellt.

Wir werden uns nicht weiter mit dem Problem des Findens von Nullstellen komplexer Polynome beschäftigen, weil die vielen verfügbaren (algebraischen und/oder numerischen) Methoden zum Finden solcher Nullstellen über den Rahmen dieser Vorlesung weit hinausgehen würden. Im Fall von Polynomen vom Grad ≥ 3 werden wir uns auf Beispiele beschränken, in denen man genügend viele Nullstellen erraten kann, um das Problem durch Divisionen auf quadratische Polynome zurückführen zu können.

7.12. Nullstellen reeller Polynome. Der Fundamentalsatz der Algebra liefert uns auch Informationen über reelle Polynome. Ist nämlich $p \in \mathbb{R}[x]$ ein Polynom vom Grad $n \geq 1$, dann können wir p wegen $\mathbb{R} \subset \mathbb{C}$ natürlich auch als komplexes Polynom auffassen. Nach Korollar 7.11 finden wir als n (nicht notwendig verschiedene) komplexe Zahlen $\lambda_1, \dots, \lambda_n$ und $a_n \in \mathbb{C}$, sodass $p = a_n(x - \lambda_1) \cdots (x - \lambda_n)$ gilt. Aus dieser Formel ist offensichtlich, dass a_n gerade der führende Koeffizient von p ist, also gilt $a_n \in \mathbb{R}$. Ist nun $p = a_n x^n + \cdots + a_1 x + a_0$, dann ist $a_i \in \mathbb{R}$, also $\bar{a}_i = a_i$. Ist $\lambda \in \mathbb{C}$ eine Nullstelle von p, dann gilt $0 = a_n\lambda^n + \cdots + a_1\lambda + a_0$. Da die komplexe Konjugation mit allen Körperoperationen auf $\mathbb{C}$ verträglich ist, erhalten wir daraus

$$0 = \bar{0} = \bar{a}_n\bar{\lambda}^n + \cdots + \bar{a}_1\bar{\lambda} + \bar{a}_0 = a_n\bar{\lambda}^n + \cdots + a_1\bar{\lambda} + a_0,$$

also ist mit λ automatisch auch $\bar{\lambda}$ eine Nullstelle von p. Andererseits ist $(x-\lambda)(x-\bar{\lambda}) = x^2 - (\lambda+\bar{\lambda})x + \lambda\bar{\lambda}$, und $\lambda + \bar{\lambda} = 2\operatorname{Re}(\lambda) \in \mathbb{R}$ und $\lambda\bar{\lambda} = |\lambda|^2 \in \mathbb{R}$, also hat dieses Polynom reelle Koeffizienten. Somit erhalten wir

Korollar 7.12. *Sei $p \in \mathbb{R}[x]$ ein Polynom vom Grad $n \geq 1$. Dann gibt es eine Zahl $k \leq n$, sodass $n - k$ gerade ist, k (nicht notwendig verschiedene) Zahlen $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ und $\ell := \frac{n-k}{2}$ monische Polynome $q_1, \dots, q_\ell \in \mathbb{R}[x]$ mit $\deg(q_i) = 2$, die keine Nullstellen über $\mathbb{R}$ besitzen, sodass*

$$p = a_n(t - \lambda_1) \cdots (t - \lambda_k)q_1 \cdots q_\ell.$$

Ein Polynom $p \in \mathbb{R}[x]$ zerfällt also genau dann in in Produkt von Polynomen ersten Grades, wenn alle Nullstellen des komplexen Polynoms p im Teilkörper $\mathbb{R} \subset \mathbb{C}$ liegen. Die Sprechweise "alle Nullstellen von p liegen in $\mathbb{K}$", anstatt "p zerfällt in ein Produkt von Polynomen ersten Grades" ist auch für allgemeine Körper üblich. Der Grund dafür ist, dass man zu einem beliebigen Körper $\mathbb{K}$ einen (im wesentlichen eindeutigen) algebraisch abgeschlossenen Körper $\tilde{\mathbb{K}}$ konstruieren kann, der $\mathbb{K}$ als Teilkörper enthält. Dieser Körper heißt der *algebraische Abschluss* von $\mathbb{K}$. Über $\tilde{\mathbb{K}}$ zerfällt dann jedes Polynom $p \in \mathbb{K}[x]$ in ein Produkt von Polynomen ersten Grades und das gilt genau dann auch über $\mathbb{K}$ wenn alle Nullstellen von p im Teilkörper $\mathbb{K} \subset \tilde{\mathbb{K}}$ liegen. Der Beweis für die Existenz des algebraischen Abschlusses geht weit über den Rahmen dieser Vorlesung hinaus, wir werden diese Tatsache aber gelegentlich noch verwenden.

7.13. Triangulierbarkeit über $\mathbb{C}$. Der Fundamentalsatz der Algebra impliziert natürlich sofort, dass jede lineare Abbildung $f : V \to V$ auf einem endlichdimensionalen Vektorraum V über $\mathbb{C}$ mindestens einen Eigenwert (und damit auch einen Eigenvektor) besitzt. Damit können wir aber leicht ein allgemeineres Resultat beweisen, nämlich dass über $\mathbb{C}$ jede lineare Abbildung auf einem endlichdimensionalen Vektorraum durch eine obere Dreiecksmatrix dargestellt werden kann. Dabei lernen wir Konzepte kennen, die eine wichtige Rolle in der Theorie der Normalformen spielen werden.

Definition 7.13. Sei V ein Vektorraum über $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. Ein Teilraum $W \subset V$ heißt *f–invariant* wenn $f(w) \in W$ für alle Elemente $w \in W$ gilt.

Ist $f : V \to V$ eine lineare Abildung und $W \subset V$ ein f–invarianter Teilraum, dann kann man natürlich die Einschränkung $f|_W : W \to W$ bilden. Betrachten wir andererseits den Quotientenraum V/W und die kanonische Surjektion $\pi : V \to V/W$ und $\pi \circ f : V \to V/W$. Für $w \in W$ ist dann $\pi(f(w)) = 0$, weil $f(w)$ in $W = \mathrm{Ker}(\pi)$ liegt. Damit gibt es aber nach der universellen Eigenschaft des Quotienten (Satz 5.11) eine eindeutige lineare Abbildung $\underline{f} : V/W \to V/W$, sodass $\underline{f} \circ \pi = \pi \circ f$ gilt.

Die Frage der Darstellbarkeit durch eine obere Dreiecksmatrix läßt sich nun schön in Termen invarianter Teilräume charakterisieren.

Lemma 7.13. *Sei V ein n–dimensionaler Vektorraum über einem Körper $\mathbb{K}$. Dann sind für eine lineare Abbildung $f : V \to V$ äquivalent:*

(1) Es gibt eine Basis $\mathcal{B}$ von V, sodass $[f]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist.

(2) Es gibt f–invariante Teilräume $W_1 \subset W_2 \subset \dots \subset W_{n-1} \subset V$, sodass $\dim(W_i) = i$ für alle $i = 1, \dots n$ gilt.

Beweis. Das ist eine fast offensichtliche Umformulierung. Ist $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis wie in (1), dann bedeutet die Tatsache, dass $[f]_{\mathcal{B}}$ obere Dreiecksgestalt hat, genau, dass für jedes $i = 1, \dots, n$ der Vektor $f(v_i)$ eine Lineakombination von $v_1, \dots, v_i$ ist. Definiert man nun $W_i := \langle v_1, \dots, v_i \rangle$, dann ist natürlich $\dim(W_i) = i$, weil die v_j linear unabhängig sind. Außerdem gilt $W_1 \subset W_2 \subset \dots \subset W_{n-1}$ nach Konstruktion und natürlich ist $f(v_j) \in W_i$ für alle $j \leq i$. Damit folgt aber $f(W_i) \subset W_i$ sofort.

Haben wir umgekehrt die W_i gegeben, dann wählen wir eine Baisis $\{v_1\}$ des eindimensionalen Raumes W_1, erweitern sie zu einer Basis $\{v_1, v_2\}$ von W_2 und so weiter bis wir eine Basis $\mathcal{B} := \{v_1, \dots, v_n\}$ für V haben. Die Tatsache, dass jedes W_i ein f–invarianter Teilraum ist, bedeutet aber gerade, dass $f(v_i)$ jeweils als Linearkombination der v_j mit $j \leq i$ geschrieben werden kann. Damit ist aber $[f]_{\mathcal{B}}$ eine obere Dreiecksmatrix. □

PROPOSITION 7.13. *Sei V ein endlichdimensionaler Vektorraum über einem algebraisch abgeschlossenen Körper $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. Dann gibt es eine Basis $\mathcal{B}$ für V, sodass $[f]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist.*

BEWEIS. Wir beweisen den Satz mit Induktion nach $n = \dim(V)$. Für $n = 1$ ist jede $n \times n$–Matrix eine obere Dreiecksmatrix, also nehmen wir an, dass $n > 1$ ist und der Satz für Räume der Dimension $n - 1$ bereits bewiesen wurde. Wir benutzen Bedingung (2) aus Lemma 7.13.

Wie wir bereits festgestellt haben, besitzt f einen Eigenwert (weil p_f eine Nullstelle besitzt) also finden wir einen Vektor$v_1 \neq 0$ und ein Element $\lambda \in \mathbb{K}$, sodass $f(v_1) = \lambda v_1$ gilt. Setzen wir nunr $W_1 := \mathbb{K} \cdot v_1$, dann ist W_1 offensichtlich ein f–invarianter Teilraum von V. Wir bilden den Quotienten $\pi : V \to V/W_1$ und die Abbildung $\underline{f} : V/W_1 \to V/W_1$, die $\underline{f} \circ \pi = \pi \circ f$ erfüllt. Nach dem Dimensionssatz für Quotienten (Satz 5.10) ist $\dim(V/W_1) = n - 1$. Damit können wir die Induktionsvoraussetzung auf $\underline{f}$ anwenden und erhalten $\underline{f}$–invariante Teilräume

$$\underline{W}_1 \subset \cdots \subset \underline{W}_{n-2} \subset V/W_1$$

mit $\dim(\underline{W}_i) = i$. Nun setzen wir für $i = 2, \dots, n-1$ $W_i := \pi^{-1}(\underline{W}_{i-1})$. Da $W_1 = \pi^{-1}(\{0\})$ erhalten wir eine Folge

$$W_1 \subset W_2 \subset \cdots \subset W_{n-1} \subset V$$

von Teilräumen. Für $i \geq 2$ ist $\pi|_{W_i} : W_i \to \underline{W}_{i-1}$ eine offensichtlich surjektive lineare Abbildung mit Kern W_1, gilt $\dim(W_i) = i$ nach dem Dimensionssatz für lineare Abbildungen. Für $w \in W_i$ mit $i \geq 2$ gilt $\pi(f(w)) = \underline{f}(\pi(w))$. Nach Voraussetzung liegt aber $\pi(w)$ im $\underline{f}$–invariant Teilraum $\underline{W}_{i-1}$. Somit ist $\pi(f(w)) \in \underline{W}_{i-1}$, also $f(w) \in W_i$. Damit sind alle Bedingungen von Teil (2) des Lemmas erfüllt. $\square$

KAPITEL 8

Normen und innere Produkte

Mit diesem Kapitel verlassen wir vorübergehend die allgemeinen Körper und schränken uns auf die Fälle $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$ ein. Dafür kehren wir zu Themen zurück, die im Ansatz schon aus der Schule bekannt sind, nämlich zu inneren Produkten und den damit verbundenen Konzepten der Länge von Vektoren und des Winkels zwischen zwei Vektoren. Für einen Längenbegriff gibt es eine allgemeinere Version, die sogenannten Normen.

Normierte Räume

Zunächst beschäftigen wir uns mit Normen, die man als eine allgemeine Version eines Längenbegriffs für Vektoren verstehen kann. Eine Norm liefert eine Distanzfunktion, mit deren Hilfe Konvergenz und Stetigkeit definiert werden können, was eine Verbindung zur Analysis liefert. Der Normbegriff ist einerseits als Ausblick auf unendlichdimensionale Räume wichtig. Im Endlichdimensionalen ist insbesondere der Fall von Räumen der Form $L(V, V)$ interessant, wo man mit Hilfe des Normbegriffs gewisse stetige Funktionen auf lineare Abbildungen bzw. Matrizen anwenden kann. Wir werden uns in diesem Abschnitt eher kurz fassen und Grundkenntnisse aus der Analysis voraussetzen.

8.1. Normen. Die allgemeine Definition einer Norm stellt intuitiv ziemlich einsichtige Anforderungen an einen "Längenbegriff" für die Elemente eines Vektorraumes:

Definition 8.1. (1) Sei V ein Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$. Eine *Norm* auf V ist eine Funktion $\| \ \| : V \to \mathbb{R}$ mit folgenden Eigenschaften:
(N1) $\|v\| \geq 0$ für alle $v \in V$ und $\|v\| = 0 \iff v = 0$.
(N2) Für $v \in V$ und $\lambda \in \mathbb{K}$ ist $\|\lambda v\| = |\lambda| \cdot \|v\|$.
(N3) Für $v, w \in V$ ist $\|v + w\| \leq \|v\| + \|w\|$.

(2) Ein *normierter Vektorraum* über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ ist ein $\mathbb{K}$–Vektorraum V zusammen mit einer Norm auf V.

Die Bedingung (N1) wird üblicherweise als "Nichtnegativität" bezeichnet, Bedingung (N2) als "positive Homogenität". Denkt man an die geometrische Interpretation der Vektoraddition, dann sagt (N3) gerade, dass die Länge einer Seite eines Dreiecks höchstens so groß sein kein, wie die Summe der Längen der beiden anderen Seiten, weshalb diese Bedingung als *Dreiecksungleichung* bezeichnet wird.

Beispiel 8.1. (1) Das wichtigste Beispiel einer Norm auf einem endlichdimensionalen Vektorraum ist die aus der Schule und der Analysis bekannte *euklidische Norm* $\| \ \|_2$ auf $\mathbb{R}^n$ und $\mathbb{C}^n$. Diese Norm ist definiert durch $\|(x_1, \dots, x_n)\|_2 := \sqrt{\sum_{j=1}^n |x_j|^2}$, wobei man für $\mathbb{K} = \mathbb{R}$ natürlich den Betrag auf der rechten Seite weglassen kann. Die Eigenschaften (N1) und (N2) sind dann offensichtlich erfüllt. Die Dreiecksunglichung (N3) werden wir in einer allgemeineren Version im Abschnitt über innere Produkte beweisen.

(2) Betrachten wir wieder $\mathbb{R}^n$ oder $\mathbb{C}^n$ und die Funktion $\| \ \|_1$, die definiert ist durch $\|(x_1, \dots, x_n)\|_1 = \sum_{j=1}^n |x_j|$. Hier sind wieder (N1) und (N2) offensichtlich, und (N3)

folgt sofort aus der Dreiecksungleichung für den Absolutbetrag auf $\mathbb{R}$, siehe Proposition 6.4.12 von [EMA].

Analog definiert $\|(x_1,\dots,x_n)\|_\infty = \max\{|x_1|,\dots,|x_n|\}$ eine Norm auf $\mathbb{K}^n$. Auch dafür sind (N1) und (N2) offensichtlich. Für (N3) bemerkt man, dass für jedes $j = 1,\dots,n$ die Ungleichung $|x_j+y_j| \le |x_j|+|y_j|$ gilt. Geht man auf der rechten Seite zu den Maxima über, dann bleibt die Ungleichung richtig, also ist $|x_j+y_j| \le \|x\|_\infty + \|y\|_\infty$, und da das für alle j gilt, folgt (N3).

(3) Die Beispiele (1) und (2) sind tatsächlich Spezialfälle einer Familie $\| \ \|_p$ von Normen für $p \in [1,\infty) \subset \mathbb{R}$, wobei $\|(x_1,\dots,x_n)\|_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$. Hier sind immer noch (N1) und (N2) offensichtlich, (N3) ist allgemein schwieriger zu beweisen (Hölder–Ungleichung). Alle diese Normen besitzen Verallgemeinerungen auf geeignete (unendlichdimensionale) Räume von Folgen. Insbesondere macht jede dieser Normen auf dem Raum aller endlichen Folgen in $\mathbb{K}$ Sinn.

(4) In der Analysis und in der Funktionalanalysis spielen verschiedene Normen auf Funktionenräumen eine wichtige Rolle. Wir bemerken hier nur zum Beispiel, dass auf dem Raum $C([a,b],\mathbb{R})$ aller stetigen Funktionen auf einem abgeschlossenen Intervall $[a,b] \subset \mathbb{R}$ sowohl $\sup_{x\in[a,b]} |f(x)|$ als auch $\int_a^b |f(x)|dx$ eine Norm definiert.

Hat man eine Norm $\| \ \|$ auf V, dann kann man eine zugehörige Distanzfunktion $d : V \times V \to \mathbb{R}$ definieren, indem man $d(v,w) := \|w-v\|$ setzt. Es ist leicht zu zeigen, dass diese Distanzfunktion die aus der Analysis bekannten Eigenschaften erfüllt:

Proposition 8.1. *Sei V ein Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, sei $\| \ \|$ eine Norm auf V und sei $d : V\times V \to \mathbb{R}$ die zugehörige Distanzfunktion. Dann ist (V,d) ein metrischer Raum (im Sinne der Analysis), d.h. es für alle $u,v,w \in V$ gilt:*

(D1) *$d(v,w) \ge 0$ und $d(v,w) = 0$ gilt genau dann, wenn $v = w$ gilt.*
(D2) *$d(v,w) = d(w,v)$ (Symmetrie)*
(D3) *$d(u,w) \le d(u,v) + d(v,w)$ (Dreiecksungleichung)*

Beweis. Nach (N1) ist $d(v,w) = \|w-v\| \ge 0$, und Gleichheit gilt nur für $0 = w-v$, also für $v = w$, und (D1) ist gezeigt. Für (D2) bemerken wir, dass $v-w = (-1)(w-v)$ gilt, also nach (N2) $\|v-w\| = |-1|\|w-v\|$ gilt. Wegen $w-u = (w-v)+(v-u)$ folgt auch (D3) sofort aus (N3). □

Damit kann man nun Analoga von einigen Grundkonzepten der Analysis definieren:

- Eine Folge $(v_n)_{n\in\mathbb{N}}$ mit $v_n \in V$ *konvergiert* gegen ein Element $v \in V$, wenn es zu jedem $\varepsilon > 0$ einen Index $N \in \mathbb{N}$ gibt, sodass $d(v_n,v) < \varepsilon$ für alle $n \ge N$ gilt.
- Eine Folge $(v_n)_{n\in\mathbb{N}}$ mit $v_n \in V$ heißt *Cauchy–Folge*, wenn es zu jedem $\varepsilon > 0$ einen Index $N \in \mathbb{N}$ gibt, sodass $d(v_n,v_m) < \varepsilon$ für alle $m,n \ge N$ gilt.
- Ein normierter Raum $(V,\| \ \|)$ heißt *vollständig*, wenn jede Cauchy–Folge in V konvergiert.
- Eine Funktion $f : V \to W$ heißt stetig ein einem Punkt $v_0 \in V$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, sodass aus $d(v_0,v) < \varepsilon$ immer $d(f(v_0),f(v)) < \varepsilon$ folgt (wobei die erste Distanz in V und die zweite in W ist).

8.2. Äquivalenz von Normen. Als nächstes untersuchen wir, ob die Begriffe von Konvergenz und Stetigkeit tatsächlich von der Norm abhängen. Die "richtige" Antwort auf diese Frage wird in der Vorlesung über Topologie gegeben, aber einiges können wir hier schon leicht sehen.

Sei V ein Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$. Zwei Normen $\| \ \|$ und $\| \ \|'$ auf V heißen *äquivalent*, wenn es Konstante $a, b \in \mathbb{R}$ mit $a, b > 0$ gibt, sodass $a\|v\| \leq \|v\|' \leq b\|v\|$ für alle $v \in V$ gilt. Ist das der Fall, so schreibt man $\| \ \| \sim \| \ \|'$.

Bemerkung 8.2. (1) Obwohl die Bedingung in der Definition der Äquivalenz von Normen auf den ersten Blick etwas unsymmetrisch aussieht, ist Äquivalenz von Normen eine Äquivalenzrelation. Jede Norm ist natürlich äquivalent zu sich selbst (mit $a = b = 1$). Ist $\| \ \| \sim \| \ \|'$, dann ist $\| \ \|' \sim \| \ \|$ (mit Konstanten $1/b$ und $1/a$) und auch die Transitivität ist leicht zu verifizieren.
(2) Der Begriff der Äquivalenz von Normen ist leicht geometrisch zu verstehen, wenn man für eine Norm $\| \ \|$ und $r \in \mathbb{R}_{>0}$ die *Kugeln* $B_r(0) := \{v \in V : \|v\| \leq r\}$ betrachtet. Nach Eigenschaft (N2) gilt klarerweise $B_r(0) = \{rv : v \in B_1(0)\}$. Man erhält also die a–Kugel durch "Aufblasen" bzw. "Einschrumpfen" der Einheitskugel $B_1(0)$. Äquivalenz von $\| \ \|$ und $\| \ \|'$ bedeutet in diesem Bild gerade, dass man die Einheitskugel von $\| \ \|$ so einschrumpfen kann, dass sie ganz in der Einheitskugel von $\| \ \|'$ liegt, und so aufblasen kann, dass sie die ganze Einheitskugel von $\| \ \|'$ enthält.

Satz 8.2. *Sei V ein Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und seien $\| \ \|$ und $\| \ \|'$ äquivalente Normen auf V. Dann haben $\| \ \|$ und $\| \ \|'$ die selben konvergenten Folgen, die selben Cauchy–Folgen und die selben stetigen Funktionen in beliebige normierte Räume. Insbesondere ist $(V, \| \ \|)$ genau dann vollständig, wenn $(V, \| \ \|')$ vollständig ist.*

Beweis. Wegen der Äquivalenz der Normen finden wir Konstante $a, b \in \mathbb{R}_{>0}$, sodass $a\|v\| \leq \|v\|' \leq b\|v\|$ für alle $v \in V$ gilt. Sei $(v_n)_{n\in\mathbb{N}}$ eine Folge in V. Konvergiert diese Folge bezüglich $\| \ \|$ gegen $v_0 \in V$ und ist $\varepsilon > 0$, dann finden wir zu ε/b einen Index $N \in \mathbb{N}$, sodass $\|v_n - v_0\| < \varepsilon/b$ für alle $n \geq N$ gilt. Für diese n ist dann aber $\|v_n - v_0\|' \leq b\|v_n - v_0\| < \varepsilon$, also konvergiert $(v_n)_{n\in\mathbb{N}}$ auch bezüglich $\| \ \|'$ gegen v_0. Analog sieht man, daß eine Folge, die Cauchy bezüglich $\| \ \|$ ist auch Cauchy bezüglich $\| \ \|'$ ist. Wegen der Symmetrie der Äquivalenz von Normen folgen auch die umgekehrten Implikationen und damit die Behauptungen über Folgen und Vollständigkeit. Auch die Behauptung über stetige Funktionen zeigt man ganz analog. □

Beispiel 8.2. (1) Betrachten wir $\mathbb{R}^n$ und die Normen $\| \ \|_1$, $\| \ \|_2$ und $\| \ \|_\infty$ aus den Beispielen (1) und (2) von 8.1. Für $x_1, \dots, x_n \in \mathbb{R}$ gilt $\sum |x_i|^2 \leq \big(\sum |x_i|\big)^2$ und zieht man auf beiden Seiten die Wurzel, dann erhält man $\|x\|_2 \leq \|x\|_1$ für alle $x \in \mathbb{R}^n$. Andererseits ist offensichtlich $|x_i| = \sqrt{|x_i|^2} \leq \sqrt{\sum |x_i|^2}$ für jedes i und $\sum |x_i| \leq n \cdot \max\{|x_1|, \dots, |x_n|\}$. Damit erhalten wir aber $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty$ für alle $x \in \mathbb{R}^n$, und die drei Normen sind äquivalent.

(2) Wie wir in Kürze sehen werden, sind auf endlichdimensionalen Vektorräumen alle Normen äquivalent. Um ein interessantes Beispiel von nicht äquivalenten Normen zu finden, müssen wir also unendlichdimensionale Räume betrachten. Betrachten wir etwa den Raum $\mathbb{R}^{(\mathbb{N})}$ aller endlichen Folgen, und darauf die Normen $\|(x_n)\|_1 := \sum_n |x_n|$ und $\|(x_n)\|_\infty := \max\{|x_n|\}$, die natürlich für endliche Folgen beide wohldefiniert sind. Für $k \in \mathbb{N}$ sei $a_k \in \mathbb{R}^{(\mathbb{N})}$ definiert durch $a_k = e_1 + \cdots + e_k$, also als die Folge, deren erste k Glieder Eins sind, während alle weiteren Glieder Null sind. Dann ist aber $\|a_k\|_\infty = 1$ und $\|a_k\|_1 = k$, also kann es keine Konstante C geben, sodass $\|(x_n)\|_1 \leq C\|(x_n)\|_\infty$ für alle Folgen $(x_n) \in \mathbb{R}^{(\mathbb{N})}$ gilt.

Analog sieht man leicht, dass die Normen $\|f\|_\infty = \sup_{x\in[a,b]} |f(x)$ und $\|f\|_1 := \int_a^b |f(x)|dx$ auf $C([a, b], \mathbb{R})$ nicht äquivalent sein können, weil man ganz leicht nichtnegative Funktionen mit Supremum 1 und beliebig kleinem Integral konstruieren kann.

8.3. Der Fall endlicher Dimension. Bevor wir zeigen können, dass auf einem endlichdimensionalen Raum alle Normen äquivalent sind, benötigen wir noch zwei kleine Schritte. Einerseits müssen wir die sogenannte *umgekehrte Dreiecksungleichung* beweisen. Sei dazu $(V, \| \ \|)$ ein normierter Vektorraum und seien $v, w \in V$. Dann ist $v = (v - w) + w$, also gilt nach der Dreiecksungleichung $\|v\| \leq \|v - w\| + \|w\|$. Bringen wir w auf die andere Seite, dann erhalten wir $\|v\| - \|w\| \leq \|v - w\|$ und wegen der Symmetrie der rechten Seite können wir auch links v und w vertauschen und daraus die umgekehrte Dreiecksungleichung $|\|v\| - \|w\|| \leq \|v - w\|$ folgern.

Andererseits benötigen wir noch einen Begriff: Sind $(V, \| \ \|_V)$ und $(W, \| \ \|_W)$ normierte Vektorräume über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, dann heißt eine lineare Abbildung $f : V \to W$ *beschränkt*, wenn es eine Konstante $C \in \mathbb{R}$ gibt, sodass $\|f(v)\|_W \leq C\|v\|_V$ für alle $v \in V$ gilt. Man kann leicht zeigen (siehe Übungen) daß eine lineare Abbildung $f : V \to W$ genau dann beschränkt ist, wenn sie stetig ist.

Satz 8.3. *Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und V ein endlichdimensionaler $\mathbb{K}$–Vektorraum. Dann gilt:*

(1) Je zwei Normen auf V sind äquivalent.

(2) Ist W ein beliebiger normierter Raum über $\mathbb{K}$ und $f : V \to W$ linear, dann ist f bezüglich jeder Norm auf V beschränkt.

Insbesondere ist jeder endlichdimensionale normierte Raum über $\mathbb{R}$ und $\mathbb{C}$ vollständig.

Beweis. Durch Anwenden eines linearen Isomorphismus dürfen wir annehmen, dass $V = \mathbb{K}^n$ für ein $n \in \mathbb{N}$ gilt. Jede Norm auf einem komplexen Vektorraum ist natürlich auch eine Norm auf dem unterliegenden reellen Vektorraum und Äquivalenz von Normen und Beschränktheit linearer Abbildungen bedeuten im reellen und im komplexen Bild das gleiche, also genügt es, den Fall $\mathbb{K} = \mathbb{R}$ zu betrachten.

(1) Es genügt zu zeigen, dass eine beliebige gegebene Norm $\| \ \|$ auf $\mathbb{R}^n$ äquivalent zur euklidischen Norm $\| \ \|_2$ ist. Ist nun $\{e_1, \dots, e_n\}$ die Standardbasis von $\mathbb{R}^n$ und ist $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, dann ist $x = \sum x_i e_i$ und aus der Dreiecksungleichung erhalten wir $\|x\| \leq \sum \|x_i e_i\| = \sum |x_i|\|e_i\|$. Insbesondere ist also $\|x\| \leq \|x\|_1 \max\{\|e_1\|, \dots, \|e_n\|\}$. Aus Beispiel (1) von 8.2 wissen wir, dass $\| \ \|_1$ und $\| \ \|_2$ äquivalent sind, also gibt es eine Konstante C, sodass $\|x\| \leq C\|x\|_2$ gilt. Die umgekehrte Dreiecksungleichung für $\| \ \|$ liefert für $x, y \in \mathbb{R}^n$ nun $|\|y\| - \|x\|| \leq \|y - x\| \leq C\|y - x\|_2$.

Diese Gleichung besagt aber gerade, dass $\| \ \| : \mathbb{R}^n \to \mathbb{R}$ eine (sogar gleichmäßig) stetige Funktion ist. Betrachten wir insbesondere die Einheitssphäre $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$, dann ist das eine beschränkte abgeschlossene Teilmenge von $\mathbb{R}^n$. Nach dem Satz von Heine–Borel ist diese Teilmenge kompakt, also nimmt die stetige Funktion $\| \ \|$ auf ihr ein Minimum und ein Maximum an. Wegen (N1) muss das Minimum strikt positiv sein, also gibt es Konstante $0 < a, b \in \mathbb{R}$, sodaß $a \leq \|x\| \leq b$ für alle x mit $\|x\|_2 = 1$ gilt. Ist nun $x \neq 0$ beliebig, dann gelten für $y = \frac{1}{\|x\|_2} x$ die Gleichungen $\|y\|_2 = 1$ und $\|x\| = \|x\|_2 \|y\|$, und daraus folgt sofort $a\|x\|_2 \leq \|x\| \leq b\|x\|_2$.

(2) Sei $\| \ \|$ eine Norm auf $V = \mathbb{R}^n$ und $\| \ \|_W$ die Norm auf W. Für $x = (x_1, \dots, x_n)$ ist $f(x) = \sum x_i f(e_i)$ und damit erhält man sofort

$$\|f(x)\|_W \leq \sum |x_i| \|f(e_i)\|_W \leq \|x\|_1 \max\{\|f(e_1)\|_W, \dots, \|f(e_n)\|_W\}.$$

Nach Teil (1) ist $\| \ \|$ äquivalent zu $\| \ \|_1$, also folgt sofort die Existenz einer Konstanten C, für die $\|f(x)\|_W \leq C\|x\|$ gilt.

Der letzte Teil des Satzes folgt aus Teil (1), Satz 8.2 und der aus der Analysis bekannten Vollständigkeit von $(\mathbb{R}^n, \| \ \|_2)$. □

8.4. Die Operatornorm. Zum Abschluss erwähnen wir noch kurz einen wichtigen Spezialfall einer Norm. Sei $(V, \| \ \|)$ ein endlichdimensionaler normierter Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, und $f : V \to V$ eine lineare Abbildung. Dann ist f nach Satz 8.3 automatisch beschränkt, also gibt es eine Konstante $C \in \mathbb{R}$, sodass $\|f(v)\| \leq C\|v\|$ für alle $v \in V$ gilt. Die kleinste Konstante, für die das gilt, heißt die *Operatornorm* von f und wird mit $\|f\|$ bezeichnet, man definiert also $\|f\| := \sup_{v \in V \setminus \{0\}} \frac{\|f(v)\|}{\|v\|}$.

Aus dieser Definition folgt sofort, dass $\|f(v)\| \leq \|f\|\|v\|$ für alle $v \in V$ gilt. Das liefert leicht die wichtigste Eigenschaft der Operatornorm. Ist nämlich $g : V \to V$ noch eine lineare Abbildung, dann ist auch $g \circ f : V \to V$ linear und nach Definition gilt für jedes Element $v \in V$ die Ungleichung

$$\|(g \circ f)(v)\| = \|g(f(v))\| \leq \|g\|\|f(v)\| \leq \|g\|\|f\|\|v\|.$$

Damit folgt aber aus der Definition der Operatornorm sofort, dass $\|g \circ f\| \leq \|g\|\|f\|$.

Eine Idee, die später noch eine wichtige Rolle spielen wird, ist lineare Abblidungen oder Matrizen in Polynome einzusetzen. Sei $p = \sum_i a_i x^i \in \mathbb{K}[x]$ ein Polynom, $f : V \to V$ eine lineare Abbildung auf einem $\mathbb{K}$–Vektorraum. Dann können wir $f^2 := f \circ f$, $f^3 = f \circ f \circ f$ und analog f^k für alle $k \in \mathbb{N}$ bilden. Naheliegenderweise setzt man dann noch $f^1 := f$ und $f^0 := \mathrm{id}_V$. Dann können wir aber einfach $p(f) := \sum a_i f^i \in L(V, V)$ definieren. Analog definieren wir für eine quadratische Matrix $A \in M_n(\mathbb{K})$ die Potenzen durch $A^0 := \mathbb{I}_n$, $A^1 = A$, $A^2 = AA$ und induktiv $A^k = AA^{k-1}$ durch die Matrizenmultiplikation. Dann können wir $p(A) := \sum a_i A^i \in M_n(\mathbb{K})$ definieren. Wir werden diese Konstruktionen später noch ausgiebiger besprechen. Hier bemerken wir nur noch, dass die Anwendung von Polynomen auf lineare Abblidungen und auf Matrizen einander entsprechen. Sei $\dim(V) = n$ und $\mathcal{B}$ eine Basis für V, sodass $A = [f]_{\mathcal{B}}$, also die Matrixdarstellung von f bezüglich der Basis $\mathcal{B}$ ist. Dann ist nach Satz 4.16 $[f \circ f]_{\mathcal{B}} = A \cdot A$, also $[f^2]_{\mathcal{B}} = A^2$. Induktiv folgt sofort, dass $[f^k]_{\mathcal{B}} = A^k$ für alle $k \in \mathbb{N}$ gilt und weil die Abbildung $f \mapsto [f]_{\mathcal{B}}$ linear ist, erhalten wir $[p(f)]_{\mathcal{B}} = p(A)$.

Mit Hilfe der Operatornorm kann man diese Idee über Polynome hinaus auf gewisse Potenzreihen verallgemeinern. Von oben sehen wir nämlich, dass für $f : V \to V$ und $f^2 = f \circ f$ sofort $\|f\|^2 \leq \|f\|^2$, und daraus induktiv $\|f^k\| \leq \|f\|^k$ für alle $k \in \mathbb{N}$ folgt. Wenden wir das auf $(V, \| \ \|) = (\mathbb{R}^n, \| \ \|_2)$ an, dann erhalten wir eine Norm auf $M_n(\mathbb{R})$, die $\|A^k\| \leq \|A\|^k$ für alle $k \in \mathbb{N}$ erfüllt.

Ein wichtiger Spezialfall einer konvergenten Potenzreihe ist die Exponentialfunktion. Wie aus der Analysis bekannt, konvergiert die Reihe $\sum_{k=0}^{\infty} \frac{1}{k!} x^k$ für jedes $x \in \mathbb{R}$ gegen $e^x \in \mathbb{R}$. Betrachtet man nun eine Matrix $A \in M_n(\mathbb{R})$, dann kann man für jeden Wert $N \in \mathbb{N}$ natürlich $B_N := \sum_{k=0}^{N} \frac{1}{k!} A^k \in M_n(\mathbb{R})$ bilden, und dann die Folge $(B_N)_{N \in \mathbb{N}}$ betrachten. Für $\varepsilon > 0$ folgt nun aus der Konvergenz der reellen Exponentialreihe für $x = \|A\| \in \mathbb{R}$, dass es einen Index $M \in \mathbb{N}$ gibt, sodass $\sum_{k=\ell}^{m} \frac{1}{k!} \|A\|^k < \varepsilon$ für alle $m \geq \ell \geq M$ gilt. Dann gilt aber auch

$$\|B_m - B_\ell\| = \|\sum_{k=\ell}^{m} \tfrac{1}{k!} A^k\| \leq \sum_{k=\ell}^{m} \tfrac{1}{k!} \|A^k\| \leq \sum_{k=\ell}^{m} \tfrac{1}{k!} \|A\|^k < \varepsilon.$$

Das bedeutet aber gerade, dass $(B_N)_{N \in \mathbb{N}}$ ein Cauchy–Folge in $M_n(\mathbb{R})$ (bezüglich der Operatornorm) ist. Damit konvergiert diese Folge in $M_n(\mathbb{R})$ gegen eine Matrix, die man sinnvollerweise mit e^A bezeichnet. Aus unseren Überlegungen folgt sofort, dass $\|e^A\| \leq e^{\|A\|}$ gilt. Somit kann man die Exponentialfunktion allgemein für Matrizen (und ganz analog für lineare Abbildungen auf beliebigen endlichdimensionalen Vektorräumen) definieren.

Analog zu oben kann man aus der Stetigkeit der reellen Exponentialfunktion leicht schließen, dass das Matrizenexponential stetig als Funktion $M_n(\mathbb{R}) \to M_n(\mathbb{R})$ ist. Auch andere schöne Eigenschaften der reellen Exponentialfunktion lassen sich leicht auf das Matrizenexponential übertragen. Interessant sind hier insbesondere Exponentialkurven $t \mapsto e^{tA}$ für $t \in \mathbb{R}$ und eine fixe Matrix $A \in M_n(\mathbb{R})$, die man als Funktion $\mathbb{R} \to M_n(\mathbb{R})$ betrachten kann. Da $M_n(\mathbb{R})$ ja mit $\mathbb{R}^{n^2}$ identifiziert werden kann, können die aus der Analysis bekannten Konzepten von Differenzierbarkeit problemlos auf diese Kurven angewandt werden. Man zeigt, dass für jede Matrix A die Kurve $c(t) := e^{tA}$ differenzierbar ist, und für die Ableitung gilt $c'(t) = A \cdot c(t)$, wobei der Punkt das Matrizenprodukt bezeichnet. Daraus folgt dann sofort, dass die Kurve $c(t)$ beliebig oft differenzierbar ist. Aus der Definition folgt sofort, dass $c(0) = \mathbb{I}_n$ gilt.

Betrachten wir nun einen Vektor $v \in \mathbb{R}^n$, dann können wir natürlich für jedes $t \in \mathbb{R}$ die Matrix e^{tA} auf den Vektor v anwenden. Bezeichnen wir den resultierenden Vektor mit $v(t)$, dann definiert das eine Kurve $t \mapsto v(t)$ in $\mathbb{R}^n$. Man schließt sofort, dass diese Kurve stetig und sogar beliebig oft differenzierbar ist, wobei die Ableitung durch $v'(t) = A(v(t))$ gegeben ist. Außerdem gilt natürlich $v(0) = \mathbb{I}_n v = v$. Das bedeutet aber gerade, dass die Kurve $v(t) = e^{tA}(v)$ eine Lösung der Differentialgleichung $v'(t) = A(v(t))$ mit Anfangsbedingung $v(0) = v$ ist. Aus grundlegenden Resultaten der Analysis folgt, dass eine Lösung dieser Gleichung durch ihren Wert in einem Punkt eindeutig bestimmt ist. Damit liefert das Matrizenexponential alle Lösungen von linearen gewöhnlichen Differentialgleichungen in $\mathbb{R}^n$ mit konstanten Koeffizienten.

Innere Produkte

Eine Norm auf einem Vektorraum liefert einen Längenbegriff für Vektoren. Spezielle Normen, etwa die euklidische Norm $\| \|_2$ auf $\mathbb{R}^n$, erlauben es, darüberhinaus auch einen Winkel zwischen zwei Vektoren zu definieren. Diese speziellen Normen kommen von inneren Produkten, die dann eine Vielzahl an zusätzlichen Strukturen auf einem Vektorraum liefern.

8.5. Grundlegende Definitionen. Sei V ein reeller Vektorraum. Eine *Bilinearform* auf V ist eine Abbildung $b : V \times V \to \mathbb{R}$, $(v, w) \mapsto b(v, w)$, sodass $b(v + \lambda v', w) = b(v, w) + \lambda b(v', w)$ und $b(v, w + \lambda w') = b(v, w) + \lambda b(v, w')$ für alle $v, v', w, w' \in V$ und $\lambda \in \mathbb{K}$ gelten. Diese Bedingungen sagen gerade, dass für fixes $v \in V$ die Funktionen $w \mapsto b(v, w)$ und $w \mapsto b(w, v)$ lineare Abbildungen $V \to \mathbb{R}$ sind.

Sei $b : V \times V \to \mathbb{R}$ eine Bilinearform.

(1) b heißt *symmetrisch* falls $b(w, v) = b(v, w)$ für alle $v, w \in V$ gilt
(2) Ist b symmetrisch, dann heißt b *positiv semidefinit* falls $b(v, v) \geq 0$ für alle $v \in V$ gilt, und *positiv definit*, falls zusätzlich $b(v, v) = 0$ nur für $v = 0$ gilt.
(3) Ein *inneres Produkt* ist eine positiv definite symmetrische Bilinearform.
(4) Ein *euklidischer Vektorraum* ist ein reeller Vektorraum V zusammen mit einem inneren Produkt auf V.

Bemerkung 8.5. (1) Ist $b : V \times V \to \mathbb{R}$ eine Bilinearform, dann folgt aus der Linearität in der ersten Variable sofort, dass $b(0, w) = 0$ für alle $w \in V$ gilt. Ist $0 \in V$ der einzige Vektor, für den das vorkommt, dann nennt man die Form b *nicht ausgeartet* oder *nicht degeneriert*. Innere Produkte sind natürlich nicht ausgeartet, weil in diesem Fall schon $b(v, v) = 0$ nur für $v = 0$ möglich ist.

Sei $b : V \times V \to \mathbb{R}$ nicht ausgeartet und seien $v, v' \in V$ zwei Vektoren, sodass $b(v, w) = b(v', w)$ für alle $w \in V$ gilt. Dann folgt aus der Linearität in der ersten

Variable sofort, dass $0 = b(v - v', w)$ für alle $w \in V$ gilt, was nach Voraussetzung $v - v' = 0$, also $v = v'$ impliziert.

(2) Innere Produkte werden üblicherweise mit $(v, w) \mapsto \langle v, w\rangle$ bezeichnet.

Beispiel 8.5. (1) Das wichtigst Beispiel eines inneren Produktes ist das *Standard innere Produkt* auf $\mathbb{R}^n$, das durch

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n)\rangle := x_1y_1 + \cdots + x_ny_n.$$

definiert ist. Man sieht sofort, dass dieser Ausdruck bilinear und symmetrisch ist. Außerdem ist $\langle x, x\rangle = x_1^2 + \cdots + x_n^2$, also folgt sofort, dass $\langle\ ,\ \rangle$ tatsächlich ein inneres Produkt definiert.

(2) Vor allem wegen ihrer Bedeutung in der Physik erwähnen wir kurz eine interessante Klasse von Beispielen von nicht degenerierten symmetrischen Bilinearformen, die nicht positiv definit sind. Betrachten wir wieder $V = \mathbb{R}^n$ und sei $0 < p < n$. Dann definieren wir

$$b_p((x_1, \dots, x_n), (y_1, \dots, y_n)) := x_1y_1 + \cdots + x_py_p - x_{p+1}y_{p+1} - \cdots - x_ny_n.$$

Das ist offensichtlich eine symmetrische Bilinearform. Wegen $b_p(e_n, e_n) = -1$ ist diese Bilinearform aber nicht positiv semidefinit. Klarerweise ist aber $b_p(x, e_i)$ für $i \leq p$ gleich x_i und für $i > p$ gleich $-x_i$. Ist also $x \in \mathbb{R}^n$ so, dass $b_p(x, e_i) = 0$ für alle $i = 1, \dots, n$ gilt, dann ist $x = 0$. Damit ist b_p nicht degeneriert.

In der Physik ist der Fall $n = 4$, $p = 1$ besonders wichtig. Man bezeichnet $(\mathbb{R}^4, b_1)$ als *Minkovski–Raum*, wobei die 4 Koordinaten mit ct sowie x_1, x_2 und x_3 bezeichnet werden (Lichtgeschwindigkeit mal Zeit und Ort). Das Analogon von Euklidischer Geometrie mit diesem unterliegenden Vektorraum und dieser unterliegenden Bilinearform liefert die mathematische Formulierung der speziellen Relativitätstheorie.

(3) Ebenfalls wegen seine fundamentalen Bedeutung in der Physik besprechen wir das folgende Beispiel einer schiefsymmetrischen, nicht degenerierten Bilinearform. Sei $V = \mathbb{R}^{2n}$, bezeichnen wir Punkte in V mit (q, p) für $q, p \in \mathbb{R}^n$ ("Phasenraum" mit "Orts–" und "Impulskoordinaten"). Dann setzt man

$$\omega((q, p), (\tilde{q}, \tilde{p})) = \tilde{p}_1q_1 + \dots \tilde{p}_nq_n - p_1\tilde{q}_1 - \cdots - p_n\tilde{q}_n.$$

Dann ist ω offensichtlich eine Bilinearform und $\omega((\tilde{q}, \tilde{p}), (q, p)) = -\omega((q, p), (\tilde{q}, \tilde{p}))$, also ist ω "schiefsymmetrisch". Für $i = 1, \dots, n$ gilt nach Definition $\omega((q, p), (e_i, 0)) = q_i$ und $\omega((q, p), (0, e_i)) = -p_i$, also ist ω nicht degeneriert. Man bezeichnet (R^{2n}, ω) als den *Standard symplektischen Vektorraum* der Dimension $2n$.

Die "symplektische Form" ω ist ein zentrales Element in der mathematischen Beschreibung der klassischen Mechanik, wobei der Zusammenhang wie folgt aussieht: Angenommen ein Punktteilchen der Masse m bewegt sich auf einer Bahn in $\mathbb{R}^3$, die wir als $q(t) = (q_1(t), q_2(t), q_3(t))$ schreiben. Weiters betrachten wir ein Kraftfeld, das durch ein Potential beschrieben werden kann. Das bedeutet, dass es eine stetig differenzierbare Funktion $V : \mathbb{R}^3 \to \mathbb{R}$ gibt, sodass die im Punkt $q = (q_1, q_2, q_3)$ wirkende Kraft durch Minus den Gradienten von V, also $(-\frac{\partial V}{\partial q_1}, -\frac{\partial V}{\partial q_2}, -\frac{\partial V}{\partial q_3})$ gegeben ist. (Das Vorzeichen kommt daher, dass sich das Teilchen von selbst so bewegt, dass die potentielle Energie sinkt.)

Dann definiert man die Impulse $p_j(t) := mq_j'(t)$ (Masse mal Geschwindigkeit) und betrachtet sie als zusätzliche Koordinaten. Die Gesamtenergie (kinetische plus potentielle Energie) des Teilchens ist dann gegeben durch die Funktion $E : \mathbb{R}^6 \to \mathbb{R}$ die definiert ist durch $E = \frac{\sum p_j^2}{2m} + V(q)$. Das dritte Newton'sche Axiom (Kraft ist Masse mal Beschleunigung) liefert dann $p_j'(t) = -\frac{\partial V}{\partial q_j}(q(t))$. Das bedeutet aber, dass die Kurve $u(t) := (q(t), p(t))$ in $\mathbb{R}^6$ bestimmt ist durch $q_j'(t) = \frac{p_j(t)}{m} = \frac{\partial E}{\partial p_j}(u(t))$ und $p_j'(t) = -\frac{\partial E}{\partial q_j}(u(t))$ ("Hamilton'sche Gleichungen"). Diese Gleichungen sind aber äquivalent zu $DE(u(t))(v) = \omega(u'(t), v)$, wobei $DE : \mathbb{R}^6 \times \mathbb{R}^6 \to \mathbb{R}$ die Ableitung der Energiefunktion ist.

(4) Ähnlich wie bei Normen sind auch innere Produkte auf unendlichdimensionalen Räumen sehr interessant. Wir erwähnen hier nur ein Beispiel, nämlich den Raum $C([a,b], \mathbb{R})$ aller stetigen, reellwertigen Funktionen auf einem abgeschlossenen Intervall $[a,b] \subset \mathbb{R}$. Aus den aus der Analysis bekannten Eigenschaften des Integrals kann man leicht schließen, dass

$$\langle f, g \rangle := \int_a^b f(t)g(t)dt$$

ein inneres Produkt auf diesem Raum definiert.

8.6. Sesquilinearformen. Natürlich kann ganz analog zu Bilinearformen $V \times V \to \mathbb{R}$ auf einem reellen Vektorraum V auch $\mathbb{C}$–wertige komplexe Bilinearformen auf komplexen Vektorräumen betrachten, und alle Begriffe aus 8.5 machen auch über $\mathbb{C}$ Sinn. Es gibt allerdings ein gravierendes Problem, nämlich dass komplexe Bilinearformen niemals positiv definit sein können. Für $v \in V$ und $b : V \times V \to \mathbb{C}$ komplex bilinear gilt nämlich $b(iv, iv) = i^2 b(v,v) = -b(v,v)$. Daher benötigen wir einen etwas anderen Begriff um ein gutes Analogon reeller innerer Produkte zu erhalten, nämlich den Begriff der *Sesquilinearform*.

Definition 8.6. Sei V ein Vektorraum über $\mathbb{C}$. Eine Sesquilinearform auf V ist eine Funktion $s : V \times V \to \mathbb{C}$, die $s(v + \lambda v', w) = s(v,w) + \lambda s(v', w)$, sowie $s(w,v) = \overline{s(v,w)}$ für alle $v, v', w \in V$ und alle $\lambda \in \mathbb{C}$ erfüllt.

Aus den beiden Bedingungen in der Definition folgt sofort, dass $s(v, w + \lambda w') = s(v,w) + \bar{\lambda} s(v, w')$ gilt. Man sagt, eine Sesquilinearform ist *konjugiert linear* in der zweiten Variable. Andererseits gilt natürlich $s(v,v) = \overline{s(v,v)}$, also $s(v,v) \in \mathbb{R} \subset \mathbb{C}$ für alle $v \in V$. Man sagt nun eine Sesquilinearform s ist *positiv semidefinit*, wenn $s(v,v) \geq 0$ für alle $v \in V$ gilt und s heißt *positiv definit*, wenn zusätzlich $s(v,v) = 0$ nur für $v = 0$ gilt. Der Begriff "nicht degeneriert" macht natürlich für Sesquilinearformen in genau der selben Form Sinn, wie für Bilinearformen, und wieder sind positiv definite Sesquilinearformen automatisch nicht degeneriert.

Positiv definite Sesquilinearformen werden *hermitische innere Produkte* genannt und meist mit $\langle\ ,\ \rangle$ bezeichnet. Ein *unitärer Vektorraum* ist ein komplexer Vektorraum V zusammen mit einem hermitischen inneren Produkt auf V.

Formal kann man symmetrische Bilinearformen und Sesquilinearformen gleichzeitig behandeln, indem man einfach immer mit Sesquilinearformen arbeitet und im reellen Fall die Konjugation als Identität definiert.

Beispiel 8.6. Betrachte $V = \mathbb{C}^n$, und definiere $\langle\ ,\ \rangle : V \times V \to \mathbb{C}$ durch $\langle z, w \rangle = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n$. Offensichtlich ist das eine Sesquilinearform und wegen $\langle z, z \rangle = \sum z_j \bar{z}_j = \sum |z_j|^2$ ist diese Form positiv definit. Das ist das *Standard hermitische innere Produkt* auf $\mathbb{C}^n$.

8.7. Vom inneren Produkt zur Norm. Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum. Dann ist $\langle v, v\rangle \geq 0$ für jedes $v \in V$, also können wir $\|v\| := \sqrt{\langle v, v\rangle}$ bilden. Wir wollen zeigen, dass dies eine Norm auf V definiert. Dazu brauchen wir ein Resultat, das noch öfters sehr nützlich sein wird, nämlich die sogenannte *Cauchy–Schwarz–Ungleichung*.

Lemma 8.7 (Cauchy–Schwarz–Ungleichung). *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum. Dann ist $|\langle v, w\rangle|^2 \leq \langle v, v\rangle\langle w, w\rangle$. Gleichheit gilt genau dann, wenn v und w linear abhängig sind.*

Beweis. Für $w = 0$ sind offensichtlich beide Seiten gleich Null, also nehmen wir $w \neq 0$ an. Für jedes $\lambda \in \mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$ gilt

$$0 \leq \langle v - \lambda w, v - \lambda w\rangle = \langle v, v\rangle - \lambda\langle w, v\rangle - \bar{\lambda}\langle v, w\rangle + \lambda\bar{\lambda}\langle w, w\rangle.$$

Setzt man $\lambda = \langle v, w\rangle/\langle w, w\rangle$, so erhält man $0 \leq \langle v, v\rangle - \frac{|\langle v,w\rangle|^2}{\langle w,w\rangle}$, und die Ungleichung folgt durch Multiplizieren mit der positiven reellen Zahl $\langle w, w\rangle$. In der obigen Ungleichung gilt außerdem genau dann Gleichheit, wenn $v - \lambda w = 0$ gilt, also v und w linear abhängig sind. □

Satz 8.7. *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum. Dann definiert $\|v\| := \sqrt{\langle v, v\rangle}$ eine Norm auf V. Diese Norm erfüllt die Parallelogrammidentität*

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

Weiters gelten die Polarisierungsformeln:

(1) für $(V, \langle\ ,\ \rangle)$ euklidisch: $\langle v, w\rangle = \frac{1}{4}(\|v + w\|^2 - \|v - w\|^2)$.

(2) für $(V, \langle\ ,\ \rangle)$ unitär: $\langle v, w\rangle = \frac{1}{4}(\|v + w\|^2 - \|v - w\|^2 + i\|v + iw\|^2 - i\|v - iw\|^2)$.

Beweis. Nach Konstruktion ist $\|v\| \geq 0$ und $\|v\| = 0$ impliziert $\langle v, v\rangle = 0$ und damit $v = 0$, also ist (N1) erfüllt. Für $\lambda \in \mathbb{K}$ ist $\langle\lambda v, \lambda v\rangle = \lambda\bar{\lambda}\langle v, v\rangle$, also folgt $\|\lambda v\| = |\lambda| \cdot \|v\|$ wegen $\lambda\bar{\lambda} = |\lambda|^2$. Wiederum wegen der Bilinearität gilt

$$(*) \qquad \langle v \pm w, v \pm w\rangle = \langle v, v\rangle \pm \langle v, w\rangle \pm \langle w, v\rangle + \langle w, w\rangle.$$

Für positives Vorzeichen liefert diese Gleichung $\|v + w\|^2 = \|v\|^2 + 2\,\mathrm{Re}(\langle v, w\rangle) + \|w\|^2$. Natürlich ist $\mathrm{Re}(\langle v, w\rangle) \leq |\langle v, w\rangle|$ und nach der Cauchy–Schwarz Ungleichung ist $|\langle v, w\rangle| \leq \|v\| \cdot \|w\|$, also erhalten wir $\|v + w\|^2 \leq \|v\|^2 + 2\|v\| \cdot \|w\| + \|w\|^2$, und das impliziert natürlich die Dreiecksungleichung für $\|\ \|$. Addiert man die Gleichungen $(*)$ für positives und negatives Vorzeichen, dann erhält man direkt die Parallelogrammidentität.

Die Differenz der beiden Gleichungen $(*)$ für positives und negatives Vorzeichen liefert $\|v + w\|^2 - \|v - w\|^2 = 2\langle v, w\rangle + 2\langle w, v\rangle$. Im reellen Fall liefert das sofort die Polarisierungsformel. Im komplexen Fall muss man nur noch bemerken, dass das auch $\|v+iw\|^2 - \|v-iw\|^2 = -2i\langle v, w\rangle + 2i\langle w, v\rangle$ impliziert. Multipliziert man diese Gleichung mit i und addiert sie zur obigen, dann erhält man die komplexe Polarisierungsformel. □

Definition 8.7. Ein reeller bzw. komplexer *Hilbertraum* ist ein euklidischer bzw. unitärer Vektorraum $(V, \langle\ ,\ \rangle)$, der bezüglich der Norm $\|v\| = \sqrt{\langle v, v\rangle}$ vollständig ist.

Bemerkung 8.7. Ein inneres Produkt liefert also eine Norm auf V und wegen der Polarisierungsformel ist das innere Produkt durch diese Norm vollständig bestimmt. Es zeigt sich, dass die Parallelogrammidentität charakteristisch für Normen ist, die von inneren Produkten kommen. Man kann leicht elementar-geometrisch zeigen, dass diese Identität äquivalent zum Satz von Pythagoras ist. Sie hat aber den Vorteil, dass man den Begriff des rechten Winkels (bzw. der Orthogonalität) in der Formulierung nicht benötigt. Man kann allgemein beweisen, dass für eine Norm, die die Parallelogrammidentität erfüllt, durch die Polarisierungsformel ein inneres Produkt auf V definiert wird, das gerade die gegebene Norm liefert.

Beispiel 8.7. (1) Das Standard innere Produkt $\langle x, y\rangle = \sum x_i y_i$ auf $\mathbb{R}^n$ liefert als zugehörige Norm die euklidische Norm $\|x\|_2 = \sqrt{\sum x_i^2}$ (daher auch der Name "euklidische Norm"). Analog liefert das Standard hermitische innere Produkt $\langle z, w\rangle = \sum z_j \bar{w}_j$ die euklidische Norm auf $\mathbb{C}^n = \mathbb{R}^{2n}$. Wie wir bereits in 8.1 bemerkt haben, sind endlichdimensionale normierte Räume immer vollständig, also erhalten wir hier jeweils einen Hilbertraum.

(2) Für $p \neq 2$ und $n > 1$ kommt keine der Normen $\| \ \|_p$ auf $\mathbb{R}^n$ aus Beispiel (3) von 8.1 von einem inneren Produkt. Betrachten wir nämlich die Standardbasis $\{e_1, \dots, e_n\}$, dann ist $\|e_j\|_p = 1$ für alle j und p. Andererseits ist $\|e_1 \pm e_2\|_p = (1^p + 1^p)^{1/p} = 2^{1/p}$, also $\|e_1 + e_2\|^2 + \|e_1 - e_2\|^2 = 2^{1+2/p}$. Damit die Parallelogrammidentität erfüllt ist, muss das gleich 4, also $2/p = 1$ sein. Für $p = \infty$ ist $\|e_1 \pm e_2\|_\infty = 1$ und damit liefert eine Seite der Parallelogrammidentität zwei, die andere vier.

(3) Das Standardbeispiel eines unendlichdimensionalen Hilbertraumes ist der Raum ℓ^2 aller komplexen Folgen $(a_n)_{n\in\mathbb{N}}$, für die $\sum_{n\in\mathbb{N}} |a_n|^2 < \infty$ gilt. Man kann direkt verifizieren, dass diese Folgen einen Teilraum des Raumes aller komplexwertigen Folgen bilden. Weiters zeigt man, dass für zwei Folgen $(a_n), (b_n) \in \ell^2$ die Summe $\sum_{n\in\mathbb{N}} a_n \bar{b}_n$ immer konvergiert. Damit kann man

$$\langle (a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}\rangle := \textstyle\sum_{n\in\mathbb{N}} a_n \bar{b}_n$$

definieren, und man zeigt, dass das ein inneres Produkt auf ℓ^2 ist. Dieses innere Produkt liefert dann die Norm $\|(a_n)\| = \sqrt{\sum_{n\in\mathbb{N}} |a_n|^2}$ und man verifiziert, dass ℓ^2 mit dieser Norm ein vollständiger Raum ist.

Orthogonalität und Orthonormalbasen

Innere Produkte liefern den Begriff der Orthogonalität (der allgemeiner auch für nicht entartete Bilinear– bzw. Sesquilinearformen Sinn macht), der zum Begriff des Winkels zwischen zwei Vektoren verfeinert werden kann.

8.8. Orthogonalität. Sei $(V, \langle \ , \ \rangle)$ ein euklidischer oder unitärer Vektorraum. Dann nennen wir zwei Elemente $v, w \in V$ *orthogonal* und schreiben $v \perp w$, wenn $\langle v, w\rangle = 0$ gilt. Ist $A \subset V$ eine beliebige Teilmenge, dann definieren wir den *Orthogonalraum* $A^\perp$ von A durch $A^\perp := \{v \in V : \forall a \in A : \langle v, a\rangle = 0\}$. Offensichtlich ist das ein Teilraum von V (auch wenn A selbst kein Teilraum ist).

Ein *Orthogonalsystem* in V ist eine Teilmenge $\{v_i : i \in I\} \subset V$, sodass $v_i \neq 0$ für alle $i \in I$ und $\langle v_i, v_j\rangle = 0$ für $i \neq j$ gilt. Gilt zusätzlich $\langle v_i, v_i\rangle = 1$ (also $\|v_i\| = 1$) für alle $i \in I$, dann spricht man von einem *Orthonormalsystem* in V. Eine *Orthonormalbasis* ist ein Orthonormalsystem in V, das eine Basis von V ist.

Proposition 8.8. *Sei $(V, \langle \ , \ \rangle)$ ein Euklidischer oder unitärer Vektorraum und sei $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Orthonormalbasis für V. Dann gilt:*

(1) Sind $v, w \in V$ mit $v \perp w$, dann ist $\|v + w\|^2 = \|v\|^2 + \|w\|^2$.

(2) Jedes Orthogonalsystem in V ist eine linear unabhängige Teilmenge.

(3) Für Skalare $r_1, \dots, r_n, s_1, \dots, s_n \in \mathbb{K}$ gilt $\langle \sum r_j v_j, \sum s_k v_k\rangle = \sum_{j=1}^n r_j \bar{s}_j$.

(4) Für $v \in V$ ist die eindeutige Darstellung als Linearkombination der Elemente von $\mathcal{B}$ gegeben durch $v = \sum_{j=1}^n \langle v, v_j\rangle v_j$.

(5) Sei $(W, \langle\ ,\ \rangle_W)$ ein weiterer Vektorraum der gleichen Art, $\mathcal{C} = \{w_1, \dots, w_m\}$ eine Orthonormalbasis für W und $f : V \to W$ eine lineare Abbildung. Dann ist die Matrixdarstellung $[f]^{\mathcal{B}}_{\mathcal{C}} = (a_{ij})$ gegeben durch $a_{ij} = \langle f(v_j), w_i\rangle_W$.

Beweis. (1) Wegen $v \perp w$ gilt $\langle v, w\rangle = \langle w, v\rangle = 0$ und damit $\langle v + w, v + w\rangle = \langle v, v\rangle + \langle w, w\rangle$.

(2) Sei $A \subset V$ ein Orthogonalsystem, $a_1, \dots, a_m \in A$ und $\lambda_1, \dots, \lambda_m \in \mathbb{K}$, sodass $\sum \lambda_j a_j = 0$ gilt. Dann gilt für jedes $k = 1, \dots, m$ die Gleichung

$$0 = \langle \textstyle\sum_j \lambda_j a_j, a_k\rangle = \sum_j \lambda_j \langle a_j, a_k\rangle = \lambda_k \langle a_k, a_k\rangle.$$

Nach Definition ist $a_k \neq 0$, also $\langle a_k, a_k\rangle \neq 0$. Damit muss aber $\lambda_k = 0$ für alle $k = 1, \dots, m$ gelten, also ist A linear unabhängig.

(3) Wegen der Sesquilinearität von $\langle\ ,\ \rangle$ ist

$$\langle \textstyle\sum_j x_j v_j, \sum_k y_k v_k\rangle = \sum_j x_j \langle v_j, \sum_k y_k v_k\rangle = \sum_{j,k} x_j \bar{y}_k \langle v_j, v_k\rangle,$$

und nach Definition ist $\langle v_j, v_k\rangle$ gleich Null für $j \neq k$ und gleich Eins für $j = k$.

(4) Da $\{v_1, \dots, v_n\}$ eine Basis für V ist, können wir den Vektor $v \in V$ eindeutig in der Form $v = \sum_{i=1}^n \lambda_i v_i$ schreiben, und nach Teil (3) ist $\langle v, v_i\rangle = \lambda_i$.

(5) Nach Proposition 4.15 ist die j–te Spalte der Matrix $[f]^{\mathcal{B}}_{\mathcal{C}}$ der Koordinatenvektor $[f(v_j)]_{\mathcal{C}}$. Nach Teil (4) ist aber $f(v_j) = \sum_i \langle f(v_j), w_i\rangle_W w_i$ und die Behauptung folgt. □

Ein zentrales Resultat über euklidische und unitäre Vektorräume ist nun, dass linear unabhängige Teilmengen eindeutig Orthonormalsysteme liefern. Der Beweis besteht in einer expliziten Prozedur, dem sogenannten *Gram–Schmidt Orthonormalisierungsverfahren.*

Satz 8.8. *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum und sei $A \subset V$ linear unabhängig mit $A = \{a_1, \dots, a_n\}$. Dann gibt es ein eindeutiges Orthonormalsystem $\{v_1, \dots, v_n\} \subset V$, das folgende Bedingung erfüllt: Für jedes $j = 1, \dots, k$ kann a_j als Linearkombination von $v_1, \dots, v_j$ geschrieben werden, wobei der Koeffizient von v_j reell und positiv ist.*

Beweis. Zunächst beweisen wir die Existenz durch eine induktive Konstruktion. Da A linear unabhängig ist, ist $a_1 \neq 0$, also $\|a_1\| \neq 0$, und wir setzen $v_1 := \frac{1}{\|a_1\|} a_1$. Dann ist $\langle v_1, v_1\rangle = \frac{1}{\|a_1\|^2}\langle a_1, a_1\rangle = 1$, also $\{v_1\}$ ein Orthonormalsystem, und $a_1 = \|a_1\| v_1$.

Nehmen wir induktiv an, dass $k \geq 2$ gilt, und wir $\{v_1, \dots, v_{k-1}\}$ bereits konstruiert haben. Setze nun $\tilde{v}_k := a_k - \sum_{j=1}^{k-1} \langle a_k, v_j\rangle v_j$. Für $\ell < k$ ist dann

$$\langle \tilde{v}_k, v_\ell\rangle = \langle a_k, v_\ell\rangle - \textstyle\sum_{j=1}^{k-1} \langle a_k, v_j\rangle \langle v_j, v_\ell\rangle.$$

Nach Induktionsvoraussetzung ist $\{v_1, \dots, v_{k-1}\}$ ein Orthonormalsystem, also $\langle \tilde{v}_k, v_\ell\rangle = 0$ für alle $\ell < k$.

Wäre $\tilde{v}_k = 0$, dann könnte man a_k als Linearkombination von $v_1, \dots, v_{k-1}$ schreiben. Das diese $k - 1$ Elemente ein Orthogonalsystem bilden, erzeugen sie nach Teil (3) der Proposition einen $k - 1$–dimensionalen Teilraum, in dem a_k liegen würde. Aber ebenfalls nach Induktionsvoraussetzung liegt $\{a_1, \dots, a_{k-1}\}$ in diesem Teilraum und ist linear unabhängig, also eine Basis. Somit könnte man a_k als Linearkombination der a_j für $j < k$ schreiben, ein Widerspruch zur linearen Unabhängigkeit von A.

Somit ist $\tilde{v}_k \neq 0$, also können wir $v_k := \frac{1}{\|\tilde{v}_k\|} \tilde{v}_k$ setzen. Dann gilt natürlich immer noch $\langle v_k, v_j\rangle = 0$ für $j < k$ und außerdem $\langle v_k, v_k\rangle = 1$, also ist $\{v_1, \dots, v_k\}$ ein Orthonormalsystem. Außerdem ist wiederum nach Konstruktion $a_k = \|\tilde{v}_k\| v_k + \sum_{j=1}^{k-1} \langle a_k, v_j\rangle v_j$, also ist die letzte Bedingung verifiziert.

Zur Eindeutigkeit: Es muss $a_1 = av_1$ für $a \in \mathbb{R}$, $a > 0$ und $\|v_1\| = 1$ gelten. Daraus folgt aber $a = \|a_1\|$ und damit $v_1 = \frac{a_1}{\|a_1\|}$. Haben wir gezeigt, dass $v_1, \dots, v_{k-1}$ eindeutig bestimmt sind, dann muss nach Teil (2) der Proposition $a_k = \langle a_k, v_k\rangle v_k + \sum_{j=1}^{k-1} \langle a_k, v_j\rangle v_j$ und damit ist v_k bis auf Vielfache eindeutig bestimmt. Nun soll aber auch noch $\langle a_k, v_k\rangle$ reell und positiv sein, was v_k bis auf einen positiven reellen Faktor bestimmt, der durch $\|v_k\| = 1$ dann eindeutig festgelegt wird. □

KOROLLAR 8.8. *Sei $(V, \langle\ ,\ \rangle)$ ein endlichdimensionaler euklidischer oder unitärer Vektorraum. Dann gilt:*

(1) V besitzt eine Orthonormalbasis.

(2) Jedes Orthonormalsystem in V kann zu einer Orthonormalbasis erweitert werden.

BEWEIS. (1) Nach Korollar 4.4 besitzt V eine Basis $\mathcal{B} = \{a_1, \dots, a_n\}$. Wenden wir darauf das Orthonormalisierungsverfahren aus Satz 8.8 an, dann erhalten wir ein Orthonormalsystem $\{v_1, \dots, v_n\}$ mit $\dim(V)$ vielen Elementen. Nach Teil (2) von Proposition 8.8 ist das eine linear unabhängige Teilmenge in V, also eine Basis nach Korollar 4.5.

(2) Ist A ein Orthonormalsystem in V, dann ist A nach Teil (2) von Proposition 8.8 linear unabhängig, kann also nach Korollar 4.5 zu einer Basis von V erweitert werden. Wendet man auf diese Basis das Orthonormalisierungsverfahren an, dann passiert in den ersten Schritten (bei den Elementen von A) nichts, also erhält man ein Orthonormalsystem mit $\dim(V)$ Elementen, das A enthält, und wie in Teil (1) folgt, dass dieses eine Orthonormalbasis für V ist. □

BEMERKUNG 8.8. (1) Nach dem Korollar kann man in einem endlichdimensionalen euklidischen oder unitären Vektorraum immer mit Orthonormalbasen arbeiten. In einer Orthonormalbasis sieht aber das innere Produkt auf V nach Teil (3) von Proposition 8.8 genau so aus wie das Standard innere Produkt auf $\mathbb{K}^n = \mathbb{R}^n$ bzw. $\mathbb{C}^n$. Man kann das auch so ausdrücken, dass der lineare Isomorphismus $V \to \mathbb{K}^n$, $v \mapsto [v]_{\mathcal{B}}$, der durch eine Orthonormalbasis induziert wird, verträglich mit den inneren Produkten ist. (Wir werden solche lineare Isomorphismen später "orthogonal" bzw. "unitär" nennen.) Man kann also die Untersuchung von euklidischen (unitären) Vektorräumen immer auf $\mathbb{R}^n$ ($\mathbb{C}^n$) mit den Standard inneren Produkten zurückführen. Für unendlichdimensionale Hilberträume gelten ähnliche Resultate, hier gibt es bis auf Isomorphie auch jeweils nur einen Hilbertraum für jede Kardinalzahl.

(2) Unsere Resultate liefern uns auch eine schöne geometrische Interpretation des inneren Produktes in einem euklidischen Vektorraum. Sind nämlich $v, w \in V$ linear unabhängig, dann erzeugen die beiden Vektoren eine Ebene, und wir können (etwa durch Orthonormalisieren von $\{v, w\}$) einen Vektor $\tilde{v} \neq 0$ in dieser Ebene finden der orthogonal auf v steht. Dann können wir aber w eindeutig als $av + b\tilde{v}$ für $a, b \in \mathbb{R}$ schreiben, und av kann als die Projektion von w auf v interpretiert werden. Nach Teil (3) von Proposition 8.8 ist aber $a = \frac{\langle v,w\rangle}{\langle v,v\rangle}$ und damit ist $|\langle v, w\rangle| = |a| \cdot \|v\|^2 = \|av\| \cdot \|v\|$, und wir können den Betrag des inneren Produktes als das Produkt der Länge von v und der Länge der Projektion von w auf v interpretieren.

(3) Mit kleinen Änderungen machen die Konzepte dieses Abschnitts auch allgemeiner für eine nicht entartete symmetrische Bilinearform bzw. Sesquilinearform $b : V \times V \to \mathbb{K}$ Sinn. Man muss dann nur Orthogonalsysteme durch $b(v_i, v_j) = 0$ und $b(v_i, v_i) \neq 0$ definieren und für ein Orthonormalsystem zusätzlich $b(v_i, v_i) = \pm 1$ verlangen. Dann gelten auch die Resultate dieses Abschnitts mit kleinen Modifikationen weiter.

8.9. Orthogonales Komplement und Orthgonalprojektion. Als zeigen wir, dass für einen Teilraum W eines Euklidischen oder unitären Vektorraumes, der Orthogonalraum $W^\perp$ ein Komplement zu W darstellt, siehe 4.9. Deshalb wird $W^\perp$ meist als das *orthogonale Komplement* von W bezeichnet. (Hier ist entscheidend, dass wir mit inneren Produkten arbeiten, für nicht degenerierte symmetrische Bilinearformen gilt das nicht.)

PROPOSITION 8.9. *Sei $(V, \langle\ ,\ \rangle)$ ein Euklidischer oder unitärer Vektorraum und $W \subset V$ ein Teilraum mit Orthogonalraum $W^\perp$. Dann ist $V = W \oplus W^\perp$, also $V = W + W^\perp$ und $W \cap W^\perp = \{0\}$.*

BEWEIS. Nach Korollar 8.8 können wir eine Orthonormalbasis $\{w_1, \dots, w_k\}$ für W wählen und sie dann zu einer Orthonormalbasis für V erweitern. Sind $v_1, \dots, v_\ell$ die hinzugekommenen Elemente, dann gilt nach Konstruktion $\langle v_i, w_j\rangle = 0$ für alle i und j. Da jedes $w \in W$ als Linearkombination der w_j geschrieben werden kann, folgt $\langle v_i, w\rangle = 0$ für alle $w \in W$, also $v_i \in W^\perp$ für $i = 1, \dots, \ell$. Weiters ist $W \cap W^\perp = \{0\}$, weil für $w \in W \cap W^\perp$ insbesondere $\langle w, w\rangle = 0$ gelten muss. Nach der Dimensionsformel für Summen ist $\dim(W^\perp) \leq \dim(V) - \dim(W) = \ell$, also muss Gleichheit gelten. Damit folgt $V = W \oplus W^\perp$ aus Proposition 4.9. □

Eine Zerlegung eines Vektorraumes in eine direkte Summe läßt sich schön durch eine spezielle lineare Abbildung, eine sogenannte *Projektion* beschreiben. Ist nämlich $V = W \oplus \tilde{W}$ für Teilräume $W, \tilde{W} \subset V$, dann kann man nach Proposition 4.9 jedes Element $v \in V$ eindeutig als $v = w + \tilde{w}$ mit $w \in W$ und $\tilde{w} \in \tilde{W}$ schreiben. Damit kann man eine Funktion $\pi : V \to V$ definieren, indem man $\pi(v)$ als die W–Komponente in dieser eindeutigen Darstellung definiert.

LEMMA 8.9. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$.*

Ist $V = W \oplus \tilde{W}$ für Teilräume $W, \tilde{W} \subset V$ und $\pi : V \to V$ die zugehörige Projektion auf die W–Komponente, dann ist π eine lineare Abbildung und es gilt $\pi \circ \pi = \pi$, $\mathrm{Im}(\pi) = W$ und $\mathrm{Ker}(\pi) = \tilde{W}$.

Ist umgekehrt $\varphi : V \to V$ eine lineare Abbildung, die $\varphi \circ \varphi = \varphi$ erfüllt, dann ist $V = \mathrm{Im}(\varphi) \oplus \mathrm{Ker}(\varphi)$ und φ ist genau die Projektion auf die erste Komponente dieser Zerlegung.

BEWEIS. Ist $V = W \oplus \tilde{W}$, dann ist für $v = w + \tilde{w}$ und $r \in \mathbb{K}$ natürlich $rv = rw + r\tilde{w}$ die eindeutige Darstellung als Summe, also folgt $\pi(rv) = rw = r\pi(v)$. Analog sieht man, dass $\pi(v_1 + v_2) = \pi(v_1) + \pi(v_2)$ gilt, also ist π linear. Für $w \in W$ ist natürlich $w = w + 0$ die eindeutige Darstellung als Summe, also gilt $\pi(w) = w$ für alle $w \in W$. Wegen $\pi(v) \in W$ folgt sofort $\pi \circ \pi = \pi$ und $\mathrm{Im}(\pi) = W$. Analog sieht man $\pi(\tilde{w}) = 0$ für alle $\tilde{w} \in \tilde{W}$, also $\tilde{W} \subset \mathrm{Ker}(\pi)$. Wegen $\dim(\tilde{W}) = \dim(V) - \dim(W) = \dim(V) - \dim(\mathrm{Im}(\pi)) = \dim(\mathrm{Ker}(\pi))$ folgt dann $\mathrm{Ker}(\pi) = \tilde{W}$.

Sei umgekehrt $\varphi : V \to V$ linear mit $\varphi \circ \varphi = \varphi$. Dann sind $\mathrm{Ker}(\varphi)$ und $\mathrm{Im}(\varphi)$ Teilräume von V und nach dem Dimensionssatz für lineare Abbildungen (Satz 4.11) ist $\dim(\mathrm{Ker}(\varphi)) + \dim(\mathrm{Im}(\varphi)) = \dim(V)$. Ist $v \in \mathrm{Ker}(\varphi) \cap \mathrm{Im}(\varphi)$, dann ist $\varphi(v) = 0$ und $v = \varphi(\tilde{v})$ für ein Element $\tilde{v} \in V$. Damit ist aber $0 = \varphi(\varphi(\tilde{v})) = \varphi(\tilde{v}) = v$. Somit ist $\mathrm{Ker}(\varphi) \cap \mathrm{Im}(\varphi) = \{0\}$, also $V = \mathrm{Im}(\varphi) \oplus \mathrm{Ker}(\varphi)$ nach Proposition 4.9. Aus $\varphi \circ \varphi = \varphi$ folgt sofort, dass $\varphi(v - \varphi(v)) = 0$ gilt. Damit ist aber $v = \varphi(v) + (v - \varphi(v))$ genau die Zerlegung von v bezüglich $V = \mathrm{Im}(\varphi) \oplus \mathrm{Ker}(\varphi)$. □

Kehren wir nun zum Fall eines Euklidischen oder unitären Vektorraumes $(V, \langle\ ,\ \rangle)$ und eines Teilraumes $W \subset V$ zurück. Die Projektion π auf die erste Kompoenente bezüglich der Zerlegung $V = W \oplus W^\perp$ wir als die *Orthogonalprojektion von V auf W* bezeichnet. Wir können diese nun schön charakterisieren und explizit beschreiben.

Satz 8.9. *Sei $(V, \langle\ ,\ \rangle)$ ein Euklidischer oder unitärer Vektorraum, $W \subset V$ ein Teilraum und $\pi : V \to V$ die zugehörige Orthogonalprojektion.*

(1) Der Punkt $\pi(v) \in W$ ist der eindeutig bestimmte Punkt von W, der minimalen Abstand von v hat.

(2) Ist $\{w_1, \dots, w_k\}$ eine Orthonormalbasis für W, dann ist $\pi(v) = \sum_{i=1}^k \langle v, w_i\rangle w_i$.

Beweis. (1) Nach Definition ist $v - \pi(v) \in W^\perp$. Für einen Punkt $w \in W$ ist $v - w = (v - \pi(v)) + (\pi(v) - w)$, wobei der erste Summand in $W^\perp$ liegt und der zweite in W. Nach dem Satz von Pythagoras (Teil (1) von Proposition 8.8) folgt

$$\|v - w\|^2 = \|v - \pi(v)\|^2 + \|\pi(v) - w\|^2.$$

Damit ist aber $\|v - w\|^2 \geq \|v - \pi(v)\|^2$ und Gleichheit gilt genau für $\|\pi(v) - w\| = 0$, also $w = \pi(v)$.

(2) Erweitern wir wieder durch $v_1, \dots, v_\ell$ zu einer Orthonormalbasis für V, dann bilden die v_i eine Orthonormalbasis für $W^\perp$. Entwickeln wir v bezüglich dieser Basis, dann erhalten wir eine Komponente in W und eine in $W^\perp$. Nach Teil (3) von Proposition 8.8 ist die W–Komponente gerade durch den behaupteten Ausdruck gegeben. □

8.10. Der Winkel. Wie schon angekündigt kann man in euklidischen Vektorräumen den Winkel zwischen zwei Vektoren definieren. Grundlage dafür ist die Cauchy–Schwarz Ungleichung: Für einen euklidischen Raum V und $v, w \in V$ mit $v, w \neq 0$ impliziert diese, dass $-\|v\| \cdot \|w\| \leq \langle v, w\rangle \leq \|v\| \cdot \|w\|$ gilt. Somit ist $\frac{\langle v,w\rangle}{\|v\|\cdot\|w\|}$ eine reelle Zahl im Intervall $[-1, 1]$. Weiters wissen wir, dass diese Zahl nur dann 1 (bzw. -1) sein kann, wenn w ein positives (bzw. negatives) Vielfaches von v ist, sowie dass sie genau dann gleich Null ist, wenn $v \perp w$ gilt. Wie aus der Analysis bekannt ist, gibt es einen eindeutigen Winkel $\alpha \in [0, \pi]$, sodass $\frac{\langle v,w\rangle}{\|v\|\cdot\|w\|} = \cos(\alpha)$ gilt, und wir definieren dieses α als den Winkel zwischen den Vektoren v und w.

Diese Definition entspricht genau der intuitiven Vorstellung des Winkels, und stimmt in $\mathbb{R}^n$ mit dem Standard inneren Produkt mit dem üblichen Begriff des Winkels überein. Insbesondere ist der Winkel zwischen v und w genau dann 0, wenn w ein positives Vielfaches von v ist, er ist genau dann gleich π, wenn w ein negatives Vielfaches von v ist, und er ist genau dann $\frac{\pi}{2}$, wenn $v \perp w$ gilt.

Die Definition des Winkels α zwischen zwei Vektoren $v, w \in V$ mit $v, w \neq 0$ kann man nun auch als $\langle v, w\rangle = \|v\| \cdot \|w\| \cdot \cos(\alpha)$ schreiben. Die Gleichung

$$\langle v + w, v + w\rangle = \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle$$

liefert nun den *Cosinussatz* $\|v+w\|^2 = \|v\|^2 + \|w\|^2 + 2\cos(\alpha)\|v\| \cdot \|w\|$ der euklidischen Geometrie, der den Satz von Pythagoras aus Teil (1) von Proposition 8.8 verallgemeinert.

Beispiel 8.10. Mit Hilfe der bis jetzt entwickelten Theorie über innere Produkte können wir einige der fundamentalen Konzepte der beschreibenden Statistik verstehen. Nehmen wir also an, dass wir eine Eigenschaft haben, die durch eine relle Zahl beschrieben wird (etwa Körpergröße oder ähnliches). Nehmen wir weiters an, dass wir für eine Zahl $N \in \mathbb{N}$ diese Eigenschaft für eine Stichprobe von N Elementen erheben. Dann können wir die erhaltenen Werte x_i als Komponenten eines Vektors in $\mathbb{R}^N$ betrachten.

Betrachten wir nun den eindimensionalen Teilraum $V \subset \mathbb{R}^N$, der von jenen Vektoren gebildet wird, in denen alle Komponenten gleich sind, also $V = \{(t, \dots, t) : t \in \mathbb{R}\}$. Das orthogonale Komplement $V^\perp \subset \mathbb{R}^N$ besteht dann offensichtlich aus all jenen Vektoren $(r_1, \dots, r_N) \in \mathbb{R}^N$, für die $r_1 + \dots + r_N = 0$ gilt. Das entspricht also Stichproben mit Mittelwert Null.

Wir können nun unsere Stichprobe $x = (x_1, \dots, x_N)$ eindeutig als Summe eines Elements von V und eines Elements von $V^\perp$ schreiben. Eine Orthonormalbasis von V wird natürlich durch den Vektor $v := \frac{1}{\sqrt{N}}(1, \dots, 1)$ gebildet. Damit ist aber die Orthogonalprojektion von x auf V gegeben durch

$$\pi(x) = \langle x, v\rangle v = \tfrac{1}{N}(\textstyle\sum x_i)(1, \dots, 1) = (\bar{x}, \dots, \bar{x}),$$

gegeben, wobei $\bar{x}$ der übliche Mittelwert der Stichprobe ist. Die Projektion auf $V^\perp$ ist dann natürlich durch $x_0 := (x_1 - \bar{x}, \dots, x_N - \bar{x})$ gegeben. Dieser Vektor beschreibt gerade, wie weit die einzelnen Komponenten der Stichprobe vom Mittelwert entfernt liegen. Nach Definition ist $\frac{1}{N}\langle x_0, x_0\rangle = \frac{1}{N}\sum(x_i - \bar{x})^2 = s_x^2$ die empirische Varianz der Stichprobe und $\frac{1}{\sqrt{N}}\|x_0\| = s_x$ ist ihre Standardabweichung. Wendet man den Satz von Pythagoras auf die Zerlegung $x = \bar{x}\cdot(1, \dots, 1) + x_0$ an, dann erhält man $\|x\|^2 = N\bar{x}^2 + \|x_0\|^2$ und daraus sofort die alternative Formel $s_x^2 = (\frac{1}{N}\sum_i x_i^2) - \bar{x}^2$ für die Varianz.

Nehmen wir nun an, dass wir in der gleichen Stichprobe auch eine zweite Eigenschaft erheben, die ebenfalls durch eine reelle Zahl beschrieben werden kann. Dann erhalten wir einen zweiten Vektor $(y_i) \in \mathbb{R}^N$ den wir analog als $(\bar{y}, \dots, \bar{y}) + y_0$ zerlegen können. Wiederum nach Definition ist nun $\frac{1}{N}\langle x_0, y_0\rangle$ genau die empirische Kovarianz s_{xy} der beiden Stichproben. Gilt $x_0, y_0 \neq 0$, dann kann man den Winkel zwischen den beiden "normalisierten" Vektoren x_0 und y_0 berechnen. Da sich die Faktoren $\frac{1}{N}$ und $\frac{1}{\sqrt{N}}$ wegheben, erhält man

$$\cos(\alpha) = \frac{\langle x_0, y_0\rangle}{\|x_0\| \cdot \|y_0\|} = \frac{s_{xy}}{s_x s_y},$$

und das ist genau der empirische Korrelationskoeffizient r_{xy} der beiden Eigenschaften in der Stichprobe. Insbesondere ist $r_{xy} = \pm 1$ genau dann, wenn die beiden Vektoren linear abhängig sind, es also eine Konstante $\lambda \in \mathbb{R}$ mit $y_0 = \lambda x_0$ gibt. (Dabei stimmt das Vorzeichen von λ mit dem Vorzeichen von r_{xy} überein.) das bedeutet aber gerade, dass für alle $i = 1, \dots, N$ die Gleichung $y_i - \bar{y} = \lambda(x_i - \bar{x})$ und damit $y_i = kx_i + d$ gilt, wobei $k = \lambda$ und $d = \bar{y} - \lambda\bar{x}$. Stellt man die Stichprobe als "Punktwolke" in $\mathbb{R}^2$ (also als die Punktmenge $\{(x_i, y_i) : i = 1, \dots, N\} \subset \mathbb{R}^2$) dar, dann liegen also in diesem Fall alle Punkte auf einer Geraden. (Allgemeiner erhält man so die Regressionsgerade.)

Dualraum und adjungierte Abbildung

Als nächsten Schritt zeigen wir, dass ein inneres Produkt auf einem Vektorraum V eine Identifikation von V mit seinem Dualraum liefert. Damit kann man für eine lineare Abbildung $f : V \to V$ die duale Abbildung ebenfalls als Abbildung von V nach V betrachten, die in diesem Bild die adjungierte Abbildung zu f heißt. Wir werden dieses Konzept aber ohne die Benutzung von Resultaten über Dualräume herleiten.

8.11. Identifikation mit dem Dualraum. Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum. Nach Definition ist für jeden Vektor $v \in V$ die Zuordnung $w \mapsto \langle w, v\rangle$ eine lineare Abbildung $\varphi_v : V \to \mathbb{K}$, also ein Element des Dualraumes $V^* = L(V, \mathbb{K})$ von V.

PROPOSITION 8.11. *Sei $(V, \langle\ ,\ \rangle)$ ein (endlichdimensionaler) euklidischer bzw. unitärer Vektorraum über $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$ und $\varphi : V \to \mathbb{K}$ ein lineares Funktional. Dann gibt es ein eindeutiges Element $v_0 \in V$, sodass $\varphi(v) = \langle v, v_0\rangle$ für alle $v \in V$ gilt.*

Beweis. Betrachte zunächst den Fall eines euklidischen Vektorraumes. Für $v \in V$ setze $\varphi_v : V \to \mathbb{R}$, $\varphi_v(w) = \langle w, v\rangle$ und betrachte $v \mapsto \varphi_v$ als Funktion $V \to L(V, \mathbb{R})$. Nach Definition gilt für $v_1, v_2 \in V$, $r \in \mathbb{R}$ und alle $w \in V$

$$\varphi_{v_1+rv_2}(w) = \langle w, v_1 + rv_2\rangle = \langle w, v_1\rangle + r\langle w, v_2\rangle = \varphi_{v_1}(w) + r\varphi_{v_2}(w) = (\varphi_{v_1} + r\varphi_{v_2})(w).$$

Damit ist aber $\varphi_{v_1+rv_2} = \varphi_{v_1} + r\varphi_{v_2}$, also ist $v \mapsto \varphi_v$ eine lineare Abbildung zwischen den $\mathbb{R}$–Vektorräumen V und $L(V, \mathbb{R})$. Außerdem gilt $\varphi_v = 0$ genau dann, wenn $0 = \varphi_v(w) = \langle w, v\rangle$ für alle $w \in V$ gilt, und das gilt nur für $v = 0$. Damit ist $v \mapsto \varphi_v$ injektiv und da $\dim(V) = \dim(L(V, \mathbb{R}))$ ist ist die Abbildung sogar bijektiv und die Behauptung folgt.

Im Fall unitärer Vektorräume ist die resultierende Abbildung $V \to L(V, \mathbb{C})$ nicht linear sondern konjugiert linear, die Argumentationsweise funktioniert aber auch für solche Abbildungen. □

Identifiziert man in dieser Weise den Vektorraum V mit seinem Dualraum V^*, dann können wir zu einer Teilmenge $A \subset V$ den Annihilator A° (siehe 5.13) bilden, und diesen als Teilraum von V betrachten. Nach Definition liegt ein Element $v \in V$ genau dann in diesem Teilraum, wenn $0 = \varphi_v(a) = \langle a, v\rangle$ für alle $a \in A$ gilt, man erhält also genau den Orthogonalraum $A^\perp$ aus 8.8. Damit kann man alle Resultate über Annihilatoren aus 5.13 auf Orthogonalräume übertragen. Wir geben hier aber direkte Beweise, was ziemlich einfach ist.

Satz 8.11. *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum, $A, A' \subset V$ Teilmengen und $W, W' \subset V$ Teilräume. Dann gilt:*
(1) $A \subset A' \implies (A')^\perp \subset A^\perp$.
(2) $\dim(W^\perp) = \dim(V) - \dim(W)$.
(3) $(A^\perp)^\perp$ *ist der von A erzeugte Teilraum von V. Insbesondere ist* $(W^\perp)^\perp = W$.
(4) $(W + W')^\perp = W^\perp \cap (W')^\perp$.
(5) $(W \cap W')^\perp = W^\perp + (W')^\perp$.

Beweis. (1) ist offensichtlich und (2) haben wir schon in Proposition 8.9 bewiesen. Für (3) bemerken wir zunächst, dass $(A^\perp)^\perp$ ein Teilraum von V ist, der A enthält, weil für $a \in A$ und $v \in A^\perp$ natürlich $\langle a, v\rangle = 0$ gilt. Ist U der von A erzeugte Teilraum, dann kann jedes Element von U als Linearkombination von Elementen von A geschrieben werden. Damit folgt aber, dass jedes Element $v \in A^\perp$ auch schon in $U^\perp$ liegt, und mittels (1) erhalten wir $A^\perp = U^\perp$ und damit $(A^\perp)^\perp = (U^\perp)^\perp$. Von vorher wissen wir, dass $U \subset (U^\perp)^\perp$ gilt und nach (2) haben die beiden Teilräume gleiche Dimension, also folgt $U = (A^\perp)^\perp$.

(4) Da $W \subset (W + W')$ gilt, folgt $(W + W')^\perp \subset W^\perp$ nach (1). Analog gilt das für W' und damit folgt $(W + W')^\perp \subset W^\perp \cap (W')^\perp$. Liegt umgekehrt ein Vektor $v \in V$ sowohl in $W^\perp$ als auch in $(W')^\perp$, dann liegt v offensichtlich in $(W \cup W')^\perp$ und von oben wissen wir, dass das mit $(W + W')^\perp$ übereinstimmt.

(5) Nach (4) und (3) wissen wir, dass $(W^\perp + (W')^\perp)^\perp = W \cap W'$ gilt. Daraus folgt die Behauptung, indem man auf beiden Seiten den Orthogonalraum bildet. □

8.12. Anwendungen. Wie wollen hier zwei geometrische Anwendungen der Identifikation von V mit V^* durch ein inneres Produkt geben, nämlich einerseits die Hessesche Normalform für affine Hyperebenen und andererseits die Konstruktion des Normalvektors in $\mathbb{R}^2$ und des Kreuzprodukts in $\mathbb{R}^3$.

(1) Betrachte $\mathbb{R}^n$ mit dem Standard inneren Produkt $\langle\ ,\ \rangle$. Ist $v \in \mathbb{R}^n$ mit $v \neq 0$ beliebig, dann der von v erzeugt Teilraum gerade $\mathbb{R}v = \{\lambda v : \lambda \in \mathbb{R}\}$ und nach Teil

(2) des Satzes ist das orthogonale Komplement $\{v\}^\perp = \{w \in \mathbb{R}^n : \langle w, v\rangle = 0\}$ ein Teilraum der Dimension $n-1$, eine sogenannte *Hyperebene*. Ist umgekehrt $W \subset \mathbb{R}^n$ eine Hyperebene, dann ist $W^\perp$ ein eindimensionaler Teilraum von $\mathbb{R}^n$. Eine Baisis für diesen Teilraum ist durch jedes Element $v \neq 0$ darin gegeben und nimmt man zusätzlich $\|v\| = 1$ an, dann ist v bis auf das Vorzeichen eindeutig bestimmt. Man kann also jede Hyperebene in der Form $\{v\}^\perp$ mit $\|v\| = 1$ darstellen. Insbesondere ist in $\mathbb{R}^2$ jede Gerade durch Null und in $\mathbb{R}^3$ jede Ebene durch Null von dieser Form.

Betrachten wir nun eine *"affine Hyperebene"*, also eine Teilmenge der Form $a+W := \{a+w : w \in W\}$ für einen fixen Punkt $a \in \mathbb{R}^n$ und eine Hyperebene $W \subset \mathbb{R}^n$, vergleiche mit 5.2. Ist $v \in W^\perp$ mit $\|v\| = 1$, dann gilt für alle $w \in W$ natürlich $\langle v, w\rangle = 0$, also $\langle v, a+w\rangle = \langle v, a\rangle$. Damit gilt $\langle v, y\rangle = \langle v, a\rangle$ für alle $y \in a+W$. Ist umgekehrt $y \in \mathbb{R}^n$ so, dass $\langle v, y\rangle = \langle v, a\rangle$ gilt, dann ist $\langle x, y-a\rangle = 0$, also $y - a \in W$ und damit $y \in a + W$. Das bedeutet aber gerade, dass man $a+W$ als $\{y \in \mathbb{R}^n : \langle v, y\rangle = d\}$ für einen Einheitsnormalvektor $v \in \mathbb{R}^n$ zu W und eine geeignete Zahl $d \in \mathbb{R}$ beschreiben kann.

Die Bedeutung von d is leicht zu verstehen: Nach Proposition 8.9 ist $\mathbb{R}^n = W \oplus W^\perp$, also kann man jedes Element $y \in a + W$ eindeutig als $w + \lambda v$ für $w \in W$ und $\lambda \in \mathbb{R}$ schreiben. Da $y + w \in a + W$ für alle $y \in a + W$ und $w \in W$ gilt hat $(a + W) \cap W^\perp$ genau ein Element, also gibt es ein eindeutiges $\lambda \in \mathbb{R}$, sodass $\lambda v \in a + W$ gilt, und wegen $\|v\| = 1$ erhalten wir sofort $\lambda = d$. Wählen wir v so, dass $d \geq 0$ gilt, dann ist d genau der *Normalabstand* der affinen Hyperebene $a + W$ vom Nullpunkt.

(2) Betrachten wir wieder $\mathbb{R}^n$ mit dem Standard inneren Produkt $\langle\ ,\ \rangle$ und sei $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ die Determinante. Fixieren wir beliebige Vektoren $v_1, \dots, v_{n-1} \in \mathbb{R}^n$, dann ist $y \mapsto \det(v_1, \dots, v_{n-1}, y)$ eine lineare Abbildung, also gibt es einen eindeutigen Vektor $x \in \mathbb{R}^n$, sodass $\langle x, y\rangle = \det(v_1, \dots, v_{n-1}, y)$ für alle $y \in \mathbb{R}^n$ gilt.

Im Fall $n = 2$ bedeutet das, dass wir für einen fix gewählten Vektor $v \in \mathbb{R}^2$ einen eindeutigen Vektor $x \in \mathbb{R}^2$ finden, der $\langle x, y\rangle = \det(v, y)$ für alle $y \in \mathbb{R}^2$ erfüllt. Insbesondere ist $\langle x, v\rangle = \det(v, v) = 0$, also x ein Normalvektor zu v. Ist $v = (v_1, v_2)$ und $x = (x_1, x_2)$, dann ist $x_1 = \langle x, e_1\rangle = \det(v, e_1) = -v_2$ und $x_2 = \langle x, e_2\rangle = \det(v, e_2) = v_1$, also ist $x = (-v_2, v_1)$. Insbesondere ist $\|x\| = \|v\|$. Nun ist $|\langle x, y\rangle|$ das Produkt der Länge der von x mit der Länge der Projektion von y auf x. Da $x \perp v$ gilt, ist die Länge dieser Projektion genau die Höhe des von y und v aufgespannten Parallelogramms, also ist $|\det(v, y)| = |\langle x, y\rangle|$ genau die Fläche des von v und y aufgespannten Parallelogramms. Man zeigt leicht (siehe Übungen), dass $|\det(v, y)| = \|v\| \cdot \|y\| \sin(\alpha)$ gilt, wobei α der Winkel zwischen v und y ist.

Im Fall $n = 3$ erhalten wir zu zwei Vektoren $v, w \in \mathbb{R}^3$ einen eindeutigen Vektor $u \in \mathbb{R}^3$, sodass $\langle u, y\rangle = \det(v, w, y)$ für alle $y \in \mathbb{R}^3$ gilt. Bezeichnet man diesen Vektor mit $v \times w$, dann definiert $(v, w) \mapsto v \times w$ eine Abbildung $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, das sogenannte *Kreuzprodukt*. Aus der Linearität der Determinante in jeder Eintragung folgt sofort, dass $(v + \lambda v') \times w = v \times w + \lambda(v' \times w)$ und $v \times (w + \lambda w') = v \times w + \lambda(v' \times w)$ gelten. Da die Determinante alternierend ist, erhalten wir $w \times v = -v \times w$ und $\langle v \times w, v\rangle = \langle v \times w, w\rangle = 0$.

Sind v und w linear abhängig, dann ist $\det(v, w, x) = 0$ für alle $x \in \mathbb{R}^3$, also $v \times w = 0$. Ist andererseits $\{v, w\}$ linear unabhängig, dann finden wir ein $y \in \mathbb{R}^3$, sodass $\{v, w, y\}$ eine Basis, und damit $\det(v, w, y) = \langle v \times w, y\rangle \neq 0$ gilt. Somit gilt $v \times w = 0$ genau dann, wenn v und w linear abhängig sind.

Schließlich ist nach Konstruktion $\det(v, w, v \times w) = \|v \times w\|^2$. Ist also $\{v, w\}$ linear unabhängig, dann ist $\{v, w, v \times w\}$ eine Basis für $\mathbb{R}^3$. Eine explizite Formel für $v \times w$ erhält man, indem man benutzt, dass die $i$te Komponente von $v \times w$ gerade als $\langle v \times w, e_i\rangle =$

$\det(v, w, e_i)$ berechnet werden kann. Entwickelt man diese Determinante jeweils nach der letzten Spalte, dann erhält man

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ -v_1 w_3 + v_3 w_1 \\ v_1 w_2 - v_2 w_1 \end{pmatrix}.$$

Aus dieser Formel verifiziert man direkt, dass $\|v \times w\|^2 = \|v\|^2 \cdot \|w\|^2 - \langle v, w\rangle^2$ gilt. Sind v und w linear unabhängig, dann wählen wir wie in Bemerkung (2) von 8.8 einen Vektor $\tilde{v} \neq 0$ in der von v und w erzeugten Ebene mit $\langle v, \tilde{v}\rangle = 0$, und schreiben w als $av + b\tilde{v}$. Dann wissen wir, dass $\langle v, w\rangle^2 = \|v\|^2 \cdot \|av\|^2$ gilt, also erhalten wir $\|v \times w\|^2 = \|v^2\|(\|w\|^2 - \|av\|^2)$. Nach dem Satz von Pythagoras liefert die Klammer $\|b\tilde{v}\|^2$, also erhalten wir $\|v \times w\| = \|v\| \cdot \|b\tilde{v}\|$, und das kann man offensichtlich als die Fläche des von v und w aufgespannten Parallelogramms interpretieren. Nun ist $|\langle v \times w, y\rangle|$ gerade das Produkt der Länge von $v \times w$ mit der Länge der Projektion von y auf $v \times w$. Da $v \times w$ normal auf die von v und w erzeugte Ebene steht, ist die Länge dieser Projektion genau die Höhe des von v, w und y erzeugten Parallelepipeds. Somit sehen wir, dass $|\det(v, w, y)| = |\langle v \times w, y\rangle|$ genau das Volumen des von v, w und y aufgespannten Parallelepipeds ist.

8.13. Die adjungierte Abbildung. Wir kommen jetzt zu der Interpretation der dualen Abbildung, die durch die Identifikation eines euklidischen oder unitären Vektorraumes mit seinem Dualraum ermöglicht wird. Wir werden allerdings wieder nicht auf Resultate über duale Abbildungen aufbauen, sondern direkte Beweise geben. Für ein tieferes Verständnis ist aber sehr hilfreich, sich die Zusammenhänge zur Dualität bewusst zu machen.

Satz 8.13. *Seien $(V, \langle\ ,\ \rangle_V)$, $(W, \langle\ ,\ \rangle_W)$ und $(Z, \langle\ ,\ \rangle_Z)$ endlichdimensionale euklidische bzw. unitäre Vektorräume über $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$. Dann gilt:*

(1) Zu einer linearen Abbildung $f : V \to W$ gibt es eine eindeutige lineare Abbildung $f^* : W \to V$, sodass $\langle f(v), w\rangle_W = \langle v, f^*(w)\rangle_V$ für alle $v \in V$ und $w \in W$ gilt.

(2) Die adjungierte Abbildung $(f^*)^* : V \to W$ zu $f^* : W \to V$ ist durch $(f^*)^* = f$ gegeben.

(3) $\mathrm{Ker}(f^*) = \mathrm{Im}(f)^\perp \subset W$ *und* $\mathrm{Im}(f^*) = \mathrm{Ker}(f)^\perp \subset V$.

(4) Sind $\mathcal{B} \subset V$ und $\mathcal{C} \subset W$ Orthonormalbasen, und ist $[f]^{\mathcal{B}}_{\mathcal{C}} = (a_{ij})$ die Matrixdarstellung von f bezüglich dieser Basen, dann ist die Matrixdarstellung $[f^*]^{\mathcal{C}}_{\mathcal{B}} = (b_{ij})$ charakterisiert durch $b_{ij} = \bar{a}_{ji}$.

(5) Ist $\tilde{f} : V \to W$ eine weitere lineare Abbildung und $\lambda \in \mathbb{K}$, dann gilt $(f + \lambda\tilde{f})^* = f^* + \bar{\lambda}\tilde{f}^*$.

(6) Ist $g : W \to Z$ eine weitere lineare Abbildung, dann ist $(g \circ f)^* = f^* \circ g^*$.

Beweis. (1) Fixieren wir zunächst ein Element $w_0 \in W$ und betrachten die Funktion $V \to \mathbb{K}$, die durch $\langle f(v), w_0\rangle_W$ gegeben ist. Aus der Linearität von f folgt sofort, dass auch diese Abbildung linear ist. Somit gibt es nach Proposition 8.11 ein eindeutig bestimmtes Element $\tilde{v} \in V$, sodass $\langle f(v), w_0\rangle_W = \langle v, \tilde{v}\rangle_V$. Setzt man $f^*(w_0) := \tilde{v}$, dann können wir so jedem Vektor $w \in W$ einen Vektor $f^*(w) \in V$ zuordnet, sodass die gewünschte Gleichung gilt. Das definiert eine Funktion $f^* : W \to V$.

Es bleibt also nur noch zu zeigen, dass f^* linear ist: Seien dazu $w_1, w_2 \in W$ und $\lambda \in \mathbb{K}$. Dann rechnen wir für beliebiges $v \in V$

$$\begin{aligned} \langle v, f^*(w_1 + \lambda w_2)\rangle_V &= \langle f(v), w_1 + \lambda w_2\rangle_W = \langle f(v), w_1\rangle_W + \bar{\lambda}\langle f(v), w_2\rangle_W \\ &= \langle v, f^*(w_1)\rangle_V + \bar{\lambda}\langle v, f^*(w_2)\rangle_V = \langle v, f^*(w_1) + \lambda f^*(w_2)\rangle_V, \end{aligned}$$

und da $\langle\ ,\ \rangle_V$ nicht entartet ist, folgt $f^*(w_1 + \lambda w_2) = f^*(w_1) + \lambda f^*(w_2)$.

(2) Für $v \in V$ und $w \in W$ rechnen wir einfach

$$\langle w, (f^*)^*(v)\rangle_W = \langle f^*(w), v\rangle_V = \overline{\langle v, f^*(w)\rangle_V} = \overline{\langle f(v), w\rangle_W} = \langle w, f(v)\rangle_W.$$

Da das für alle $w \in W$ gilt, folgt $(f^*)^*(v) = f(v)$ für alle $v \in V$, also $(f^*)^* = f$.

(3) Da $\langle\ ,\ \rangle_V$ nicht entartet ist, ist $f^*(w) = 0$ äquivalent zu $0 = \langle v, f^*(w)\rangle_V = \langle f(v), w\rangle_V$ für alle $v \in V$, also zu $w \in \mathrm{Im}(f)^\perp$.

Ist andererseits $v \in \mathrm{Ker}(f)$, dann gilt $0 = \langle f(v), w\rangle_W = \langle v, f^*(w)\rangle_V$, also gilt $f^*(w) \in \mathrm{Ker}(f)^\perp$. Das gilt aber für alle $w \in W$, also folgt $\mathrm{Im}(f^*) \subset \mathrm{Ker}(f)^\perp$. Nach Satz 8.11 ist $\dim(\mathrm{Ker}(f)^\perp) = \dim(V) - \dim(\mathrm{Ker}(f)) = \dim(\mathrm{Im}(f))$. Nach dem Dimensionssatz für lineare Abbildungen ist $\dim(\mathrm{Im}(f^*)) = \dim(W) - \dim(\mathrm{Ker}(f^*)) = \dim(W) - \dim(\mathrm{Im}(f)^\perp) = \dim(\mathrm{Im}(f))$. Damit hat der Teilraum $\mathrm{Im}(f^*) \subset \mathrm{Ker}(f)^\perp$ die gleiche Dimension wie $\mathrm{Ker}(f)^\perp$, also folgt Gleichheit.

(4) Ist $\mathcal{B} = \{v_1, \dots, v_n\}$ und $\mathcal{C} = \{w_1, \dots, w_m\}$, dann wissen wir aus Proposition 8.8, dass $a_{ij} = \langle f(v_j), w_i\rangle_W$ gilt. Analog ist

$$b_{ij} = \langle f^*(w_j), v_i\rangle_V = \overline{\langle v_i, f^*(w_j)\rangle_V} = \overline{\langle f(v_i), w_j\rangle_W} = \bar{a}_{ji}.$$

(5) Für beliebige Elemente $v \in V$ und $w \in W$ rechnen wir

$$\begin{aligned}\langle v, (f + \lambda\tilde{f})^*(w)\rangle_V &= \langle (f + \lambda\tilde{f})(v), w\rangle_W = \langle f(v), w\rangle_W + \langle \lambda\tilde{f}(v), w\rangle_W \\ &= \langle v, f^*(w)\rangle_V + \lambda\langle v, \tilde{f}^*(w)\rangle_V = \langle v, f^*(w) + \bar{\lambda}\tilde{f}^*(w)\rangle_V,\end{aligned}$$

und die Behauptung folgt analog wie in (2).

(6) Für $v \in V$ und $z \in Z$ rechnen wir

$$\begin{aligned}\langle v, (g \circ f)^*(z)\rangle_V &= \langle (g \circ f)(v), z\rangle_Z = \langle g(f(v)), z\rangle_Z \\ &= \langle f(v), g^*(z)\rangle_W = \langle v, f^*(g^*(z))\rangle_V,\end{aligned}$$

und die Behauptung folgt wie in (2). □

Definition 8.13. (1) Für eine lineare Abbildung $f : V \to W$ zwischen zwei euklidischen bzw. unitären Vektorräumen heißt die lineare Abbildung $f^* : W \to V$ aus Teil (1) des Satzes die *zu f adjungierte Abbildung*.

(2) Für eine Matrix $A = (a_{ij}) \in M_{m,n}(\mathbb{C})$ nennt man die Matrix $(b_{ij}) \in M_{n,m}(\mathbb{C})$ die gegeben ist durch $b_{ij} = \overline{a_{ji}}$ die *adjungierte Matrix* zu A und bezeichnet sie mit A^*.

Bemerkung 8.13. (1) Im Fall euklidischer Vektorräume erhält man in Teil (4) des Satzes als Matrixdarstellung für f^* einfach die Transponierte der Matrixdarstellung von f, siehe 5.12. Deshalb spricht man in diesem Fall oft auch von der *Transponierten* der linearen Abbildung f und schreibt dafür auch f^t statt f^*. Die Tatsache, dass im Komplexen nicht einfach die transponierte Matrix auftritt hängt damit zusammen, dass die Identifikation mit dem Dualraum in diesem Fall konjugiert linear ist.

(2) Sei $f : V \to W$ eine lineare Abbildung zwischen euklidischen oder unitären Vektorräumen und sein $\mathcal{B}$ und $\mathcal{C}$ Orthonormalbasen für V und W. Setzt $A = [f]^{\mathcal{B}}_{\mathcal{C}}$, dann gilt $[f^*]^{\mathcal{C}}_{\mathcal{B}} = A^*$ (im komplexen Fall) bzw. $[f^*]^{\mathcal{C}}_{\mathcal{B}} = A^t$ im reellen Fall.

Betrachtet man insbesondere den Fall $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$ für $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{C}$, dann sind die Standardbasen orthonormal bezüglich der Standard inneren Produkte. Damit ist die adjungierte Abbildung zu $x \mapsto Ax$ (bezüglich des Standard inneren Produkts) einfach durch $x \mapsto A^*x$ bzw. $x \mapsto A^tx$ gegeben.

Als Beispiel für die Verwendung adjungierter Abbildungen können wir die Orthogonalprojektion auf einen Teilraum eines Euklidischen oder unitären Vektorraums charakterisieren. In Lemma 8.9 haben wir gesehen, dass für ein lineare Abbildung $\varphi : V \to V$, die $\varphi^2 = \varphi$ erfüllt, immer $V = \mathrm{Im}(\varphi) \oplus \mathrm{Ker}(\varphi)$ gilt und φ die Projektion auf die erste Komponente in dieser Zerlegung ist.

PROPOSITION 8.13. *Sei $(V, \langle\ ,\ \rangle)$ ein Euklidischer oder unitärer Vektorraum und $\varphi : V \to V$ eine Projektion, also eine lineare Abbildung, die $\varphi^2 = \varphi$ erfüllt. Dann ist φ genau dann die Orthogonalprojektion auf* $\mathrm{Im}(\varphi)$*, wenn zusätzlich $\varphi^* = \varphi$ gilt.*

BEWEIS. Nach Definition ist φ genau dann die Orthogonalporjektion auf $\mathrm{Im}(\varphi)$, wenn $\mathrm{Ker}(\varphi) = \mathrm{Im}(\varphi)^\perp$ ist. Ist $\varphi^* = \varphi$, dann folgt das sofort aus Teil (3) von Satz 8.13.

Sei umgekehrt $\mathrm{Ker}(\varphi) = \mathrm{Im}(\varphi)^\perp$. Dann gilt wegen $w - \varphi(w) \in \mathrm{Ker}(\varphi)$ immer $\langle \varphi(v), w - \varphi(w) \rangle = 0$, also $\langle \varphi(v), w \rangle = \langle \varphi(v), \varphi(w) \rangle$. Analog ist aber auch $\langle \varphi(v), \varphi(w) \rangle = \langle v, \varphi(w) \rangle$. Damit erhalten wir aber für alle $v, w \in V$

$$\langle \varphi(v), w \rangle = \langle v, \varphi(w) \rangle = \langle v, \varphi^*(w) \rangle$$

und damit $\varphi^*(w) = \varphi(w)$ für alle w. $\square$

KAPITEL 9

Spezielle lineare Abbildungen – euklidische Geometrie

In diesem Kapitel studieren wir lineare Abbildungen zwischen euklidischen und unitären Vektorräumen. Das führt einerseits zu einem Begriff von Diagonalisierbarkeit, der an ein inneres Produkt angepasst und deutlich einfacher zu verstehen ist, als die allgemeine Diagonalisierbarkeit. Andererseits habe die Inhalte dieses Kapitels enge Bezüge zur euklidischen Geometrie.

Orthogonale und unitäre Abbildungen

Es gibt einen offensichtlichen Verträglichkeitsbegriff zwischen einer linearen Abbildung und einem inneren Produkt, mit dem wir unsere Überlegungen beginnen. Das liefert auch den richtigen Begriff von Isomorphie von euklidischen und unitären Vektorräumen.

9.1. Definitionen und grundlegende Eigenschaften. Der Begriff der orthogonalen bzw. unitären Abbildung ist ganz naheliegend:

Definition 9.1. Seien $(V, \langle\ ,\ \rangle_V)$ und $(W, \langle\ ,\ \rangle_W)$ zwei euklidische (bzw. unitäre) Vektorräume.

(1) Eine lineare Abbildung $f : V \to W$ heißt *orthogonal* (bzw. *unitär*), wenn $\langle f(v_1), f(v_2)\rangle_W = \langle v_1, v_2\rangle_V$ für alle $v_1, v_2 \in V$ gilt.

(2) $f : V \to W$ heißt ein *euklidischer (bzw. unitärer) Isomorphismus*, falls f orthogonal (bzw. unitär) und bijektiv ist.

(3) Eine Matrix $A \in M_n(\mathbb{R})$ heißt *orthogonal*, wenn die lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto Ax$ orthogonal bezüglich des Standard inneren Produktes ist.

(4) Eine Matrix $A \in M_n(\mathbb{C})$ heißt *unitär*, wenn die lineare Abbildung $\mathbb{C}^n \to \mathbb{C}^n$, $z \mapsto Az$ unitär bezüglich des Standard hermitischen inneren Produktes ist.

Für einen euklidischen (unitären) Isomorphismus $f : V \to W$ ist auch die inverse Abbildung $f^{-1} : W \to V$ ein euklidischer (unitärer) Isomorphismus. Es gilt nämlich

$$\langle f^{-1}(w_1), f^{-1}(w_2)\rangle_V = \langle f(f^{-1}(w_1)), f(f^{-1}(w_2))\rangle_W = \langle w_1, w_2\rangle_W.$$

Allgemein können wir orthogonale (unitäre) lineare Abbildungen wie folgt charakterisieren:

Satz 9.1. *Seien $(V, \langle\ ,\ \rangle_V)$ und $(W, \langle\ ,\ \rangle_W)$ zwei endlichdimensionale euklidische (bzw. unitäre) Vektorräume. Dann sind für eine lineare Abbildung $f : V \to W$ mit adjungierter Abbildung $f^* : W \to V$ äquivalent:*

(1) f ist orthogonal (unitär).

(2) $\|f(v)\| = \|v\|$ für alle $v \in V$.

(3) Es gibt eine Orthonormalbasis $\{v_1, \dots, v_n\}$ für V, sodass $\{f(v_1), \dots, f(v_n)\} \subset W$ ein Orthonormalsystem ist.

(4) Für jedes Orthonormalsystem $\{v_1, \dots, v_k\} \subset V$ ist $\{f(v_1), \dots, f(v_k)\} \subset W$ ein Orthonormalsystem.

(5) $f^* \circ f = \mathrm{id}_V$.

Beweis. (1) $\Rightarrow$ (2) ist klar wegen $\|v\| = \sqrt{\langle v, v\rangle}$ und (2)$\Rightarrow$ (1) folgt sofort aus den Polarisierungsformeln aus Satz 8.7. (1) $\Rightarrow$ (4) ist klar nach Definition eines Orthonormalsystems, und (4) $\Rightarrow$ (3) ist offensichtlich.

(3) $\Rightarrow$ (1): Sind $x = x_1v_1 + \cdots + x_nv_n$ und $y = y_1v_1 + \cdots + y_nv_n$ beliebige Elemente von V, dann ist nach Proposition 8.8 $\langle x, y\rangle_V = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n$. Andererseits ist $f(x) = x_1f(v_1) + \cdots + x_nf(v_n)$ und analog für y und damit folgt

$$\langle f(x), f(y)\rangle_W = \textstyle\sum_{i,j} x_i\bar{y}_j\langle f(v_i), f(v_j)\rangle = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n.$$

(1) $\Leftrightarrow$ (5) Nach Definition ist $\langle v_1, f^*(f(v_2))\rangle_V = \langle f(v_1), f(v_2)\rangle_W$. Da das innere Produkt nicht entartet ist, ist $f^*(f(v_2)) = v_2$ äquivalent zu $\langle v_1, f^*(f(v_2))\rangle_V = \langle v_1, v_2\rangle_V$ für alle $v_1 \in V$, und damit folgt die Behauptung. □

Korollar 9.1. *(1) Orthogonale und unitäre lineare Abbildungen sind automatisch injektiv, also bei gleicher Dimension automatisch euklidische bzw. unitäre Isomorphismen.*

(2) Jeder n–dimensionale euklidische (unitäre) Vektorraum ist orthogonal (unitär) isomorph zu $\mathbb{R}^n$ ($\mathbb{C}^n$) mit dem Standard (hermitischen) inneren Produkt.

(3) $f : V \to V$ ist genau dann orthogonal (unitär) wenn $f^* = f^{-1}$ gilt.

(4) Ist $f : V \to V$ orthogonal (unitär) und $\lambda \in \mathbb{K}$ ein Eigenwert von f, dann ist $|\lambda| = 1$.

(5) Ist $f : V \to V$ orthogonal (unitär), dann ist $|\det(f)| = 1$.

(6) Für eine Matrix $A \in M_n(\mathbb{R})$ ($M_n(\mathbb{C})$) sind äquivalent:

(a) *A ist orthogonal (unitär)*
(b) *die Spaltenvektoren von A bilden eine Orthonormalbasis von $\mathbb{R}^n$ ($\mathbb{C}^n$).*
(c) *$A^t = A^{-1}$ ($A^* = A^{-1}$).*

Beweis. (1) Nach Teil (2) des Satzes folgt aus $f(v) = 0$ schon $\|v\| = 0$, also $v = 0$, also ist $\mathrm{Ker}(f) = \{0\}$, also f injektiv.

(2) Nach Korollar 8.8 findet man immer eine Orthonormalbasis. Die eindeutige lineare Abbildung nach $\mathbb{K}^n$, die diese Orthonormalbasis auf die Standardbasis abbildet ist nach Teil (4) des Satzes ein orthogonaler (unitärer) Isomorphismus.

(3) Ist f orthogonal, dann ist f nach (1) invertierbar und nach Teil (5) des Satzes ist $f^* \circ f = \mathrm{id}$. Komponiert man diese Gleichung von rechts mit f^{-1}, so erhält man $f^* = f^{-1}$.

(4) folgt sofort aus $\|f(v)\| = \|v\|$.

(5) Ist $\mathcal{B}$ eine Orthonormalbasis für V und $A = [f]_{\mathcal{B}}$, dann ist nach Satz 8.13 $[f^*]_{\mathcal{B}} = A^*$ (bzw. A^t). Nach Satz 6.9 ist $\det(A^t) = \det(A)$ und daraus folgt natürlich $\det(A^*) = \overline{\det(A)}$. Damit erhalten wir aus $f^* \circ f = \mathrm{id}$ im reellen Fall $\det(A)^2 = 1$ und im komplexen Fall $\det(A)\overline{\det(A)} = 1$, also in beiden Fällen $|\det(A)| = 1$.

(6) Die Äquivalenz der ersten beiden Bedingungen ist die Matrizenversion von Teil (4) des Satzes, die Äquivalenz zur letzten die Matrizenversion von (3). □

Bemerkung 9.1. Klarerweise ist die Komposition von zwei orthogonalen (unitären) linearen Abbildungen wiederum orthogonal (unitär), und wir haben bereits gesehen, dass für eine invertierbare orthogonale (unitäre) lineare Abbildung die Inverse ebenfalls orthogonal (unitär) ist. Damit bilden aber die orthogonalen linearen Abbildungen auf einem euklidischen Vektorraum V eine Gruppe $O(V)$, die *orthogonale Gruppe* von V. Darin ist $SO(V) := \{f \in O(V) : \det(f) = 1\}$ eine Untergruppe, die *spezielle orthogonale Gruppe* von V. Betrachten wir insbesondere $\mathbb{R}^n$ mit dem Standard inneren Produkt, dann können wir diese Gruppen mit den Gruppen $O(n) := \{A \in M_n(\mathbb{R}) : A^t = A^{-1}\}$ bzw. $SO(n) := \{A \in O(n) : \det(A) = 1\}$ identifizieren.

Analog erhalten wir für einen unitären Vektorraum V die *unitäre Gruppe* $U(V)$ und die *spezielle unitäre Gruppe* $SU(V)$, die wir im Fall $V = \mathbb{C}^n$ mit $U(n) := \{A \in M_n(\mathbb{C}) : A^* = A^{-1}\}$ bzw. $SU(n) = \{A \in U(n) : \det(A) = 1\}$ identifizieren können.

9.2. Die QR–Zerlegung. Eine Matrix A ist also genau dann orthogonal bzw. unitär, wenn die Spaltenvektoren eine Orthonormalbasis bilden. Damit können wir aber das Gram–Schmidt Orthonormalisierungsverfahren aus 8.8 auch in Termen von Matrizen formulieren. Das liefert eine Zerlegung für beliebige invertierbare Matrizen, die in der Theorie der Matrizengruppen unter dem Namen *Iwasawa–Zerlegung* ziemlich bedeutsam ist, aber auch in der angewandten Mathematik (unter dem Name *QR–Zerlegung*) eine wichtige Rolle spielt.

Satz 9.2. *Sei A eine invertierbare $n \times n$–Matrix über $\mathbb{R}$ (bzw. $\mathbb{C}$). Dann gibt es eine orthogonale (bzw. unitäre) Matrix Q und eine obere Dreiecksmatrix R mit reellen positiven Eintragungen auf der Hauptdiagonale, sodass $A = QR$ gilt. Die Matrizen Q und R sind eindeutig bestimmt.*

Beweis. Wir beweisen zuerst die Existenz der Zerlegung. Sind $a_1, \dots, a_n$ die Spaltenvektoren von A, dann bilden diese Vektoren eine Basis, weil A invertierbar ist. Damit können wir das Gram–Schmidt Orthonormalisierungsverfahren aus Satz 8.8 anwenden. Das liefert eine Orthonormalbasis $\{v_1, \dots, v_n\}$, sodass für jeden Index j, der Vektor a_j als Linearkombination der Form $r_{1j}v_1 + \dots + r_{jj}v_j$ geschrieben werden kann, wobei r_{jj} reell und positiv ist. Setzen wir nun $R := (r_{ij})$, dann ist $r_{ij} = 0$ für $i > j$, also ist R eine obere Dreiecksmatrix, und wir haben schon bemerkt, dass die Elemente auf der Hauptdiagonale reell und positiv sind.

Sei andererseits Q die Matrix, deren Spaltenvektoren die Vektoren v_j sind. Dann ist Q nach Korollar 9.1 orthogonal (bzw. unitär), weil die v_i eine Orthonormalbasis bilden. Setzen wir nun $Q = (q_{ij})$ und $QR = (b_{ij})$, dann ist $b_{ij} = \sum_k q_{ik}r_{kj} = \sum_k r_{kj}q_{ik}$. Diese Gleichung sagt aber genau, dass man die $j$te Spalte von QR erhält, indem man die $k$te Spalte von Q mit r_{kj} multiplziert und dann die erhaltenen Vektoren addiert. Da die Spalten von Q gerade die v_j sind, folgt $QR = A$ nach Konstruktion.

Zur Eindeutigkeit: Sind Q_1, Q_2 und R_1, R_2 Matrizen der verlangten Form, sodass $Q_1R_1 = Q_2R_2$ gilt, dann ist $Q_2^{-1}Q_1 = R_2R_1^{-1}$. Dann ist aber $Q_2^{-1}Q_1$ als Produkt orthogonaler bzw. unitärer Matrizen selbst orthogonal bzw. unitär. Andererseits behaupten wir, dass $R_2R_1^{-1}$ eine obere Dreiecksmatrix mit positiven reellen Eintragungen auf der Hauptdiagonale ist. Haben wir das gezeigt, dann sind wir fertig, weil die Hauptdiagonalelemente dann Eigenwerte einer orthogonalen (unitären) linearen Abbildung sind. Damit haben sie nach Teil (4) von Korollar 9.1 Betrag Eins, also sind alle Hauptdiagonalelemente gleich Eins, weil sie reell und positiv sein müssen. Da die Spaltenvektoren von eine orthogonalen (unitären) Matrix ein Orthonormalsystem bilden, folgt sofort $R_2R_1^{-1} = \mathbb{I}$ und damit $R_1 = R_2$, also auch $Q_1 = Q_2$.

Betrachten wir eine obere Dreiecksmatrix $R = (r_{ij})$, sodass alle $r_{ii} \neq 0$ sind und setzen $R^{-1} = (s_{ij})$. Dann ist die erste Spalte von $\mathbb{I} = R^{-1}R$ gerade r_{11} mal die erste Spalte von R, also $s_{11} = (r_{11})^{-1}$ und $s_{i1} = 0$ für $i > 1$. Analog überlegt man leicht direkt, dass $s_{ij} = 0$ für $i > j$ und $s_{ii} = (r_{ii})^{-1}$ gilt. Damit folgt aber, dass R_1^{-1} eine obere Dreiecksmatrix mit postiven reellen Eintragungen auf der Hauptdaigonale ist. Schließlich folgt aus der Formel für die Matrizenmultiplikation sofort, dass das Produkt zweier oberen Dreiecksmatrizen selbst eine obere Dreiecksmatrix ist, wobei in der Hauptdiagonale die Produkte der Hauptdiagonalelemente der beiden Matrizen stehen. □

9.3. Spiegelungen. Das einfachste Beispiel einer nichttrivialen orthogonalen (unitären) Abbildung ist die Spiegelung an einer Hyperebene. Nach 8.12 kann man jede Hyperebene W in einem euklidischen oder unitären Vektorraum V als $W = \{a\}^\perp$ für einen Einheitsvektor a (also einen Vektor mit $\|a\| = 1$) in V schreiben. Definieren wir nun für so ein $a \in V$ die *Spiegelung* $s_a : V \to V$ an $\{a\}^\perp$ durch $s_a(v) := v - 2\langle v, a\rangle a$.

LEMMA 9.3. *(1) Für jedes $a \in V$ mit $\|a\| = 1$ ist die Spiegelung s_a an $\{a\}^\perp$ eine orthogonale (unitäre) lineare Abbildung mit $(s_a)^{-1} = s_a$ und $\det(s_a) = -1$. Die Spiegelung ist charakterisiert durch $s_a(a) = -a$ und $s_a(v) = v$ falls $\langle v, a\rangle = 0$.*

(2) Für $v \neq w \in V$ gibt es genau dann einen Einheitsvektor $a \in V$ mit $s_a(v) = w$, wenn $\|v\| = \|w\|$ sowie $\langle v, w\rangle \in \mathbb{R}$ gilt.

BEWEIS. (1) Die Abbildung s_a ist offensichtlich linear, und nach Definition gilt $s_a(a) = a - 2\langle a, a\rangle a = -a$ und $s_a(v) = v$ falls $\langle v, a\rangle = 0$. Nach Proposition 8.9 ist $V = \mathbb{K}{\cdot}a \oplus \{a\}^\perp$, also kann man jedes $v \in V$ eindeutig als $\lambda a + w$ mit $\langle w, a\rangle = 0$ schreiben. Damit ist aber $s_a(v) = -\lambda a + w$, also ist s_a durch die obigen Eigenschaften eindeutig bestimmt. Nach dem Satz von Pythagoras liefert diese Zerlegung $\|v\|^2 = |\lambda|^2 + \|w\|^2 = \|s_a(v)\|^2$, also ist s_a orthogonal nach Satz 9.1. Erweitern wir das Orthonormalsystem $\{a\} \subset V$ zu einer Orthonormalbasis $\mathcal{B} = \{a, v_2, \dots, v_n\}$, dann ist $\langle v_i, a\rangle = 0$ für alle $i = 2, \dots, n$, also ist $[s_a]_\mathcal{B}$ eine Diagonalmatrix mit einer -1 und sonst lauter Einsen auf der Hauptdiagonale, also ist $s_a \circ s_a = \mathrm{id}$ und $\det(s_a) = -1$.

(2) Das s_a orthogonal bzw. unitär ist, ist $\|s_a(v)\| = \|v\|$. Außerdem ist nach Definition $\langle v, s_a(v)\rangle = \langle v, v\rangle - 2\overline{\langle v, a\rangle}\langle v, a\rangle \in \mathbb{R}$. Damit bleibt nur noch zu zeigen, dass es für $v \neq w \in V$ mit $\|v\| = \|w\|$ und $\langle v, w\rangle \in \mathbb{R}$ einen Einheitsvektor $a \in V$ gibt, sodass $s_a(v) = w$ gilt. Geometrisch bedeuten die Bedingungen $\|v\| = \|w\|$ und $\langle v, w\rangle \in \mathbb{R}$ genau, dass $(v + w) \perp (v - w)$ gilt. Es ist nämlich

$$\langle v + w, v - w\rangle = (\|v\|^2 - \|w\|^2) - (\langle v, w\rangle - \overline{\langle v, w\rangle}),$$

und die letzte Klammer liefert genau den $2i$ mal dem Imaginärteil von $\langle v, w\rangle$. Nach Voraussetzung ist $v - w \neq 0$, also können wir $a := \frac{1}{\|v-w\|}(v - w)$ setzen. Dann ist

$$s_a(v) = v - 2\langle v, a\rangle a = v - 2\tfrac{\langle v, v-w\rangle}{\langle v-w, v-w\rangle}(v - w).$$

Aus $\langle v+w, v-w\rangle = 0$ folgt aber sofort $\langle w, v-w\rangle = -\langle v, v-w\rangle$, und damit $\frac{\langle v,v-w\rangle}{\langle v-w,v-w\rangle} = \frac{1}{2}$, also $s_a(v) = w$. □

Daraus erhalten wir sofort eine Beschreibung von orthogonalen Abbildungen:

SATZ 9.3. *Sei $(V, \langle\ ,\ \rangle)$ ein n–dimensionaler euklidischer Vektorraum und $f : V \to V$ eine orthogonale lineare Abbildung. Dann gibt es eine Zahl $k \leq n$ und Einheitsvektoren $a_1, \dots, a_k \in V$, sodass $f = s_{a_1} \circ \dots \circ s_{a_k}$ gilt. Insbesondere ist* $\det(f) = (-1)^k$.

BEWEIS. Fixieren wir eine Orthonormalbasis $\{v_1, \dots, v_n\}$ von V. Nach Teil (4) von Satz 9.1 ist $\{f(v_1), \dots, f(v_n)\}$ ebenfalls eine Orthonormalbasis für V. Sei j der kleinste Index, sodass $f(v_j) \neq v_j$ gilt. Wegen $\|f(v_j)\| = \|v_j\| = 1$ (und weil wir über $\mathbb{R}$ arbeiten) ist nach dem Lemma $s_{a_1}(f(v_j)) = v_j$, wobei $a_1 = \frac{f(v_j)-v_j}{\|f(v_j)-v_j\|}$.

Für $i < j$ ist $v_i \perp v_j$ und $v_i = f(v_i) \perp f(v_j)$, also auch $v_i = f(v_i) \perp a_1$, also $s_{a_1}(f(v_i)) = f(v_i) = v_i$. Damit gilt aber $(s_{a_1} \circ f)(v_i) = v_i$ für alle $i \leq j$ und $s_{a_1} \circ f$ ist wiederum orthogonal. Induktiv finden wir Einheitsvektoren $a_1, \dots, a_k$, sodass $(s_{a_k} \circ \dots \circ s_{a_1} \circ f)(v_i) = v_i$ für alle $i = 1, \dots, n$. Damit ist diese Komposition aber die Identitätsabbildung, und wegen $(s_{a_j})^{-1} = s_{a_j}$ liefert das $f = s_{a_1} \circ \dots \circ s_{a_k}$. Wegen $\det(s_{a_j}) = -1$ für alle j folgt $\det(f) = (-1)^k$ sofort. □

9.4. Orthogonale Abbildungen in Dimension 2 und 3. Als nächstes wollen wir orthogonale Abbildungen in $\mathbb{R}^2$ und $\mathbb{R}^3$ (jeweils mit dem Standard inneren Produkt) beschreiben. Nach Korollar 9.1 liefert das auch die Beschreibung für beliebige euklidische Vektorräume der Dimensionen zwei und drei.

Sei $f : \mathbb{R}^2 \to \mathbb{R}^2$ orthogonal. Dann ist nach Satz 9.3 entweder f eine Spiegelung, oder f ist eine Komposition von zwei Spiegelungen und $\det(f) = 1$. Ein Einheitsvektor in $\mathbb{R}^2$ hat die Form (x, y) mit $x^2 + y^2 = 1$. Damit gibt es aber einen eindeutigen Winkel $\varphi \in [0, 2\pi)$, sodass $x = \cos(\varphi)$ und $y = \sin(\varphi)$ gilt. Die zwei möglichen Einheitsnormalvektoren zu (x, y) sind nach 8.12 durch $(-y, x)$ und $(y, -x)$ gegeben. Ist nun f orthogonal mit $\det(f) = 1$ und ist $f(e_1) = a = (\cos(\varphi), \sin(\varphi))$, dann muss $f(e_2)$ ein Normalvektor zu $f(e_1)$ sein, der außerdem noch $\det(f(e_1), f(e_2)) = 1$ erfüllt. Damit entspricht f aber der Matrix $\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$, also ist f genau die Drehung um den Winkel φ.

Um andererseits eine Matrixdarstellung für die Spiegelung s_a für $a = (\cos(\varphi), \sin(\varphi))$ zu erhalten, betrachten wir die Rotation f um den Winkel φ. Diese bildet e_1 auf a und e_2 auf ein Element von $a^\perp$ ab. Damit folgt aber sofort, dass $f^{-1} \circ s_a \circ f$ den Basisvektor e_1 auf $-e_1$ und e_2 auf sich selbst abbildet. Damit ist aber die Matrix zu s_a gegeben durch

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix} = \begin{pmatrix} -\cos(2\varphi) & -\sin(2\varphi) \\ -\sin(2\varphi) & \cos(2\varphi) \end{pmatrix}.$$

In drei Dimensionen ist die Situation ähnlich einfach. Eine orthogonale lineare Abbildung $f : \mathbb{R}^3 \to \mathbb{R}^3$ ist entweder eine Spiegelung, oder eine Komposition von zwei Spiegelungen und erfüllt $\det(f) = 1$, oder eine Komposition von drei Spiegelungen. Die wesentliche Beobachtung hier ist aber, dass f mindestens einen reellen Eigenwert besitzen muss, weil jedes Polynom dritten Grades mindestens eine reelle Nullstelle hat. Nach Korollar 9.1 muss dieser Eigenwert gleich 1 oder -1 sein.

Ist v ein Eigenvektor dazu, dann betrachten wir den Teilraum $W = \{v\}^\perp \subset \mathbb{R}^3$. Für $w \in W$ ist $\langle v, f(w)\rangle = \pm\langle f(v), f(w)\rangle = \pm\langle v, w\rangle = 0$, also ist $f(w) \in W$, also W ein f–invarianter Teilraum. Wir können das innere Produkt auf W einschränken und natürlich ist die Einschränkung von f auf W orthogonal bezüglich dieses inneren Produktes. Damit können wir diese Einschränkung aus der Beschreibung der zweidimensionalen orthogonalen Abbildungen von oben ablesen: Ist der Eigenwert von oben gleich 1, dann ist für $\det(f) = 1$ die Abbildung f eine Drehung um die Achse $\{tv : t \in \mathbb{R}\}$ während wir für $\det(f) = -1$ ein Spiegelung an einer Ebene erhalten, die v enthält.

Ist der Eigenwert von oben gleich -1 und ist $\det(f) = 1$, dann muss die Einschränkung von f auf W Determinante -1 haben, also eine Spiegelung sein. Damit erhalten wir aber einen zweidimensionalen Eigenraum zum Eigenwert -1 und einen eindimensionalen Eigenraum zum Eigenwert 1. Geometrisch können wir f als "Spiegelung" an der Geraden interpretieren, die durch den Eigenraum zum Eigenwert Eins beschrieben wird. Im letzten Fall, also Eigenwert gleich -1 und $\det(f) = -1$ muss die Einschränkung von f auf W Determinante Eins haben, also eine Drehung sein. Damit ist f aber die Komposition einer Drehung um die Achse $\{tv : t \in \mathbb{R}\}$ mit der Spiegelung an $W = \{v\}^\perp$. Im Spezialfall des Winkels π erhalten wir $f = -\mathrm{id}$, was wir auch als Punktspiegelung im Nullpunkt interpretieren können.

Bewegungen und euklidische Geometrie

Die Beschreibung von orthogonalen Abbildungen hat direkte Anwendungen in der euklidischen Geometrie. Der richtige Rahmen dafür wäre eigentlich der *euklidische Raum* $\mathbb{E}^n$, den man erhält indem man den affinen Raum $\mathbb{A}^n$ aus 5.2 zusammen mit dem Standard inneren Produkt auf dem Raum $\mathbb{R}^n$ der Verbindungsvektoren betrachtet. Wir werden hier aber der Einfachheit halber auf dem euklidischen Vektorraum $\mathbb{R}^n$ arbeiten.

9.5. Bewegungen. Betrachte man $(\mathbb{R}^n, \langle\ ,\ \rangle)$ als affinen Raum, dann können wir wie in der affinen Geometrie für zwei Punkte $x, y \in \mathbb{R}^n$ den Verbindungsvektor $\overrightarrow{xy} := y - x$ von x nach y bilden. Nun können wir aber auch die Länge $\|\overrightarrow{xy}\|$ dieses Verbindungsvektors betrachten, die nach Definition gerade die Distanz $d(x, y)$ der Punkte x und y ist. Andererseits können wir natürlich auch Winkel messen. Für drei verschiedene Punkte $x, y, z \in \mathbb{R}^n$ definieren wir $\angle(x, y, z)$ als den Winkel zwischen $\overrightarrow{yx}$ und $\overrightarrow{yz}$. Das ist also gerade der Winkel bei y des Dreiecks mit den Ecken x, y und z. Wegen der Polarisierungsformeln kann man aber Winkel aus Distanzen ausrechnen, also sind die Distanzen der wesentliche Begriff. Setzt man nämlich die Parallelogrammidentität in die Polarisierungsformel ein (siehe Satz 8.7), dann erhält man

$$\langle v, w\rangle = \tfrac{1}{2}\left(\|v\|^2 + \|w\|^2 - \|v - w\|^2\right),$$

was sofort $\langle\overrightarrow{yx}, \overrightarrow{yz}\rangle = \frac{1}{2}\left(d(x, y)^2 + d(y, z)^2 - d(x, z)^2\right)$ liefert. Daraus erhalten wir aber $\cos(\angle(x, y, z)) = \frac{d(x,y)^2+d(y,z)^2-d(x,z)^2}{2d(x,y)d(y,z)}$. Damit liegt es nahe, als fundamentalen Abbildungen in der euklidischen Geometrie jene Funktionen zu betrachten, die Distanzen bewahren.

Definition 9.5. Eine *Bewegung* oder eine *Kongruenzabbildung* von $\mathbb{R}^n$ ist eine Funktion $f : \mathbb{R}^n \to \mathbb{R}^n$, die $d(f(x), f(y)) = d(x, y)$ für alle $x, y \in \mathbb{R}^n$ erfüllt.

Obwohl in der Definition einer Bewegung die affine Struktur nicht vorkommt, kann man Bewegungen durch orthogonale lineare Abbildungen beschreiben, denn es gilt

Satz 9.5. *Sei $f : \mathbb{R}^n \to \mathbb{R}^n$ eine Bewegung. Dann gibt es eine orthogonale lineare Abbildung $\overrightarrow{f} : \mathbb{R}^n \to \mathbb{R}^n$, sodass $f(y) = f(x) + \overrightarrow{f}(\overrightarrow{xy})$ für alle $x, y \in \mathbb{R}^n$ gilt.*

Sind umgekehrt $a, b \in \mathbb{R}^n$ beliebig und ist $\overrightarrow{f} : \mathbb{R}^n \to \mathbb{R}^n$ eine orthogonale lineare Abbildung, dann definiert $f(x) := b + \overrightarrow{f}(\overrightarrow{ax})$ eine Bewegung $f : \mathbb{R}^n \to \mathbb{R}^n$.

Beweis. Definiere $\overrightarrow{f} : \mathbb{R}^n \to \mathbb{R}^n$ durch $\overrightarrow{f}(x) := \overrightarrow{f(0)f(x)} = f(x) - f(0)$. Da $x = \overrightarrow{0x}$ gilt, bedeutet diese Gleichung gerade $f(x) = f(0) + \overrightarrow{f}(\overrightarrow{0x})$. Weiters gilt

$$\|\overrightarrow{f}(x) - \overrightarrow{f}(y)\| = \|\overrightarrow{f(0)f(x)} - \overrightarrow{f(0)f(y)}\| = d(f(y), f(x)) = d(y, x) = \|y - x\|.$$

Für $y = 0$ erhalten wir wegen $\overrightarrow{f}(0) = 0$ insbesondere $\|\overrightarrow{f}(x)\| = \|x\|$. Benutzt man die Formel $\langle x, y\rangle = \frac{1}{2}(\|x\|^2 + \|y\|^2 - \|x - y\|^2)$ von oben, dann liefert das $\langle\overrightarrow{f}(x), \overrightarrow{f}(y)\rangle = \langle x, y\rangle$ für alle $x, y \in \mathbb{R}^n$. Insbesondere ist $\{\overrightarrow{f}(e_1), \dots, \overrightarrow{f}(e_n)\}$ eine Orthonormalbasis für $\mathbb{R}^n$. Nach Proposition 8.8 ist $v = \sum_{i=1}^n \langle v, e_i\rangle e_i$ und

$$\overrightarrow{f}(v) = \sum_{i=1}^n \langle\overrightarrow{f}(v), \overrightarrow{f}(e_i)\rangle \overrightarrow{f}(e_i) = \sum_{i=1}^n \langle v, e_i\rangle \overrightarrow{f}(e_i).$$

Daher ist aber $\overrightarrow{f}(x_1, \dots, x_n) = \sum x_i \overrightarrow{f}(e_i)$, also ist $\overrightarrow{f}$ eine lineare Abbildung. Wegen $\overrightarrow{0y} = \overrightarrow{0x} + \overrightarrow{xy}$ und der Linearität von $\overrightarrow{f}$ erhalten wir schließlich

$$f(y) = f(0) + \overrightarrow{f}(\overrightarrow{0y}) = f(0) + \overrightarrow{f}(\overrightarrow{0x}) + \overrightarrow{f}(\overrightarrow{xy}) = f(x) + \overrightarrow{f}(\overrightarrow{xy}).$$

Ist umgekehrt $f(x) = b + \overrightarrow{f}(\overrightarrow{ax})$, dann ist

$$d(f(x), f(y)) = \|b + \overrightarrow{f}(\overrightarrow{ay}) - b - \overrightarrow{f}(\overrightarrow{ax})\| = \|\overrightarrow{f}(\overrightarrow{xy})\| = \|\overrightarrow{xy}\| = d(x, y),$$

also ist f eine Bewegung. □

Bemerkung 9.5. Für $\overrightarrow{f} = \mathrm{id}$ ist die Bewegung $f(v) = b + \overrightarrow{f}(v) = b + v$ einfach eine Translation. Nach dem Satz setzt sich also jede Bewegung aus eine Translation und einer orthogonalen Abbildung zusammen. Eine Bewegung $f : \mathbb{R}^n \to \mathbb{R}^n$ heißt eine *eigentliche Bewegung*, wenn die zugehörige orthogonale lineare Abbildung $\overrightarrow{f}$ in $SO(\mathbb{R}^n)$ liegt, also $\det(\overrightarrow{f}) = 1$ gilt. Ob man mit Bewegungen oder mit eigentlichen Bewegungen arbeitet, hängt davon ab, ob man Spiegelbilder als gleich oder als verschieden betrachten möchte.

9.6. Kongruenzsätze. Wir besprechen hier nur ganz kurz einige Grundzüge der euklidischen Geometrie und die Unterschiede zur affinen Geometrie. Wie in der affinen Geometrie sind auch hier alle Punkte "gleichberechtigt" weil man einen gegebenen Punkt durch Translationen auf jeden anderen Punkt abbilden kann.

In 5.6 haben wir bemerkt, dass es zu zwei Paaren $x_1 \neq y_1$ und $x_2 \neq y_2$ von verschiedenen Punkten in $\mathbb{R}^n$ immer einen affinen Isomorphismus $f : \mathbb{R}^n \to \mathbb{R}^n$ gibt, sodass $f(x_i) = y_i$ für $i = 1, 2$ gilt. Das kann für Bewegungen natürlich nicht mehr stimmen, weil die Distanz der Punkte invariant unter Bewegungen ist. Gilt aber $d(x_1, x_2) = d(y_1, y_2)$, dann haben die Verbindungsvektoren $\overrightarrow{x_1x_2}$ und $\overrightarrow{y_1y_2}$ die gleiche Länge und aus der Diskussion in 9.4 schließt man leicht, dass es eine orthogonale lineare Abbildung $A : \mathbb{R}^n \to \mathbb{R}^n$ gibt, die $A\overrightarrow{x_1x_2} = \overrightarrow{y_1y_2}$ erfüllt, und wir betrachten die Funktion $f(x) := y_1 + A\overrightarrow{x_1x}$, die nach Satz 9.5 eine Bewegung auf $\mathbb{R}^n$ definiert. Diese erfüllt nach Konstruktion $f(x_1) = y_1$ und $f(x_2) = y_1 + A\overrightarrow{x_1x_2} = y_1 + \overrightarrow{y_1y_2} = y_2$. Der Abstand ist also die einzige Invariante eines Paares von Punkten.

Für drei Punkte ist es schon im affinen Fall nötig zwischen *kollinearen Punkten* (die auf einer affinen Geraden liegen) und *Punkten in allgemeiner Lage* (bei denen das nicht der Fall ist) zu unterscheiden. Für kollineare Punkte war die wesentliche affine Invariant das Teilverhältnis, das man jetzt natürlich als das Verhältnis der Abstände betrachten kann. Damit kann man leicht charakterisieren, wann zwei Tripel von kollinearen Punkten durch eine Bewegung aufeinander abgebildet werden können.

Interessanter ist der Fall von Punkten in allgemeiner Lage. Seien $x_i, y_i \in \mathbb{R}^n$ Punkte, sodass sowohl x_1, x_2 und x_3 als auch y_1, y_2 und y_3 nicht kollinear sind. Dann haben wir in 5.6 gesehen, dass es einen affinen Isomorphismus $f : \mathbb{R}^n \to \mathbb{R}^n$ gibt, der $f(x_i) = y_i$ für $i = 1, 2, 3$ erfüllt. Es sind also alle Dreiecke affin äquivalent.

Für Bewegungen kann das natürlich nicht stimmen. Ist nämlich $f : \mathbb{R}^n \to \mathbb{R}^n$ eine Bewegung, sodass $f(x_i) = y_i$ für $i = 1, 2, 3$ gilt, dann muss zumindest $d(x_i, x_j) = d(y_i, y_j)$ für alle Paare $i, j \in \{1, 2, 3\}$ gelten. Die verschiedenen Charakterisierungen für die Existenz solcher Bewegungen sind nun die schon aus der Schule bekannten *Kongruenzsätze* für Dreiecke. Man kann diese Sätze leicht aus der Beschreibung orthogonaler Abbildungen auf $\mathbb{R}^2$ aus 9.4 ableiten. Der Seite–Winkel–Seite Satz besagt zum Beispiel, dass es eine Bewegung f wie oben gibt, falls $d(x_1, x_2) = d(y_1, y_2)$, $d(x_2, x_3) = d(y_2, y_3)$ und $\angle x_1, x_2, x_3 = \angle y_1, y_2, y_3$ gelten. Das bedeutet nämlich, dass $\overrightarrow{x_2x_1}$ und $\overrightarrow{x_2, x_3}$ und $\overrightarrow{y_2y_1}$ und $\overrightarrow{y_2y_3}$ jeweils gleiche Länge haben und den gleichen Winkel einschließen. Betrachtet man die Ebene, die diese beiden Vektoren aufspannen, dann findet man darin leicht eine orthogonale lineare Abbildung A, sodass $A\overrightarrow{x_2x_1} = \overrightarrow{y_2y_1}$ und $A\overrightarrow{x_2x_3} = \overrightarrow{y_2y_3}$ und gemeinsam mit der Identität auf dem orthogonalen Komplement dieser Ebene erhalten wir eine orthogonale Abbildung auf $\mathbb{R}^n$ mit diesen Eigenschaften. Dann ist $f(x) = y_2 + A\overrightarrow{x_2 x}$ die gesuchte Bewegung.

Daraus folgt sofort der Seite–Seite–Seite Satz, der besagt, dass auch $d(x_i, x_j) = d(y_i, y_j)$ für alle i, j hinreichend für die Existenz einer Bewegung f wie oben ist. Man kann nämlich $\angle x_1, x_2, x_3$ aus den paarweisen Distanzen der Punkte ausrechnen und analog für die y_i.

Normale, selbstadjungierte und symmetrische lineare Abbildungen

Wir besprechen nun eine Version von Diagonalisierbarkeit, die den Ideen von Orthogonalität angepasst ist. Überraschenderweise ist Diagonalisierbarkeit in diesem Sinn sehr einfach zu charakterisieren.

9.7. Orthogonale und unitäre Diagonalisierbarkeit. Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer (unitärer) Vektorraum. Eine lineare Abbildung $f : V \to V$ heißt *orthogonal (unitär) diagonalisierbar*, wenn es eine Orthonormalbasis von V gibt, die aus Eigenvektoren von f besteht.

Analog wie in 7.2 ist die orthogonale (unitäre) Diagonalisierbarkeit von f äquivalent dazu, dass es eine Orthonormalbasis $\mathcal{B}$ für V gibt, sodass $[f]_{\mathcal{B}}$ eine Diagonalmatrix ist. Das Konzept der orthogonalen (unitären) Diagonalisierbarkeit macht natürlich auch für Matrizen Sinn, wobei man für $A \in M_n(\mathbb{K})$ einfach die lineare Abbildung $x \mapsto Ax$ auf $\mathbb{K}^n$ betrachtet. Da die Basiswechsel zwischen zwei Orthonormalbasen nach Teil (3) von Satz 9.1 genau den orthogonalen (unitären) Matrizen entsprechen, ist $A \in M_n(\mathbb{K})$ genau dann orthogonal (unitär) diagonalisierbar, wenn es eine orthogonale (unitäre) Matrix $U \in M_n(\mathbb{K})$ gibt (die nach Korollar 9.1 automatisch invertierbar ist), sodass UAU^{-1} eine Diagonalmatrix ist. Man bemerke, dass UAU^{-1} nach Definition für orthogonales U gleich UAU^t und für unitäres U gleich UAU^* ist.

Betrachten wir zunächst den komplexen Fall. Sei $f : V \to V$ unitär diagonalisierbar und $f^* : V \to V$ die adjungierte Abbildung zu f. Ist $\mathcal{B}$ eine Orthonormalbasis für V, die aus Eigenvektoren von f besteht, dann ist $[f]_{\mathcal{B}}$ eine Diagonalmatrix. Nach Satz 8.13 ist dann $[f^*]_{\mathcal{B}} = ([f]_{\mathcal{B}})^*$, also ebenfalls eine Diagonalmatrix, nur sind die Eintragungen auf der Hauptdiagonale gerade die Konjugierten der ursprünglichen Eintragungen. Jedenfalls gilt aber $[f]_{\mathcal{B}}[f^*]_{\mathcal{B}} = [f^*]_{\mathcal{B}}[f]_{\mathcal{B}}$, und damit ist $f \circ f^* = f^* \circ f$. Man nennt lineare Abbildungen mit dieser Eigenschaft *normal*. Man bemerke, dass insbesondere unitäre Abbildungen normal sind, weil sie nach Korollar 9.1 $f^* = f^{-1}$ und damit $f \circ f^* = f^* \circ f = \mathrm{id}$ erfüllen. Überraschenderweise charakterisiert diese Eigenschaft schon die unitär diagonalisierbaren linearen Abbildungen, denn es gilt:

Satz 9.7. *Sei $(V, \langle\ ,\ \rangle)$ ein unitärer Vektorraum. Dann ist eine lineare Abbildung $f : V \to V$ genau dann unitär diagonalisierbar, wenn sie normal ist, also $f \circ f^* = f^* \circ f$ erfüllt.*

Beweis. Wir müssen nur noch zeigen, dass eine normale lineare Abbildung $f : V \to V$ unitär diagonalisierbar ist, und dazu benutzen wir Induktion nach $n := \dim(V)$. Für $n = 1$ ist jede lineare Abbildung unitär diagonalisierbar. Nehmen wir also induktiv an, dass $n \geq 2$ gilt, und der Satz für alle $k < n$ bereits bewiesen wurde. Da $\mathbb{C}$ algebraisch abgeschlossen ist, besitzt f mindestens einen Eigenwert $\lambda \in \mathbb{C}$ und der zugehörige Eigenraum $V_\lambda^f \subset V$ ist ein Teilraum der Dimension ≥ 1. Ist nun $W := (V_\lambda^f)^\perp \subset V$, dann ist $\dim(W) < \dim(V)$.

Wir behaupten, dass der Teilraum W f–invariant ist. Zum Beweis bemerken wir zunächst, dass V_λ^f invariant unter f^* ist. Ist nämlich $v \in V_\lambda^f$, dann ist $f(f^*(v)) =$

$f^*(f(v)) = \lambda f^*(v)$ wegen der Normalität von f, also $f^*(v) \in V_\lambda^f$. Ist nun aber $w \in W$ und $v \in V_\lambda^f$, dann ist $\langle v, f(w)\rangle = \langle f^*(v), w\rangle = 0$, weil $f^*(v) \in V_\lambda^f$ gilt, also gilt $f(w) \in W$. Analog folgt aus $f(V_\lambda^f) \subset V_\lambda^f$, dass $f^*(W) \subset W$ gilt.

Nun können wir das hermitische innere Produkt auf W einschränken und die Einschränkungen von f und f^* auf W betrachten, die jeweils W nach W abbilden. Damit folgt aber leicht, dass $f^*|_W$ die adjungierte Abbildung zu $f|_W$ bezüglich des eingeschränkten inneren Produkts ist. Damit ist $f|_W : W \to W$ normal bezüglich dieses inneren Produkts und nach Induktionsvoraussetzung gibt es eine Orthonormalbasis für W, die aus Eigenvektoren für f besteht. Andererseits gibt es nach Korollar 9.6 eine Orthonormalbasis für V_λ^f. Weil $W = (V_\lambda^f)^\perp$ gilt, ist die Vereinigung der beiden Orthonormalbasen eine Orthonormalbasis für V, also folgt die Behauptung. □

Korollar 9.7. *Sei $(V, \langle\ ,\ \rangle)$ ein n–dimensionaler unitärer Vektorraum, $f : V \to V$ eine unitäre lineare Abbildung. Dann gibt es eine Orthonormalbasis $\{v_1, \dots, v_n\}$ für V und Elemente $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ mit $|\lambda_j| = 1$ für alle j, sodass $f(v_j) = \lambda_j v_j$ für alle $j = 1, \dots, n$ gilt.*

Beweis. Wie wir bereits festgestellt haben ist jede unitäre lineare Abbildung normal, also nach dem Satz unitär diagonalisierbar. Nach Korollar 9.1 hat jeder Eigenwert von f Betrag Eins, also folgt die Behauptung. □

9.8. Selbstadjungierte und symmetrische Abbildungen. Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer (unitärer) Vektorraum. Eine lineare Abbildung $f : V \to V$ heißt *symmetrisch (selbstadjungiert)*, wenn $f^* = f$ gilt.

Eine Matrix $A \in M_n(\mathbb{R})$ heißt *symmetrisch*, wenn $A^t = A$ gilt und eine Matrix $B \in M_n(\mathbb{C})$ heißt *selbstadjungiert* oder *hermitisch*, wenn $B^* = B$ gilt.

Nach Definition ist eine lineare Abbildung $f : V \to V$ also genau dann symmetrisch (selbstadjungiert), wenn $\langle f(v), w\rangle = \langle v, f(w)\rangle$ für alle $v, w \in V$ gilt. Nach Satz 8.13 ist f genau dann symmetrisch (selbstadjungiert) wenn eine (oder äquivalent jede) Matrixdarstellung von f bezüglich einer Orthonormalbasis von V eine symmetrische (selbstadjungierte) Matrix ist.

Natürlich sind selbstadjungierte lineare Abbildungen auf einem komplexen Vektorraum automatisch normal, weil ja $f^* \circ f = f \circ f = f \circ f^*$ gilt. Insbesondere sind also selbstadjungierte lineare Abbildungen nach Satz 9.7 immer unitär diagonalisierbar. Ist aber $v \in V$ ein Eigenvektor für f zum Eigenwert λ, dann ist $\langle f(v), v\rangle = \lambda\langle v, v\rangle$ aber das ist gleich $\langle v, f(v)\rangle = \bar{\lambda}\langle v, v\rangle$. Da v ein Eigenvektor ist, ist $\langle v, v\rangle \neq 0$, also muss $\lambda \in \mathbb{R}$ gelten. Somit erhalten wir

Proposition 9.8. *Sei $(V, \langle\ ,\ \rangle)$ ein unitärer Vektorraum und $f : V \to V$ eine selbstadjungierte lineare Abbildung. Dann ist f unitär diagonalisierbar und alle Eigenwerte von f sind reell.*

Man kann auch leicht direkt sehen, dass zwei Eigenvektoren v, w einer selbstadjungierten linearen Abbildung $f : V \to V$ zu *verschiedenen* Eigenwerten $\lambda \neq \mu$ automatisch orthogonal aufeinander stehen. Es gilt nämlich $\lambda\langle v, w\rangle = \langle f(v), w\rangle = \langle v, f(w)\rangle = \mu\langle v, w\rangle$ (wobei wir im letzten Schritt benutzt haben, dass μ reell ist), also $(\lambda - \mu)\langle v, w\rangle = 0$.

Damit können wir aber nun die orthogonale Diagonalisierbarkeit auf euklidischen Vektorräumen charakterisieren.

SATZ 9.8. *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer Vektorraum. Eine lineare Abbildung $f : V \to V$ ist genau dann orthogonal diagonalisierbar, wenn sie symmetrisch ist, also $f^* = f$ erfüllt.*

BEWEIS. Ist f orthogonal diagonalisierbar, dann gibt es eine Orthonormalbasis $\mathcal{B}$ aus Eigenvektoren. Dann ist $[f]_{\mathcal{B}}$ eine Diagonalmatrix, also insbesondere symmetrisch, also ist $[f^*]_{\mathcal{B}} = [f]_{\mathcal{B}}$, also $f^* = f$.

Ist umgekehrt $f : V \to V$ symmetrisch, dann beweisen wir mit Induktion nach $n = \dim(V)$, dass f diagonalisierbar ist. Für $n = 1$ ist nichts zu zeigen, also nehmen wir $n > 1$ an und der Satz sei für Vektorräume der Dimension $n - 1$ bereits bewiesen. Sei $\mathcal{B}$ eine beliebige Orthonormalbasis für V und $A = [f]_{\mathcal{B}}$. Dann ist A eine symmetrische reelle $n \times n$–Matrix, erfüllt also als komplexe Matrix betrachtet die Gleichung $A^* = A^t = A$. Damit wissen wir aber aus unseren Überlegungen über selbstadjungierte lineare Abbildungen auf unitären Vektorräumen von oben, dass alle Eigenwerte von A reell sind.

Insbesondere hat f mindestens einen reellen Eigenwert a und dazu finden wir einen Eigenvektor v mit $\|v\| = 1$. Sei nun $W = \{v\}^\perp$. Für $w \in W$ ist $\langle f(w), v\rangle = \langle w, f(v)\rangle = a\langle w, v\rangle = 0$, also ist W ein f–invarianter Teilraum. Das innere Produkt schränkt sich zu einem inneren Produkt auf W ein, bezüglich dem f symmetrisch ist. Nach Induktionsvoraussetzung finden wir eine Orthonormalbasis für W, die aus Eigenvektoren für f besteht. Geben wir zu dieser Basis v dazu, dann erhalten wir eine Orthonormalbasis für V aus Eigenvektoren. $\square$

Damit können wir nun auch ein vollständiges Schema angeben, wie man eine symmetrische Matrix $A \in M_n(\mathbb{R})$ orthogonal diagonalisiert. Zunächst bestimmt man die Eigenwerte. Für jeden Eigenraum findet man mit Hilfe des Gauß'schen Algorithmus eine Basis, die man mit Hilfe des Gram–Schmidt Verfahrens orthonormalisiert. Vereinigt man die Orthonormalbasen für die verschiedenen Eigenräume, dann erhält man eine Orthonormalbasis für V, die aus Eigenvektoren für A besteht. Schreibt man diese Basisvektoren als Spaltenvektoren in eine Matrix und invertiert diese, dann erhält man eine orthogonale Matrix U, sodass UAU^t eine Diagonalmatrix ist.

Analog kann man natürlich bei normalen Matrizen in $M_n(\mathbb{C})$ beziehungsweise bei symmetrischen (normalen) linearen Abbildungen auf euklidischen (unitären) Vektorräumen vorgehen.

9.9. Anwendung: Klassifikation von Hyperflächen zweiten Grades. Die erste Anwendung, die wir besprechen wollen kommt aus der euklidischen Geometrie. Eine Funktion zweiten Grades auf $\mathbb{R}^n$ ist eine Funktion $\kappa : \mathbb{R}^n \to \mathbb{R}$ der Form

$$\kappa(x_1, \dots, x_n) = \textstyle\sum_{j=1}^n \alpha_j x_j^2 + \sum_{j<k} \beta_{jk} x_j x_k + \sum_{\ell=1}^n \gamma_\ell x_\ell + \delta,$$

wobei $\alpha_i, \beta_{jk}, \gamma_\ell, \delta \in \mathbb{R}$, und mindestens ein α oder ein β ungleich Null ist. Eine Hyperfläche zweiten Grades in $\mathbb{R}^n$ ist eine Teilmenge der Form $\{v \in \mathbb{R}^n : \kappa(v) = 0\}$, wobei κ eine Funktion zweiten Grades ist.

Wir wollen solche Hyperflächen zweiten Grades durch euklidische Bewegungen in eine einfache Form bringen. Wir besprechen erst die allgemeinen Prinzipien für beliebige Dimension, dann führen wir eine vollständige Klassifikation im Fall $n = 2$ durch, bestimmen also alle *Kurven zweiten Grades*.

Der Schlüssel dazu ist, dass man die Funktion κ von oben mit Hilfe des Standard inneren Produktes schön anschreiben kann: Sei dazu $A = (a_{ij}) \in M_n(\mathbb{R})$ gegeben durch $a_{ii} = \alpha_i$ für $i = 1, \dots, n$, $a_{ij} = \frac{1}{2}\beta_{ij}$ für $i < j$ und $a_{ij} = a_{ji}$ für $i > j$. Somit ist A eine symmetrische Matrix. Weiters sei $b = (\gamma_1, \dots, \gamma_n) \in \mathbb{R}^n$. Dann erhalten wir für

$x = (x_1, \dots, x_n)$

$$\langle Ax, x\rangle + \langle b, x\rangle + \delta = \textstyle\sum_{i=1}^n x_i(\sum_{j=1}^n a_{ij}x_j) + \sum_{k=1}^n \gamma_k x_k + \delta = \kappa(x).$$

Wir wollen nun versuchen, die Funktion

$$(*) \qquad \kappa(x) = \langle Ax, x\rangle + \langle b, x\rangle + \delta$$

durch Translationen und orthogonale Abbildungen zu vereinfachen, um die Nullstellenmenge gut beschreiben zu können. Um den Effekt einer Translation zu bestimmen, berechnen wir

$$\kappa(x - v) = \langle Ax - Av, x - v\rangle + \langle b, x - v\rangle + \delta.$$

Dabei benutzen wir, dass $A^t = A$ und damit $\langle Ax, -v\rangle = \langle -Av, x\rangle$ gilt, und erhalten

$$\langle Ax, x\rangle + \langle b - 2Av, x\rangle + \langle Av, v\rangle - \langle b, v\rangle + \delta = \langle Ax, x\rangle + \langle \tilde{b}, x\rangle + \tilde{\delta},$$

wobei $\tilde{b} = b - 2Av$ und $\tilde{\delta} = \delta - \langle b, v\rangle + \langle Av, v\rangle$. Eine Translation ändert also nichts an der Matrix A, sondern nur b und δ.

Setzen wir $\varphi(x) = Ax$, dann ist $\varphi^* = \varphi$ und nach Satz 8.13 ist $\mathrm{Ker}(\varphi)^\perp = \mathrm{Im}(\varphi^*) = \mathrm{Im}(\varphi)$. Damit kann nach Proposition 8.9 jedes Element $w \in \mathbb{R}^n$ eindeutig in der Form $w = w_0 + w_1$ mit $w_0 \in \mathrm{Ker}(\varphi)$ und $w_1 \in \mathrm{Im}(\varphi)$ geschrieben werden. Natürlich kann man φ auf den Teilraum $\mathrm{Im}(\varphi)$ einschränken und erhält so eine lineare Abbildung $\varphi : \mathrm{Im}(\varphi) \to \mathrm{Im}(\varphi)$. Da $\mathrm{Ker}(\varphi) \cap \mathrm{Im}(\varphi) = \{0\}$ gilt, ist diese Abbildung injektiv, also ein linearer Isomorphismus. Nun zerlegen wir $b = b_0 + b_1$ und $v = v_0 + v_1$ in dieser Weise, und erhalten $\tilde{b} = b_0 + (b_1 - 2Av_1)$ und $\tilde{\delta} = \delta - \langle b_0, v_0\rangle - \langle b_1, v_1\rangle + \langle Av_1, v_1\rangle$. Nun können wir v_1 eindeutig so wählen, dass $2Av_1 = b_1$, also $\tilde{b} = b_0$ gilt. Nachdem v_1 festgelegt ist, stellt sich die Frage, ob man durch gute Wahl von v_0 eine weitere Vereinfachung erreichen kann. Dazu müssen wir zwei Fälle unterscheiden.

Fall 1) $\tilde{b} = b_0 = 0$: Hier hat die Wahl von v_0 keinen Einfluss, wir können also einfach $v_0 = 0$ setzen.

Fall 2) $\tilde{b} = b_0 \neq 0$: In diesem Fall ist $\langle b_0, b_0\rangle \neq 0$, und setzt man

$$v_0 := \frac{\delta - \frac{1}{2}\langle b_1, v_1\rangle}{\langle b_0, b_0\rangle} b_0,$$

dann erreicht man $\tilde{\delta} = 0$.

Wir sehen also, dass man durch eine Translation erreichen kann, dass in der Formel $(*)$ von oben entweder $b = 0$ oder $b \neq 0$, $Ab = 0$ und $\delta = 0$ erreichen kann.

Der zweite Schritt der Normalisierung ist noch einfacher: Wir berechnen einfach in $(*)$ $\kappa(Ux)$ für eine orthogonale Matrix U und erhalten

$$\langle AUx, Ux\rangle + \langle b, Ux\rangle + \delta = \langle U^tAUx, x\rangle + \langle U^tb, x\rangle + \delta.$$

Das ersetzt also nur die Matrix A durch $U^tAU = U^{-1}AU$ und b durch $U^tb = U^{-1}b$, und da $U^{-1}AUU^{-1}b = U^{-1}Ab$ gilt, bleiben die vereinfachenden Annahmen von oben einfach in der gleichen Form gültig. Da $A = A^t$ gilt, können wir die Matrix U nach Satz 9.8 so wählen, dass U^tAU diagonal mit den Eigenwerten von A (in beliebiger Reihenfolge) auf der Hauptdiagonale ist. Davon ausgehend kann man nun in allgemeinen Dimensionen eine Klassifikation von Hyperflächen zweiten Grades durchführen, wir werden uns aber auf den Fall $n = 2$ von *Kurven zweiten Grades* beschränken.

9.10. Der Fall von Kurven zweiten Grades. Wie in 9.9 beginnen wir mit einer Funktion zweiten Grades in der Form von Gleichung $(*)$ aus 9.9, wobei A eine symmetrische Matrix ungleich Null ist. Wir bezeichnen mit α und β die beiden reellen Eigenwerte von A, von denen mindestens einer ungleich Null ist. Wie in 9.9 beschrieben, können wir nun eine passende Translation durchführen, einer der beiden folgenden Fälle eintritt:

Fall 1) $b = 0$
Fall 2) $b \neq 0$, $Ab = 0$ und $\delta = 0$.

Dann können wir A durch eine orthogonale Transformation diagonalisieren, sodass die Hauptdiagonaleinträge α und β sind, und b passend mitbewegen, sodass die beiden Fälle erhalten bleiben. Nun können wir die Klassifikation durchführen, wobei wir eine weitere Unterscheidung nach dem Rang von A vornehmen können. Da $A \neq 0$ vorausgesetzt war, kommt für $\mathrm{rg}(A)$ nur 1 oder 2 in Frage. Da aber im Fall 2 die Matrix A nichttrivialen Kern haben muss, ist in diesem Fall nur $\mathrm{rg}(A) = 1$ möglich.

Fall 1a) $b = 0$, $\mathrm{rg}(A) = 2$: Hier sind α und β beide $\neq 0$ und $(*)$ bekommt die einfache Form

$$\alpha(x_1)^2 + \beta(x_2)^2 + \delta.$$

und wir können noch mit einem beliebigen Faktor $\neq 0$ multiplizieren, ohne die Nullstellenmenge zu ändern.

Fall 1aa) Ist $\delta = 0$, dann multiplizieren wir mit $1/\alpha$ und erhalten die Gleichung $(x_1)^2 + \frac{\beta}{\alpha}(x_2)^2 = 0$. Für $\frac{\beta}{\alpha} > 0$ (d.h. wenn α und β gleiches Vorzeichen haben) Ist das nur der Nullpunkt. Für $\frac{\beta}{\alpha} < 0$ erhält man $x_1 = \pm\sqrt{-\frac{\beta}{\alpha}}x_2$, also zwei Gerade, die sich in Null schneiden. Man erhält also entweder **einen Punkt** oder **zwei schneidende Gerade**.

Fall 1ab) Ist $\delta \neq 0$, dann dividieren wir durch $-\delta$ und erhalten dadurch die Gleichung

$$\frac{-\alpha}{\delta}(x_1)^2 + \frac{-\beta}{\delta}(x_2)^2 = 1.$$

Sind die Faktoren $\frac{-\alpha}{\delta}$ und $\frac{-\beta}{\delta}$ beide negativ, dann beschreibt das die **leere Menge**.

Ist $\frac{-\alpha}{\delta} > 0$ und $\frac{-\beta}{\delta} < 0$, dann setzt man $c = \sqrt{\frac{\delta}{-\alpha}}$ und $d = \sqrt{\frac{\delta}{\beta}}$ und erhält die Gleichung

$$\left(\frac{x_1}{c}\right)^2 - \left(\frac{x_2}{d}\right)^2 = 1,$$

die eine **Hyperbel** in erster Hauptlage mit Hauptachsenlänge c und Nebenachsenlänge d beschreibt. Analog erhält man für die umgekehrte Konfiguration der Vorzeichen eine Hyperbel in zweiter Hauptlage. Sind schließlich beide Faktoren positiv, dann erhält man analog die Gleichung

$$\left(\frac{x_1}{c}\right)^2 + \left(\frac{x_2}{d}\right)^2 = 1$$

und damit eine **Ellipse** mit Achsenlängen c und d.

Fall 1b) $b = 0$, $\mathrm{rg}(A) = 1$: In diesem Fall können wir $\alpha \neq 0$ und $\beta = 0$ annehmen und erhalten die einfache Funktion zweiten Grades

$$\alpha(x_1)^2 + \delta.$$

Wieder können wieder mit einem Faktor $\neq 0$ multiplizieren ohne die Nullstellenmenge zu ändern. Für $\delta = 0$ dividieren wir durch α und erhalten $(x_1)^2 = 0$, und damit **eine Gerade** (nämlich die x_2–Achse, die eigentlich doppelt gezählt werden sollte) als Nullstellenmenge.

Für $\delta \neq 0$ dividieren wir durch $-\delta$ und erhalten $\frac{-\alpha}{\delta}(x_1)^2 = 1$. Das liefert entweder (für $\frac{\alpha}{\delta} > 0$) die **leere Menge** oder (für $\frac{\alpha}{\delta} < 0$) $x_1 = \pm\sqrt{-\frac{\delta}{\alpha}}$, also **zwei parallele Gerade**.

Fall 2): $b \neq 0$, $Ab = 0$, $\delta = 0$: Wie schon bemerkt muss hier $\mathrm{rg}(A) = 1$ gelten, also können wir $\alpha \neq 0$ und $\beta = 0$ annehmen. Wegen $Ab = 0$ muss nun b ein Vielfaches von e_2 sein, sagen wir $b = \lambda e_2$. Damit erhalten wir die einfache Funktion zweiten Grades

$$\alpha(x_1)^2 + \lambda x_2.$$

Wir dürfen wieder durch einen beliebigen Faktor $\neq 0$ dividieren. In diesem Fall ist die übliche Konvention durch $-\frac{\lambda}{2}$ zu dividieren, was zur Gleichung $\frac{-2\alpha}{\lambda}(x_1)^2 = 2x_2$ und damit zu einer **Parabel** führt.

9.11. Anwendung: Klassifikation von symmetrischen Bilinear– und Sesquilinearformen. Um beide Arten von Formen gleichzeitig betrachten zu können, arbeiten wir wieder mit Sesquilinearformen und definieren für $\mathbb{K} = \mathbb{R}$ die Konjugation als Identität. Sei V ein endlichdimensionaler Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und seien $b, \tilde{b} : V \times V \to \mathbb{K}$ zwei Sesquilinearformen. Wir nennen die Formen b und $\tilde{b}$ *äquivalent*, wenn es einen linearen Isomorphismus $f : V \to V$ gibt, sodass $\tilde{b}(v, w) := b(f(v), f(w))$ gilt. Das definiert eine Äquivalenzrelation auf der Menge aller Sesquilinearformen auf V. Wir wollen entscheiden, wann zwei Formen in diesem Sinne äquivalent sind. Dazu suchen wir eine Normalform für solche Formen, d.h. wir suchen eine Basis, in der die Form möglichst einfach aussieht. Das kann man auf eine Frage über selbstadjungierte Matrizen zurückführen.

Der erste wesentliche Begriff an dieser Stelle ist der *Rang* einer Sesquilinearform. Haben wir $b : V \times V \to \mathbb{K}$ gegeben, dann definieren wir den *Nullraum* $\mathcal{N}$ von b durch $\mathcal{N} := \{v \in V : \forall w \in V : b(v, w) = 0\}$. Das ist offensichtlich ein Teilraum von V und wir definieren den *Rang* von b als $\mathrm{rg}(b) := \dim(V) - \dim(\mathcal{N})$. Nach Definition ist b genau dann nicht–degeneriert, wenn $\mathcal{N} = \{0\}$ bzw. $\mathrm{rg}(b) = \dim(V)$ gilt.

Sind b und $\tilde{b}$ äquivalent und ist $f : V \to V$ der entsprechende lineare Isomorphismus, dann bildet f den Nullraum von b isomorph auf den Nullraum von $\tilde{b}$ ab. Daher haben b und $\tilde{b}$ den gleichen Rang, also ist der Rang eine Invariante der Äquivalenzklasse von b.

Ist $b : V \times V \to \mathbb{K}$ eine Sesquilinearform und $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis für V, dann definiert man die *Gram Matrix* $A = (a_{ij})$ von b bezüglich $\mathcal{B}$ durch $a_{ij} := b(v_j, v_i) = \overline{b(v_i, v_j)}$ (der Grund für die Konjugation wird gleich klar werden). Weil b sesquilinear ist, gilt $a_{ji} = \overline{a_{ij}}$, also $A^* = A$. Durch Anwenden eines linearen Isomorphismus, genügt es, Formen auf $V = \mathbb{K}^n$ zu klassifizieren. Ist b so eine Form, dann können wir die Gram Matrix $A = (a_{ij})$ bezüglich der Standardbasis bilden. Für $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n) \in \mathbb{K}^n$ ist $x = \sum x_i e_i$ und $y = \sum y_j e_j$, also erhalten wir wegen der Sesquilinearität von b die Gleichung

$$b(x, y) = b(\textstyle\sum x_i e_i, \sum y_j e_j) = \sum_{i,j} x_i \overline{y_j} b(e_i, e_j) = \sum_i x_i (\sum_j \overline{a_{ij} y_j}) = \langle x, Ay\rangle = \langle Ax, y\rangle,$$

wobei wir im letzten Schritt benutzt haben, dass A selbstadjungiert ist. Damit können wir sofort den Rang von b in Termen von A verstehen. Für $x \in \mathbb{K}^n$ gilt $0 = b(x, y) = \langle Ax, y\rangle$ natürlich genau dann für alle $y \in \mathbb{K}^n$, wenn $Ax = 0$ gilt. Damit ist aber der Nullraum $\mathcal{N}$ von b genau der Kern von A, also $\mathrm{rg}(b) = \mathrm{rg}(A)$. Insbesondere ist b genau dann nicht entartet, wenn die Matrix A invertierbar ist.

Umgekehrt definiert für jede Matrix $A = (a_{ij}) \in M_n(\mathbb{K})$ natürlich $\beta(x, y) := \langle x, Ay\rangle$ eine reelle Bilinearform β auf $\mathbb{K}^n$ mit $a_{ij} = \overline{\beta(e_i, e_j)}$ und wegen $\langle y, Ax\rangle = \overline{\langle Ax, y\rangle} = \overline{\langle x, A^* y\rangle}$ ist β genau dann eine Sesquilinearform, wenn $A = A^*$ gilt (also für $\mathbb{K} = \mathbb{R}$

die Matrix symmetrisch ist). Damit erhalten wir eine Bijektion zwischen der Menge der Sesquilinearformen und der Menge der selbstadjungierten Matrizen.

Jeder lineare Isomorphismus $f : \mathbb{R}^n \to \mathbb{R}^n$ ist von der Form $x \mapsto Tx$ für eine invertierbare Matrix $T \in M_n(\mathbb{R})$. Damit ist aber $b(Tx, Ty) = \langle Tx, ATy\rangle = \langle x, T^*ATy\rangle$. Damit sind aber die Sesquilinearformen zu Hermite'schen Matrizen $A, B \in M_n(\mathbb{R})$ genau dann äquivalent, wenn es eine invertierbare Matrix T gibt, sodass $B = T^*AT$ gilt. (Beachte, dass mit T auch T^* invertierbar ist, und $(T^*)^{-1} = (T^{-1})^*$ gilt.)

Satz 9.11 (Trägheitssatz von Sylvester). *Sei V ein n–dimensionaler Vektorraum über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$ und sei $b : V \times V \to \mathbb{R}$ eine Sesquilinearform. Dann gibt es eindeutig bestimmte Zahlen $p, q \in \mathbb{N}$ mit $p+q = \operatorname{rg}(b) \leq n$ sowie eine Basis $\{v_1, \dots, v_n\}$ für V, sodass $b(v_i, v_j) = 0$ für $i \neq j$, und $b(v_i, v_i)$ gleich 1 für $1 \leq i \leq p$, gleich -1 für $p+1 \leq i \leq p+q$ und gleich 0 für $i > p+q$ gilt. Man sagt dann, b hat Signatur (p, q).*

Zwei Sesquilinearformen auf V sind genau dann äquivalent, wenn sie die gleiche Signatur (und damit insbesondere den gleichen Rang) haben.

Beweis. Wie oben erklärt dürfen wir $V = \mathbb{K}^n$ annehmen. Sei $A \in M_n(\mathbb{K})$ die Gram Matrix zu b. Da A selbstadjungiert ist, finden wir nach Proposition 9.8 eine Orthonormalbasis $\{\tilde{v}_1, \dots, \tilde{v}_n\}$ aus Eigenvektoren für A und alle Eigenwerte sind reell. Wir ordnen die Vektoren so an, dass erst Eigenvektoren zu positiven Eigenwerten, dann solche zu negativen Eigenwerten und am Ende solche zum Eigenwert Null kommen. Sind p und q die Anzahlen der positiven bzw. negativen Eigenwerte (mit Vielfachheit), dann ist $A\tilde{v}_i = \lambda_i \tilde{v}_i$ mit $\lambda_i > 0$ für $i = 1, \dots, p$, $\lambda_i < 0$ für $i = p+1, \dots, p+q$ und $\lambda_i = 0$ für $i > p+q$. Bemerke, dass $p+q$ gleich dem Rang von A, also dem Rang von b ist. Nach Konstruktion ist $b(\tilde{v}_i, \tilde{v}_j) = \langle A\tilde{v}_i, \tilde{v}_j\rangle = \lambda_i \langle \tilde{v}_i, \tilde{v}_j\rangle$, also gleich 0 für $i \neq j$ und gleich λ_i für $i = j$. Setzt man nun $v_i = \frac{1}{\sqrt{|\lambda_i|}}\tilde{v}_i$ für $1 \leq i \leq p+q$ und $v_i = \tilde{v}_i$ für $i \geq p+q$, dann hat die Basis $\{v_1, \dots, v_n\}$ die gewünschten Eigenschaften.

Zur Eindeutigkeit von p und q haben wir bereits bemerkt, dass $p+q$ gleich dem Rang von b ist, also genügt es zu zeigen, dass p eindeutig bestimmt ist. Sei also $\{w_1, \dots, w_n\}$ eine weitere Basis für V, sodass $b(w_i, w_j) = 0$ für $i \neq j$ und $b(w_i, w_i) > 0$ für $1 \leq i \leq p'$ und $b(w_i, w_i) \leq 0$ für $i > p'$. Dann behaupten wir, dass die Menge $\{v_1, \dots, v_p, w_{p'+1}, \dots, w_n\}$ linear unabhängig ist. Ist nämlich $\sum_{i=1}^p \lambda_i v_i + \sum_{j=p'+1}^n \mu_j w_j = 0$, dann ist $\sum_{i=1}^p \lambda_i v_i = -\sum_{j=p'+1}^n \mu_j w_j$, und damit

$$b(\textstyle\sum_{i=1}^p \lambda_i v_i, \sum_{i=1}^p \lambda_i v_i) = b(\sum_{j=p'+1}^n \mu_j w_j, \sum_{j=p'+1}^n \mu_j w_j).$$

Nach Konstruktion liefert aber die linke Seite $\sum_{i=1}^p |\lambda_i|^2 \geq 0$ und die rechte Seite $\sum_{j=p'+1}^n |\mu_j|^2 b(w_j, w_j) \leq 0$. Damit kann Gleichheit nur gelten, wen beide Seiten gleich Null sind. Daraus folgt $\lambda_i = 0$ für alle i, und wegen der linearen Unabhängigkeit von $\{w_{p'+1}, \dots, w_n\}$ müssen auch alle μ_j Null sein. Damit muss aber $p + (n - p') \leq n$, also $p \leq p'$ gelten. Vertauscht man die Rollen der beiden Basen, dann folgt $p' \leq p$, also $p = p'$. Damit ist die Signatur wohldefiniert, und zwei Bilinearformen können nur dann äquivalent sein, wenn sie die gleiche Signatur haben.

Sind umgekehrt b und $\tilde{b}$ symmetrische Bilinearformen mit der gleichen Signatur, dann finden wir entsprechende Basen $\{v_1, \dots, v_n\}$ und $\{w_1, \dots, w_n\}$. Ist $f : V \to V$ die eindeutige lineare Abbildung, die $f(w_i) = v_i$ für alle i erfüllt, dann gilt einerseits $b(f(w_i), f(w_j)) = b(v_i, v_j) = \tilde{b}(w_i, w_j)$ für alle i, j. Andererseits ist für $v = \sum x_i w_i$ das Bild $f(v) = \sum x_i v_i$ und daraus folgt sofort $\tilde{b}(f(v), f(w)) = b(v, w)$ für alle $v, w \in V$. $\square$

Bemerkung 9.11. Man kann auch komplexe Bilinearformen auf komplexen Vektorräumen betrachten und klassifizieren. An sich ist die Klassifikation einfacher als im reellen Fall (bzw. als bei Sesquilinearformen), der Beweis ist aber etwas aufwändiger, weil wir die Resultate über unitäre Diagonalisierbarkeit nicht ins Spiel bringen können. Deshalb skizzieren wir nur kurz das Resultat und den Beweis (der auch als elementare Beweismethode für den Trägheitssatz verwendet werden kann).

Der Rang macht natürlich auch für komplexe Bilinearformen Sinn und äquivalente komplexe Bilinearformen müssen den gleichen Rang haben. Im komplexen Fall ist der Rang aber die einzige Invariante, d.h. zwei Bilinearformen sind genau dann äquivalent, wenn sie den gleichen Rang haben. Genauer gesagt, findet man für eine komplexe Bilinearform $b : V \times V \to \mathbb{C}$ mit $\operatorname{rg}(b) = r \leq n = \dim(V)$ immer eine Basis $\{v_1, \dots, v_n\}$ für V, die $b(v_i, v_j) = 0$ für $i \neq j$, $b(v_i, v_i) = 1$ für $1 \leq i \leq r$ und $b(v_i, v_i) = 0$ für $i > r$ erfüllt.

Zum Beweis bemerkt man zunächst, dass eine komplexe Bilinearform $b : V \times V \to \mathbb{C}$, die $b(v, v) = 0$ für alle $v \in V$ erfüllt, identisch Null sein muss. Man erhält dann nämlich für alle $v, w \in V$ auch $0 = b(v + w, v + w) = b(v, v) + 2b(v, w) + b(w, w) = 2b(v, w)$. Ist also $b \neq 0$, dann findet man einen Vektor $\tilde{v} \in V$ mit $b(\tilde{v}, \tilde{v}) = z \neq 0$. Da wir über $\mathbb{C}$ arbeiten, finden wir eine Zahl $\lambda \in \mathbb{C}$, sodass $\lambda^2 = \frac{1}{z}$ gilt (schreibe $z = re^{i\varphi}$ und setze $\lambda := r^{-1/2}e^{-i\varphi/2}$). Setzt man nun $v = \lambda\tilde{v}$ dann gilt $b(v, v) = \lambda^2 z = 1$.

Damit beweist man im nächsten Schritt, die Behauptung im Fall $\operatorname{rg}(b) = \dim(V) = n$ induktiv. Der Induktionsanfang ist schon erledigt, weil wir immer ein Element v mit $b(v, v) = 1$ finden. Für $n > 1$ wählt man ein Element $v_1 \in V$ mit $b(v_1, v_1) = 1$ und betrachtet $V_1 := \{v \in V : b(v, v_1) = 0\}$. Man sieht leicht, dass $\dim(V_1) = n - 1$ gilt und damit folgt $V = V_1 \oplus \mathbb{C} \cdot v_1$. Aus dieser Summendarstellung folgt leicht, dass die Einschränkung $b : V_1 \times V_1 \to \mathbb{C}$ von b auf V_1 nicht degeneriert ist. Wendet man darauf die Induktionsvoraussetzung an, dann erhält man sofort eine passende Basis für V.

Für $r = \operatorname{rg}(b) < n$, betrachtet man den Nullraum $\mathcal{N}$ von b und den Quotientenraum $V/\mathcal{N}$. Man sieht leicht, dass b eine komplexe Bilinearform $\underline{b}$ auf $V/\mathcal{N}$ induziert, die durch $\underline{b}(v + \mathcal{N}, w + \mathcal{N}) := b(v, w)$ gegeben ist. Aus der Konstruktion folgt leicht, dass $\underline{b}$ nicht degeneriert ist. Aus dem letzten Schritt erhalten wir eine Basis für $V/\mathcal{N}$, und man wählt Urbilder $v_1, \dots, v_r$ für die Basiselemente. Wählt man dann noch eine Basis $\{v_{r+1}, \dots, v_n\}$ für $\mathcal{N}$, dann sieht man leicht, dass $\{v_1, \dots, v_n\}$ eine Basis für V ist, und nach Konstruktion hat b auf dieser Basis die richtige Form. Den Zusammenhang der einfachen Basen mit der Äquivalenz erhält man genau wie im Fall der Sesquilinearformen.

9.12. Positiv definite Matrizen. Wie wir in 9.11 gesehen haben, definiert für eine selbstadjungierte Matrix $A \in M_n(\mathbb{K})$ der Ausdruck $b(x, y) := \langle Ax, y\rangle = \langle x, Ay\rangle$ eine Sesquilinearform auf $\mathbb{K}^n$.

Definition 9.12. Eine Matrix $A \in M_n(\mathbb{K})$ heißt *positiv (semi–)definit* wenn A selbstadjungiert ist und die Sesquilinearform $b(x, y) = \langle Ax, y\rangle$ positiv (semi–)definit ist.

Wir können nun ganz leicht einige Charakterisierungen von positive semidefiniten Matrizen angeben.

Proposition 9.12. *Für eine Matrix $A \in M_n(\mathbb{K})$ sind äquivalent:*

(1) *A ist positiv semidefinit (positiv definit).*
(2) *A ist selbstadjungiert und alle Eigenwerte von A sind nicht–negativ (positiv).*
(3) *Es gibt eine (invertierbare) Matrix $B \in M_n(\mathbb{K})$ sodass $A = B^*B$ gilt.*

Beweis. (1)$\Rightarrow$(2): Ist A positiv semidefinit, dann ist A nach Definition selbstadjungiert, also unitär diagonalisierbar. Ist v ein Eigenvektor von A zum (reellen) Eigenwert

λ, dann ist $\lambda\langle v,v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = b(v,v) \geq 0$, also $\lambda \geq 0$ und $\lambda > 0$ im positiv definiten Fall.

(2)$\Rightarrow$(3): Ist A selbstadjungiert mit nicht–negativen Eigenwerten, dann ist $A = UDU^*$ für eine unitäre Matrix U und eine Diagonalmatrix D mit reellen, nicht–negativen Eintragungen $\lambda_1, \dots, \lambda_n$. Definiert man $\tilde{D}$ als die Diagonalmatrix mit Eintragungen $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$, dann gilt natürlich $\tilde{D}^2 = D$. Nun setzen wir $B := U\tilde{D}U^*$ und erhalten $B^* = U\tilde{D}^*U^* = U\tilde{D}U^*$. Da U unitär ist, gilt $U^* = U^{-1}$ und damit folgt $B^*B = U\tilde{D}^2U^* = A$. Sind alle Eigenwerte von A positiv, dann ist $\tilde{D}$ offensichtlich invertierbar und damit ist auch B invertierbar.

(3)$\Rightarrow$(1): Ist $A = B^*B$, dann ist $A^* = B^*B^{**} = A$. Außerdem ist die Bilinearform zu A gegeben durch $b(x,y) = \langle x, B^*By\rangle = \langle Bx, By\rangle$. Damit folgt $b(x,x) = \langle Bx, Bx\rangle \geq 0$ sofort. Für invertierbares B ist $b(x,x) = 0$ nur für $x = 0$ möglich. □

Die Matrix B in Teil (3) ist bei weitem nicht eindeutig. Tatsächlich kann man relativ einfach sehen, dass man für B eine spezielle Form annehmen kann. Benutzt man nämlich die QR–Zerlegung aus Satz 9.2, dann kann man $B = QR$ für eine unitäre (bzw. orthogonale) Matrix Q und eine obere Dreiecksmatrix R mit positiven Eintragungen auf der Hauptdiagonale schreiben. Dann ist aber $B^* = R^*Q^*$ und weil Q unitär ist, folgt $B^*B = R^*R$. Somit kann man jede positiv semidefinite Matrix A als R^*R für eine obere Dreiecksmatrix R mit positiven Eintragungen auf der Hauptdiagonale schreiben. Es zeigt sich, dass für eine positiv definite Matrix A die Matrix R eindeutig bestimmt ist ("Cholesky–Zerlegung von A").

Es gibt noch eine weitere Charakterisierung von positiv definiten Matrizen, die relativ einfach zu überprüfen ist, aber auf den ersten Blick etwas unnatürlich aussieht. Wir haben in 6.11 den Begriff eines Minors einer Matrix A kennen gelernt. Das ist einfach eine Matrix, die entsteht, indem man in A einige Zeilen und Spalten streicht. Die einfachsten Minoren sind die sogenannten *Hauptminoren* einer Matrix $A \in M_n(\mathbb{K})$. Das sind die $k \times k$–Teilmatrizen von A, die aus den ersten k Zeilen und Spalten von A bestehen. Ist $A = (a_{ij})$, dann ist der $k$te Hauptminor einfach die $k \times k$–Matrix mit Eintragungen a_{ij} für $1 \leq i, j \leq k$.

Wir benötigen noch eine Beobachtung. Für eine selbstadjungierte Matrix A sind auch die Hauptminoren selbstadjungiert (im Gegensatz zu allgemeinen Minoren). Außerdem ist die Determinante einer selbstadjungierten Matrix immer reell. Das kann man entweder aus Satz 6.9 folgern: Wegen $A = A^* = (\overline{A})^t$ gilt natürlich $\det(A) = \det(\overline{A}) = \overline{\det(A)}$, wobei die letzte Gleichheit aus der Verträglichkeit der Konjugation mit allen Körperoperationen folgt. Alternativ kann man A diagonalisieren und dann $\det(A)$ als die Determinante der entsprechenden Diagonalmatrix berechnen. Damit ist $\det(A)$ das Produkt aller Eigenwerte von A, also reell und für eine positiv definite Matrix positiv.

Satz 9.12 (Hauptminorenkriterium). *Eine selbstadjungierte Matrix A ist genau dann positiv definit, wenn alle Hauptminoren von A positive Determinante haben.*

Beweis. Der Schlüssel zum Beweis ist es, die scheinbare Unnatürlichkeit der Bedingung an die Hauptminoren (die ja von der gewählten Basis und sogar von der Reihenfolge der Basiselemente abhängen) "loszuwerden". Betrachten wir dazu eine Sesquilinearform b auf einem Vektorraum V und die Gram Matrix A von b bezüglich einer Basis von V (siehe 9.11). Bei Wahl einer anderen Basis erhält man eine Gram Matrix der Form T^*AT für eine invertierbare Matrix T. Nun ist aber $\det(T^*AT) = \det(A)\overline{\det(T)}\det(T) = \det(A)|\det(T)|^2$ und das ist genau dann positiv, wenn $\det(A)$ positiv ist. Somit ist "positive Determinante der Gram Matrix" eine wohldefinierte Eigenschaft für Bilinearformen.

Wir behaupten nun, dass eine nicht–degenerierte Sesquilinearform b auf einem n–dimensionalen Vektorraum V genau dann positiv definit ist, wenn es einen Teilraum $W \subset V$ der Dimension $n-1$ gibt, sodass die Einschränkung $b : W \times W \to \mathbb{K}$ positiv definit ist und die Gram Matrix von b positive Determinante hat. Die eine Implikation ist klar, weil jede Einschränkung einer positiv definiten Sesquilinearform wieder positiv definit ist und eine positiv definite Matrix positive Determinante hat, interessant ist die Umkehrung: Dazu wählen wir einen Vektor $\tilde{v} \in V \setminus W$ und betrachten die lineare Abbildung $W \to \mathbb{K}$, die gegeben ist durch $w \mapsto b(w, \tilde{v})$. Da b auf W positiv definit ist, finden wir nach Satz 8.11 einen Vektor $w_0 \in W$, sodass $b(w, \tilde{v}) = b(w, w_0)$ für alle $w \in W$ gilt. Setzt man $v := \tilde{v} - w_0$, dann ist nach Konstruktion $b(w, v) = 0$ für alle $w \in W$, aber $v \neq 0$, wegen $\tilde{v} \notin W$. Damit spannt aber v einen zu W komplementären Teilraum von V auf. Wäre $b(v, v) = 0$, dann wäre v im Nullraum von b, also b degeneriert, also ist $b(v, v) \neq 0$.

Wählt man nun eine Basis $\{w_1, \dots, w_{n-1}\}$ für W, dann ergänzt v diese zu einer Basis für V. Die Gram Matrix von b bezüglich dieser Basis hat die Form $\begin{pmatrix} B & 0 \\ 0 & b(v,v) \end{pmatrix}$, wobei B die Gram Matrix der Einschränkung von b auf W ist und damit insbesondere positive Determinante hat. Damit ist aber die Determinante der Gram Matrix genau dann positiv, wenn $b(v, v) > 0$ gilt, was natürlich äquivalent zur positiven Definitheit von b ist.

Betrachten wir nun eine selbstadjungierte Matrix $A \in M_n(\mathbb{K})$ mit zugehöriger Sesquilinearform b. Die Hauptminoren von A sind dann gerade die Matrizen zu den Einschränkungen von b auf die Teilräume $\mathbb{K}^k \subset \mathbb{K}^n$ (als erste k–Komponenten). Ist A positiv definit, dann stimmt das auch für alle diese Einschränkungen, also sind die Determinanten der Hauptminoren positiv.

Haben umgekehrt alle Hauptminoren positive Determinante, dann sind alle diese Einschränkungen nicht degeneriert. Wegen $a_{11} > 0$ ist die Einschränkung auf $\mathbb{K}^1$ positiv definit. Damit folgt aber nach unserem Kriterium positive Definitheit der Einschränkung auf $\mathbb{K}^2$ aus der Positivität der Determinante des zweiten Hauptminors und so weiter. □

Die Charakterisierung über die Determinanten der Hauptminoren ist im reellen Fall folgendermaßen leicht explizit zu überprüfen: Hat man eine symmetrische Matrix A gegeben, dann bringt man die Matrix durch elementare Zeilenoperationen auf obere Dreiecksgestalt, wobei man aber darauf achtet, dass man bei den Operationen, in denen eine Zeile mit einem Skalar multipliziert wird, nur positive Faktoren verwendet (was natürlich möglich ist). Die Hauptminoren der resultierenden Dreiecksmatrix entstehen aus den Hauptminoren von A durch elementare Zeilenoperationen, die so gewählt sind, dass das Vorzeichen der Determinante unverändert bleibt. Somit ist A genau dann positiv definit, wenn alle Diagonaleintragungen der resultierenden Dreiecksmatrix positiv sind.

9.13. Die Singulärwertzerlegung. Wir kommen nur zu einer Zerlegung für rechteckige Matrizen, die vor allem in der angewandten Mathematik ziemlich bedeutend ist. Wir werden die Zerlegung aus der Existenz von Orthonormalbasen ableiten, die einer allgemeinen linearen Abbildung angepasst sind. Betrachten wir also zwei endlichdimensionale euklidische oder unitäre Vektorräume V und W (und bezeichnen wir beide inneren Produkte einfach mit $\langle\ ,\ \rangle$) und eine lineare Abbildung $f : V \to W$.

Satz 9.13. *Sei $f : V \to W$ eine lineare Abbildung zwischen zwei endlichdimensionalen euklidischen bzw. unitären Vektorräumen mit Rang r. Dann gibt es Orthonormalbasen $\{v_1, \dots, v_n\}$ für V und $\{w_1, \dots, w_m\}$ für W sowie positive reelle Zahlen $\sigma_1, \dots, \sigma_r$, sodass $f(v_i) = \sigma_i w_i$ für $i = 1, \dots, r$ und $f(v_i) = 0$ für $i > r$ gilt.*

Die adjungierte Abbildung f^* von f ist dann durch $f^*(w_i) = \sigma_i v_i$ für $i = 1, \dots r$ und $f^*(w_i) = 0$ für $i > r$ gegeben.

Beweis. Sei $f^* : W \to V$ die adjungierte Abbildung zu f. Dann gilt nach Definition $\langle f(v), w\rangle = \langle v, f^*(w)\rangle$ für alle $v \in V$ und $w \in W$. Nach Proposition 9.12 sind dann $f^* \circ f : V \to V$ und $f \circ f^* : W \to W$ positiv semidefinite Abbildung, d.h. ihre Matrixdarstellungen bezüglich beliebiger Orthonormalbasen sind positiv semidefinit. Nach Proposition 8.9 können wir V als direkte Summe $V = \mathrm{Ker}(f) \oplus \mathrm{Ker}(f)^\perp$ zerlegen und analog ist $W = \mathrm{Im}(f) \oplus \mathrm{Im}(f)^\perp$. Klarerweise schränkt sich f zu einem linearen Isomorphismus $\mathrm{Ker}(f)^\perp \to \mathrm{Im}(f)$ ein. Andererseits ist nach Satz 8.13 $\mathrm{Im}(f)^\perp = \mathrm{Ker}(f^*)$ (also $\mathrm{Im}(f) = \mathrm{Ker}(f^*)^\perp$) und $\mathrm{Ker}(f)^\perp = \mathrm{Im}(f^*)$. Somit induziert f^* einen linearen Isomorphismus $\mathrm{Im}(f) \to \mathrm{Ker}(f)^\perp$ und als Abbildungen auf $\mathrm{Ker}(f)^\perp$ bzw. auf $\mathrm{Im}(f)$ sind $f^* \circ f$ und $f \circ f^*$ die Kompositionen dieser Isomorphismen, also sogar positiv definit.

Nach Proposition 9.12 gibt es eine Orthonormalbasis $\{v_1, \dots, v_r\}$ für $\mathrm{Ker}(f)^\perp$ die aus Eigenvektoren für $f^* \circ f$ besteht. Die entsprechenden Eigenwerte $\lambda_1, \dots, \lambda_r$ sind positiv und wir setzen $\sigma_i := \sqrt{\lambda_i}$ für $i = 1, \dots, r$. Weiters definieren wir $w_i = \frac{1}{\sigma_i} f(v_i)$ für $i = 1, \dots, r$. Dann gilt nach Konstruktion $f^*(w_i) = \frac{1}{\sigma_i} f^*(f(v_i)) = \frac{\lambda_i}{\sigma_i} v_i = \sigma_i v_i$. Andererseits rechnen wir

$$\langle w_i, w_j\rangle = \tfrac{1}{\sigma_i\sigma_j}\langle f(v_i), f(v_j)\rangle = \tfrac{1}{\sigma_i\sigma_j}\langle v_i, f^*(f(v_j))\rangle = \tfrac{\lambda_j}{\sigma_i\sigma_j}\langle v_i, v_j\rangle.$$

Das bedeutet aber gerade, dass die w_i ein Orthonormalsystem in $\mathrm{Im}(f)$ bilden, und weil $\dim(\mathrm{Im}(f)) = r$ ist, bilden sie sogar eine Orthonormalbasis für diesen Teilraum. Nun müssen wir nur noch beliebige Orthonormalbasen $\{v_{r+1}, \dots, v_n\}$ für $\mathrm{Ker}(f)$ und $\{w_{r+1}, \dots, w_m\}$ für $\mathrm{Im}(f)^\perp$ wählen um den Beweis zu vervollständigen. □

Die Zahlen σ_i heißen die *Singulärwerte* von f, sie sind nach Konstruktion genau die Wurzeln aus den Eigenwerten von $f^* \circ f$. Die Vektoren v_i und w_j nennt man *linke* bzw. *rechte singuläre Vektoren* von f.

Die Umsetzung dieses Resultats in Matrizensprache ist ziemlich offensichtlich:

Korollar 9.13 (Singulärwertzerlegung). *Sei $A \in M_{m,n}(\mathbb{K})$ eine Matrix mit Rang $r \le min\{m, n\}$. Dann gibt es unitäre (bzw. orthogonale) Matrizen $U_1 \in M_m(\mathbb{K})$ und $U_2 \in M_n(\mathbb{K})$, sowie eine Matrix $\Sigma \in M_{m,n}(\mathbb{K})$ deren einzige Eintragungen ungleich Null die ersten r Eintragungen auf der Hauptdiagonale sind, die reell und positiv sind, sodass $A = U_2 \Sigma U_1^*$ gilt.*

Beweis. Wir betrachten $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$, jeweils mit dem Standard inneren Produkt und der Standardbasis $\mathcal{S}$ und wenden Satz 9.13 auf die lineare Abbildung $f(x) := Ax$ an. Das liefert Orthonormalbasen $\mathcal{B} = \{v_1, \dots, v_n\}$ und $\mathcal{C} = \{w_1, \dots, w_n\}$ und die Singulärwerte $\sigma_1, \dots, \sigma_r$. Definieren wir $\Sigma = [f]^{\mathcal{B}}_{\mathcal{C}}$, dann hat das nach Satz 9.13 die verlangte Form. Außerdem gilt $[f]^{\mathcal{B}}_{\mathcal{C}} = [\mathrm{id}_W]^{\mathcal{S}}_{\mathcal{C}} A [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{S}}$. Weil die Basen $\mathcal{B}$ und $\mathcal{C}$ orthonormal sind, sind die Matrizen $U_1 := [\mathrm{id}_V]^{\mathcal{B}}_{\mathcal{S}}$ und $U_2 = [\mathrm{id}_W]^{\mathcal{C}}_{\mathcal{S}}$ unitär. Nach Konstruktion ist aber $\Sigma = U_2^{-1} A U_1$, also $A = U_2 \Sigma U_1^*$. □

9.14. Die Polarzerlegung. Mit Hilfe der Singulärwertzerlegung können wir nun noch eine der “klassischen” Zerlegungen für quadratische Matrizen ableiten. Motivierend für diese Zerlegung ist die Polardarstellung von komplexen Zahlen in der Form $z = re^{i\varphi}$

mit $r \in [0, \infty)$ und $\varphi \in [0, 2\pi)$. In Termen von komplexen (1×1)–Matrizen ist (r) positiv–semidefinit und $(e^{i\varphi})$ unitär. Die allgemeine Zerlegung ist völlig analog dazu:

Proposition 9.14. *Sei $A \in M_n(\mathbb{K})$ für $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$. Dann gibt es Matrizen $P, U \in M_n(\mathbb{K})$, wobei P positiv semidefinit und U unitär bzw. orthogonal ist, sodass $A = UP$ gilt.*

Beweis. Man kann die Zerlegung auf ähnliche Weise direkt beweisen, wie man die Singulärwertzerlegung konstruiert, also indem man A^*A unitär diagonalisiert. Einfacher können wir auch die Singulärwertzerlegung $A = U_2\Sigma U_1^{-1}$ bilden und beachten, dass hier alle drei Matrizen in der Zerlegung die Größe $n \times n$ haben. Daher können wir $P := U_1\Sigma U_1^{-1}$ und $U = U_2U_1^{-1}$ setzen und dann gilt offensichtlich $A = UP$. Außerdem ist Σ die unitäre bzw. orthogonale Diagonalisierung von P. Weil die Eintragungen der Diagonalmatrix Σ reell und nicht–negativ sind, ist P positiv semidefinit, während U als Produkt zweier unitärer (bzw. orthogonaler) Matrizen natürlich selbst unitär bzw. orthogonal ist. □

Ist $A = UP$ die Polarzerlegung von A, dann ist natürlich $A^* = P^*U^*$, also $A^*A = P^*P = P^2$. Die Matrix P ist also eine positiv semidefinite "Wurzel" aus der positiv semidefiniten Matrix A^*A. Man kann zeigen, dass so eine Wurzel eindeutig bestimmt ist, also ist die P–Komponente in der Polarzerlegung immer eindeutig bestimmt. Ist A invertierbar (oder äquivalent P invertierbar) dann folgt daraus, dass $U = AP^{-1}$ gelten muss, also ist in diesem Fall auch U eindeutig bestimmt. Allgemein ist aber (wie schon im Fall der komplexen Zahlen) die Polarzerlegung aber nicht eindeutig.

Allgemeiner kann man für eine positiv semidefinite Matrix A Wurzeln jeder Ordnung bilden. Dazu diagonalisiert man A unitär und erhält $A = UDU^{-1}$ für eine unitäre (bzw. orthogonale) Matrix U und eine Diagonalmatrix D mit nicht–negativen reellen Eintragungen auf der Hauptdiagonale. Jede dieser Eintragungen hat eine eindeutig, nicht–negative reelle $k$te Wurzel, und man schreibt diese als Eintragungen in eine Diagonalmatrix D_k. Dann gilt natürlich $(D_k)^k = D$ und man verifiziert induktiv, dass $(UD_kU^{-1})^k = U(D_k)^kU^{-1} = UDU^{-1} = A$ gilt. Auch die höheren Wurzeln aus A sind eindeutig festgelegt, wenn man verlangt, dass sie positiv semidefinit sind.

Die Hamilton'schen Quaternionen

Zum Abschluss dieses Kapitels besprechen wir kurz den Schiefkörper der Quaternionen, der interessante Beschreibungen von orthgonalen Abbildungen in Dimension 3 und 4 liefert.

9.15. Grundlegende Eigenschaften. Wir haben in 9.4 orthogonale Abbildungen in Dimension 2 und 3 mit direkten Methoden studiert. In Dimension 2 gibt es, vor allem für die Rotationen, einen alternativen Zugang über komplexe Zahlen. Identifiziert man nämlich $\mathbb{R}^2$ mit $\mathbb{C}$, dann liefert das auch eine Interpretation des Standard inneren Produktes auf $\mathbb{R}^2$ als $\langle z, w\rangle = \mathrm{Re}(z\overline{w}) = \frac{1}{2}(z\overline{w} + \overline{z}w)$. Das ist einfach nur die Tatsache, dass der Realteil von $(a+bi)(c-di)$ gerade $ac+bd$ ist. Daraus schließt man aber sofort, dass $\langle zw_1, zw_2\rangle = |z|^2\langle w_1, w_2\rangle$ gilt. Betrachtet man also $z \in \mathbb{C}$ mit $|z| = 1$, dann definiert $w \mapsto zw$ eine orthogonale lineare Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$. Man sieht auch leicht, dass die Determinante dieser Abbildung immer 1 sein muss, in der Terminologie von Bemerkung 9.1 erhält man immer ein Element von $SO(2)$. Gruppentheoretisch würde man sagen, dass wir einen Isomorphismus zwischen den Gruppen $U(1)$ und $SO(2)$ konstruiert haben.

Motiviert durch diese Beobachtungen versuchte R.W. Hamilton um 1840 eine analoge Beschreibung für orthogonale Abbildungen in Dimension 3 zu finden. Seine Versuche, eine entsprechende algebraische Struktur auf $\mathbb{R}^3$ zu finden, schlugen fehl (und es ist inzwischen bewiesen, dass so eine Struktur nicht existiert). Bei diesen Untersuchung fand er aber heraus, dass es auf $\mathbb{R}^4$ eine sehr schöne algebraische Struktur gibt, die man benutzen kann um orthogonale Abbildungen in den Dimensionen 3 und 4 zu beschreiben. Die entsprechende Struktur ist ein *Schiefkörper* der üblicherweise mit $\mathbb{H}$ bezeichnet und *(Hamilton'sche) Quaternionen* genannt wird.

Die Idee von Hamilton war, zu der "imaginären Einheit" i zwei weitere imaginäre Einheiten j und k hinzuzufügen, und $i^2 = j^2 = k^2 = -1$ zu verlangen. Will man dann aber etwa $ij = k$ definieren, dann sieht man, dass die Relationen mit einer kommutativen Multiplikation nicht erfüllbar sind. Dann würde nämlich $k^2 = ijij = i^2j^2 = (-1)(-1) = 1$ gelten. Hamilton sah, dass sich das Problem beheben lässt, indem man $ij = -ji = k$ definiert und damit die Kommutativität aufgibt. Tatsächlich kann man zeigen, dass die Relationen $i^2 = j^2 = k^2 = ijk = -1$ genügen um eine eindeutige, assoziative Multiplikation auf $\mathbb{H} =: \mathbb{R}^4$ mit Basis $\{1, i, j, k\}$ festzulegen.

Wir gehen hier nicht den mühsamen Weg von direkten Verifikationen sondern geben eine Beschreibung von $\mathbb{H}$, bei der viele der schönen Eigenschaften offensichtlich sind. Als Motivation betrachten wir die Menge $A := \left\{ \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) : a, b \in \mathbb{R} \right\} \subset M_2(\mathbb{R})$. Offensichtlich ist das ein zweidimensionaler Teilraum, also macht die Matrizenaddition diese Menge zu einer kommutativen Gruppe. Andererseits rechnet man sofort:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} ac - bd & -ad - bc \\ bc + ad & -bd + ac \end{pmatrix}$$

Also kann man die Matrixmultiplikation zu einer Multiplikation $A \times A \to A$ einschränken, die auf beiden Seiten distributiv bezüglich der Matrixaddition ist und wegen $\mathbb{I}_2 \in A$ besitzt diese Multiplikation ein neutrales Element. Aus der expliziten Formel folgt auch, dass diese Multiplikation (auf A!) kommutativ ist. Schließlich ist für $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) \in A$ die Determinante $a^2 + b^2$ also ungleich Null solange die Matrix ungleich Null ist, und in diesem Fall ist die Inverse Matrix $1/(a^2 + b^2) \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \in A$. Damit haben wir aber vollständig verifiziert, dass A mit der Einschränkung der üblichen Addition und Multiplikation von Matrizen ein Körper ist. Dieser Körper ist isomorph zu $\mathbb{C}$, indem man die Matrix $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ auf $a + ib$ abbildet.

Definition 9.15. Wir definieren $\mathbb{H} := \left\{ \begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix} : z, w \in \mathbb{C} \right\} \subset M_2(\mathbb{C})$.

Satz 9.15. *Die Menge $\mathbb{H}$ ist ein Teilraum des reellen Vektorraumes $M_2(\mathbb{C})$ mit reeller Dimension 4 (aber kein komplexer Teilraum). Außerdem ist die Teilmenge $\mathbb{H}$ abgeschlossen unter der Matrizenmultiplikation. Ist $A \in \mathbb{H}$ ungleich Null, dann ist A invertierbar und A^{-1} liegt ebenfalls in $\mathbb{H}$. Somit machen die übliche Addition und Multiplikation von Matrizen die Menge $\mathbb{H}$ zu einem* Schiefkörper*, d.h. es sind alle Eigenschaften eines Körpers außer der Kommutativität der Multiplikation erfüllt.*

Beweis. Für $z_1, z_2 \in \mathbb{C}$ gilt $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$ und $\overline{z_1 z_2} = \overline{z}_1 \overline{z}_2$ und für $\lambda \in \mathbb{R}$ ist $\overline{\lambda} = \lambda$. Damit folgt sofort, dass $\mathbb{H}$ ein reeller Teilraum von $M_2(\mathbb{C})$ ist, also ist $\mathbb{H}$ auch eine kommutative Gruppe unter der Matrixaddition. Es liegt zwar die Einheitsmatrix $\mathbb{I}$ in $\mathbb{H}$, aber nicht $i \cdot \mathbb{I}$, also ist $\mathbb{H}$ kein komplexer Teilraum.

Für das Produkt von zwei Elementen von $\mathbb{H}$ rechnet man einfach

$$\begin{pmatrix} z_1 & -\overline{w}_1 \\ w_1 & \overline{z}_1 \end{pmatrix} \begin{pmatrix} z_2 & -\overline{w}_2 \\ w_2 & \overline{z}_2 \end{pmatrix} = \begin{pmatrix} z_1 z_2 - \overline{w}_1 w_2 & -z_1 \overline{w}_2 - \overline{w}_1 \overline{z}_2 \\ w_1 z_2 + \overline{z}_1 w_2 & -w_1 \overline{w}_2 + \overline{z}_1 \overline{z}_2 \end{pmatrix},$$

also liegt das Produkt wieder in $\mathbb{H}$. Somit definiert die Matrixmultiplikation eine assoziative Multiplikation auf $\mathbb{H}$, die nicht kommutativ, aber distributiv bezüglich der Addition ist. Weil $\mathbb{I} \in \mathbb{H}$ gilt, besitzt diese Multiplikation ein neutrales Element. Schließlich gilt $\det \begin{pmatrix} z & -\overline{w} \\ w & \overline{z} \end{pmatrix} = |z|^2 + |w|^2$, also ist jede Matrix $\neq 0$ in $\mathbb{H}$ invertierbar. Die Inverse Matrix dazu ist gegeben durch $\frac{1}{|z|^2+|w|^2} \begin{pmatrix} \overline{z} & -w \\ \overline{w} & z \end{pmatrix}$ und liegt damit ebenfalls in $\mathbb{H}$. □

Das "klassische" Bild der Quaternionen erhält man leicht, indem man eine passende Basis wählt. Man setzt einfach

$$1 = \mathbb{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \quad i = \left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right) \quad j = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \quad k = \left(\begin{smallmatrix} 0 & -i \\ -i & 0 \end{smallmatrix}\right)$$

und beobachtet, dass diese Matrizen eine Basis für den reellen Vektorraum $\mathbb{H}$ bilden. Natürlich ist 1 ein multiplikativ neutrales Element und man rechnet einfach direkt nach, dass $i^2 = j^2 = k^2 = -1$ sowie $ij = -ji = k$ gilt. Daraus folgt dann durch rein formales Rechnen $ki = -ik = j$ und weiters $jk = -kj = i$, womit man eine vollständige Liste der Produkte der Basiselemente hat.

Man kann somit jedes Element $q \in \mathbb{H}$ als $q = a1 + bi + cj + dk$ für $a, b, c, d \in \mathbb{R}$ schreiben und Produkte einfach mittels Distributivität aus den Produkten der Basiselemente ausrechnen. Die entstehende Formel für die Multiplikation zweier Quaternionen ist etwas länglich und unhandlich, aber nicht wirklich kompliziert. In der Notation lässt man die 1 meist weg und schreibt einfach $q = a + bi + cj + dk$. Analog zu den komplexen Zahlen definiert man den *Realteil* von q als $\mathrm{Re}(q) := a$ und den *Imaginärteil* von q als $\mathrm{Im}(q) := bi + cj + dk$. Quaternionen mit Realteil Null werden als *rein imaginäre Quaternionen* bezeichnet. Man schreibt $\mathrm{Im}(\mathbb{H})$ für den Raum der rein imaginären Quaternionen.

Eine weitere nützliche Darstellung der Quaternionen erhält man, indem man den Raum $\mathrm{Im}(\mathbb{H})$ mit $\mathbb{R}^3$ identifiziert, also $bi + cj + dk$ mit dem Vektor $(b, c, d) \in \mathbb{R}^3$ identifiziert. Man schreibt dann $q = (a, X)$ mit $a \in \mathbb{R}$ und $X \in \mathbb{R}^3$ und nennt a den *Skalarteil* und X den *Vektorteil* von q. In diesem Bild ist die Addition natürlich komponentenweise definiert, die Multiplikation werden wir in Kürze beschreiben.

Parallel zu den komplexen Zahlen definiert man die *Konjugation* auf $\mathbb{H}$. Um $q \in \mathbb{H}$ zu konjugieren, lässt man den Realteil unverändert und ersetzt den Imaginärteil durch sein Negatives. Es gilt also $\overline{a + bi + cj + dk} = a - bi - cj - dk$ bzw. $\overline{(a, X)} = (a, -X)$. Auch im Bild der Matrizen hat die Konjugation eine schöne Interpretation über das Bilden der adjungierten Matrix A^*. Betrachten wir die Basiselemente von oben, dann sieht man sofort, dass $1^* = 1$, $i^* = -i$, $j^* = -j$ und $k^* = -k$ gelten. Da die Abbildung $*$ $\mathbb{R}$–linear auf $M_2(\mathbb{C})$ ist, ist für $A \in \mathbb{H}$ auch $A^* \in \mathbb{H}$ und entspricht der konjugierten Quaternione zu A. Daher kann man Real– und Imaginärteil im Matrizenbild als $\frac{1}{2}(A + A^*)$ und $\frac{1}{2}(A - A^*)$ schreiben.

Alternativ bemerkt man, dass $\mathrm{tr}(\mathbb{I}) = 2$ gilt, während die Matrizen i, j und k offensichtlich Spur Null haben. Aus der Linearität der Spur folgt damit aber sofort, dass für $A = a1 + bi + cj + dk \in \mathbb{H} \subset M_2(\mathbb{C})$ gilt, dass $\mathrm{tr}(A) = 2a$ ist. Damit erhält man aber $\mathrm{Re}(A) = \frac{1}{2}\,\mathrm{tr}(A)\mathbb{I}$ und $\mathrm{Im}(A) = A - \frac{1}{2}\,\mathrm{tr}(A)\mathbb{I}$ (der "spurfreie Teil" von A).

Damit können wir nun die Multiplikation von Quaternionen im Bild von Skalarteil und Vektorteil beschreiben.

Proposition 9.15. *Schreibt man Quaternionen in der Form $q = (a, X)$ mit $a \in \mathbb{R}$ und $X \in \mathbb{R}^3$ dann ist die Multiplikation von Quaternionen gegeben durch*

$$(a, X) \cdot (b, Y) = (ab - \langle X, Y\rangle, aY + bX + X \times Y),$$

wobei in der ersten Komponente das Standard innere Produkt auf $\mathbb{R}^3$ verwendet wird und in der zweiten Komponente das Kreuzprodukt aus 8.12. Insbesondere gilt $\overline{p \cdot q} = \overline{q} \cdot \overline{p}$.

Beweis. Für $p, q \in \mathbb{H}$ schreiben wir $p = a1+\mathrm{Im}(p)$ und $q = b1+\mathrm{Im}(q)$, sodass X und Y den Imaginärteilen entsprechen. Wegen der Distributivität und weil 1 multiplikativ neutral ist, gilt $(a1+\mathrm{Im}(p))(b1+\mathrm{Im}(q)) = ab1+a\,\mathrm{Im}(q)+b\,\mathrm{Im}(p)+\mathrm{Im}(p)\,\mathrm{Im}(q)$. Damit genügt es aber, die behauptete Formel für $(0, X)$ und $(0, Y)$ zu verifizieren. Dafür rechnet man einfach direkt $(x_1i+x_2j+x_3k)(y_1i+y_2j+y_3k)$ unter Benutzung von Distributivität und der Relationen der Basiselemente aus. Die Komponente von 1 erhält man einfach aus $i^2 = j^2 = k^2 = -1$ als $-x_1y_1 - x_2y_2 - x_3y_3$. Eine Komponente von i erhält man entweder aus $jk = i$ oder $kj = -i$, also ergibt sich $(x_2y_3 - x_3y_2)i$ und so weiter.

Die letzte Eigenschaft folgt dann einfach durch Nachrechnen. Alternativ folgt sie auch sofort aus $(AB)^* = B^*A^*$ im Matrizenbild. □

Diese Beschreibung ist der Ursprung der Namen "Skalarprodukt" für das innere Produkt und "Vektorprodukt" für das Kreuzprodukt in $\mathbb{R}^3$.

9.16. Quaternionen und Geometrie. Aus der Formel für die Quaternionenmultiplikation in Proposition 9.15 ist schon sichtbar, dass die Quaternionen die grundlegenden geometrischen Operationen auf $\mathbb{R}^3$ liefern. Man kann das aber noch schöner formalisieren. Betrachten wir nämlich die Abbildung $(p, q) \mapsto p\overline{q}$. Aus Proposition 9.15 sehen wir, dass der Realteil von $(a, X)(b, -Y)$ gerade $ab + \langle X, Y\rangle$ ist. Das liefert also gerade das Standard innere Produkt auf $\mathbb{R}^4 \cong \mathbb{H}$ und durch Einschränkung auf dem Teilraum $\mathbb{R}^3 \cong \mathrm{Im}(\mathbb{H})$. Benutzt man noch $\mathrm{Re}(q) = \frac{1}{2}(p + \overline{p})$ und die Verträglichkeit der Konjugation mit der Multiplikation, dann folgt $\langle p, q\rangle = \frac{1}{2}(p\overline{q} + q\overline{p})$.

Für ein Quaternion $q \in \mathbb{H}$ gilt nun $\overline{q\overline{q}} = q\overline{q}$, also stimmt $q\overline{q}$ mit seinem Realteil $\langle q, q\rangle = |q|^2$ überein. Damit erhält man wie für komplexe Zahlen $q\overline{q} = \overline{q}q = |q|^2 \cdot 1$, also $q^{-1} = \frac{1}{|q|^2}\overline{q}$. Für rein imaginäre Quaternionen gilt zusätzlich $\overline{q} = -q$, also $q^2 = -q\overline{q} = -|q|^2 \cdot 1$. Eine weitere Teilmenge, die im Weiteren sehr wichtig sein wird, sind die *Einheitsquaternionen*

$$Sp(1) := \{q \in \mathbb{H} : |q| = 1\} \subset \mathbb{H}$$

Proposition 9.16. *Für zwei Quaternionen $p, q \in \mathbb{H}$ gilt immer $|pq| = |p||q|$. Insbesondere macht die Multiplikation von Quaternionen die Einheitsquaternionen $Sp(1)$ zu einer (nicht kommutativen) Gruppe. Die Gruppe $Sp(1)$ ist isomorph zur Gruppe $SU(2)$ aller unitären 2×2–Matrizen mit Determinante 1.*

Beweis. Es gilt $|pq|^2 = \langle pq, pq\rangle = pq\overline{pq} = p(q\overline{q})\overline{p} = p(|q|^2 1)\overline{p} = |q|^2 p\overline{p} = |p|^2|q|^2 1$. Daraus folgt die erste Behauptung sofort durch Ziehen der Wurzel. Außerdem folgt auch sofort, dass für $p, q \in Sp(1)$ auch $pq \in Sp(1)$ gilt. Außerdem ist $q^{-1} = \bar{q} \in Sp(1)$ für $q \in Sp(1)$. Damit definiert die Quaternionenmultiplikation eine assoziative Multiplikation auf $Sp(1)$, die wegen $|1| = 1$ ein neutrales Element und nach der obigen Überlegung inverse Elemente besitzt. Also ist $Sp(1)$ eine Gruppe, die aber wegen $ij = -ji$ nicht kommutativ ist.

Für die letzte Behauptung wechseln wir wieder in das Bild der Matrizen, betrachten also $Sp(1) \subset \mathbb{H} \subset M_2(\mathbb{C})$ und zeigen, dass in diesem Bild $Sp(1)$ mit $SU(2)$ übereinstimmt. Für $A \in \mathbb{H}$ ist $|A| = 1$ äquivalent zu $A^*A = \mathbb{I}$. In diesem Fall ist aber $A \in U(2)$ nach Korollar 9.1 und $\det(A) = 1$, also $A \in SU(2)$. Ist umgekehrt $B \in U(2)$, dann bilden die Spaltenvektoren von B nach Korollar 9.1 eine Orthonormalbasis von $\mathbb{C}^2$. Bezeichnen wir den ersten Spaltenvektor von B mit $\binom{z}{w}$, dann muss $|z|^2 + |w|^2 = 1$ gelten, weil es sich um einen Einheitsvektor handelt. Offensichtlich steht dann der Vektor $\binom{-\overline{w}}{\overline{z}}$ normal auf $\binom{z}{w}$ und weil das orthogonale Komplement Dimension 1 hat, muss die zweite

Spalte von B von der Form $\binom{-\lambda\overline{w}}{\lambda\overline{z}}$ für eine komplexe Zahl λ mit $|\lambda| = 1$ sein. Dann ist aber $\det(B) = \lambda(|z|^2 + |w|^2)$, also liegt B genau dann in $SU(2)$, wenn $B \in \mathbb{H}$ gilt. $\square$

Wir können nun das innere Produkt auf $\mathbb{R}^3 = \operatorname{Im}(\mathbb{H})$ noch besser in Termen der Quaternionenmultiplikation verstehen. Seien $X, Y \in \operatorname{Im}(\mathbb{H})$ rein imaginäre Quaternionen. Dann ist nach Proposition 9.15 $X \cdot Y = (\langle X, Y\rangle, X \times Y)$ und damit $X \cdot Y + Y \cdot X = 2\langle X, Y\rangle \cdot 1$. Damit ist aber $X \perp Y$ genau dann, wenn $X \cdot Y = -Y \cdot X$ gilt. Ist zusätzlich $|X| = |Y| = 1$, dann gilt $X \cdot Y = X \times Y \in \operatorname{Im}(\mathbb{H})$ und $\{X, Y, X \times Y\}$ ist eine Orthonormalbasis, die $\det(X, Y, X \times Y) = 1$ erfüllt.

Betrachten wir nun für eine Einheitsquaternione $q \in Sp(1)$ und eine rein imaginäre Quaternione X die Quaternione $qXq^{-1} = qX\overline{q}$. Dann gilt $\overline{qX\overline{q}} = q\overline{X}\overline{q} = -qX\overline{q}$, also $qXq^{-1} \in \operatorname{Im}(\mathbb{H})$. Also können wir für fixes q die Abbildung $\varphi_q : \operatorname{Im}(\mathbb{H}) \to \operatorname{Im}(\mathbb{H})$ betrachten, die durch $\varphi_q(X) := qX\overline{q}$ gegeben ist.

Satz 9.16. *(1) Für jedes $q \in Sp(1)$ ist die Abbildung $\varphi_q : \operatorname{Im}(\mathbb{H}) \to \operatorname{Im}(\mathbb{H})$ orthogonal und $\det(\varphi_q) = 1$, also gilt $\varphi_q \in SO(3)$.*

(2) Die Abbildung $q \mapsto \varphi_q$ ist ein surjektiver Gruppenhomomorphismus $Sp(1) \to SO(3)$ und für $q_1, q_2 \in Sp(1)$ gilt $\varphi_{q_1} = \varphi_{q_2}$ genau dann, wenn $q_2 = \pm q_1$ gilt.

Beweis. (1) Wir haben bereits gesehen, dass $\varphi_q(\operatorname{Im}(\mathbb{H})) \subset \operatorname{Im}(\mathbb{H})$ gilt, also können wir φ_q als Abbildung von $\mathbb{R}^3 = \operatorname{Im}(\mathbb{H})$ auf sich selbst betrachten. Für $X, Y \in \operatorname{Im}(\mathbb{H})$ und $\lambda \in \mathbb{R}$ gilt

$$\varphi_q(X + \lambda Y) = q(X + \lambda Y)\overline{q} = (qX + \lambda qY)\overline{q} = qX\overline{q} + \lambda qY\overline{q}$$

wegen der Bilinearität der Multiplikation, also ist $\varphi_q : \mathbb{R}^3 \to \mathbb{R}^3$ eine lineare Abbildung. Nun gilt aber $|\varphi_q(X)| = |qX\overline{q}| = |q||X||\overline{q}| = |X|$, also ist φ_q nach Satz 9.1 orthogonal. Schließlich bemerken wir noch, dass wir $\varphi_q(k)$ als $\varphi_q(ij) = qij\overline{q} = qi\overline{q}qj\overline{q} = \varphi_q(i)\varphi_q(j)$ schreiben können. Schreibt man stattdessen $k = -ji$, dann erhält man genau so $\varphi_q(k) = -\varphi_q(j)\varphi_q(i)$. Aus den obigen Überlegungen schließen wir damit aber, dass $\varphi_q(k) = \varphi_q(i) \times \varphi_q(j)$ gilt und damit $\det(\varphi_q(i), \varphi_q(j), \varphi_q(k)) = 1$ ist. Das ist aber gerade die Matrix von φ_q bezüglich der Standardbasis, also ist $\det(\varphi_q) = 1$.

(2) Nach Teil (1) definiert $q \mapsto \varphi_q$ eine Funktion $Sp(1) \to SO(3)$. Für $q_1, q_2 \in Sp(1)$ und $X \in \operatorname{Im}(\mathbb{H})$ ist $\varphi_{q_1q_2}(X) = (q_1q_2)X(\overline{q_1q_2})$ Benutzt man die Assoziativität der Multiplikation und $\overline{q_1q_2} = \overline{q}_2\overline{q}_1$, dann kann man das als $q_1\varphi_{q_2}(X)\overline{q}_1 = \varphi_{q_1} \circ \varphi_{q_2}(X)$ schreiben. Damit gilt $\varphi_{q_1q_2} = \varphi_{q_1} \circ \varphi_{q_2}$, also haben wir einen Gruppenhomomorphismus definiert.

Daraus folgt dann auch, dass $\varphi_{q^{-1}} = (\varphi_q)^{-1}$ gilt. Das kann man benutzen, um zu zeigen, dass $\varphi_{q_1} = \varphi_{q_2}$ genau dann gilt, wenn $\mathrm{id} = \varphi_{q_2} \circ (\varphi_{q_1}^{-1}) = \varphi_{q_2(q_1)^{-1}}$ gilt. Damit können wir den letzten Teil beweisen, indem wir zeigen, dass $\varphi_q = \mathrm{id}$ nur für $q = \pm 1$ gilt. Nun ist aber $X = \varphi_q(X) = qXq^{-1}$ genau dann, wenn $Xq = qX$ gilt. Schreiben wir $q = a + bi + cj + dk$ und überprüfen das für $X = i, j, k$, dann folgt sofort $b = c = d = 0$. Damit ist aber $q = a \cdot 1$ und aus $|q| = 1$ folgt nun $q = \pm 1$.

Damit bleibt nur noch die Surjektivität zu beweisen. Dazu benutzen wir die Beschreibung orthogonaler Abbildungen aus 9.4. Dort haben wir gesehen, dass man jede Matrix in $SO(3)$ als Rotation um eine Achse in $\mathbb{R}^3$ schreiben kann. (Die Spiegelung an einer Geraden, die dort vorgekommen ist, kann man als Rotation um diese Gerade um den Winkel π schreiben.) Zu $A \in SO(3)$ finden wir also einen Einheitsvektor $X \in \operatorname{Im}(\mathbb{H})$ und einen Winkel α, sodass A die Rotation um die Achse $\mathbb{R} \cdot X$ mit Winkel α ist. Nun setzen wir $q := \cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}X \in \mathbb{H}$. Dann gilt natürlich $|q|^2 = \cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2}|X| = 1$, also $p \in Sp(1)$. Außerdem gilt $X \cdot X = -1$ und damit

$$\varphi_q(X) = (\cos\tfrac{\alpha}{2}X - \sin\tfrac{\alpha}{2})(\cos\tfrac{\alpha}{2} - \sin\tfrac{\alpha}{2}X) = (\cos^2\tfrac{\alpha}{2} + \sin^2\tfrac{\alpha}{2})X = X.$$

Wählt man nun einen Einheitsvektor $Y \perp X$, dann ist $\{X, Y, X \cdot Y\}$ eine Orthonormalbasis für $\operatorname{Im}(\mathbb{H})$. Unter Benutzung von $X \cdot Y = -Y \cdot X$ und $X \cdot (X \cdot Y) = -Y$ rechnet man wie oben nach, dass $\varphi_q(Y) = \cos\alpha Y + \sin\alpha X \cdot Y$ und $\varphi_q(X \cdot Y) = -\sin\alpha Y + \cos\alpha X \cdot Y$ gilt. Damit ist aber φ_q genau die Rotation um die Achse $\mathbb{R} \cdot X$ um den Winkel α. □

Mit Hilfe dieses Resultats kann man sehen, wie die Gruppe $SO(3)$ oder äquivalent die Menge aller Orthonormalbasen mit Determinante Eins "geometrisch" aussieht. Die Gruppe $Sp(1)$ besteht ja gerade aus allen Einheitsvektoren in $\mathbb{H} = \mathbb{R}^4$, ist also eine dreidimensionale Sphäre. Daraus erhält man $SO(3)$ indem man jeden Punkt q mit dem gegenüberliegenden Punkt $-q$ identifiziert. Alternativ, kann man $SO(3)$ als die Menge aller eindimensionalen Teilräume von $\mathbb{R}^4$ betrachten, denn jeder solche Teilraum schneidet die Sphäre in genau zwei Punkten, die jeweils Negative von einander sind.

9.17. Quaternionen und orthogonale Abbildungen in Dimension 4. Unsere Diskussion des Zusammenhanges zwischen $Sp(1)$ und $SO(3)$ in Satz 9.16 ist nicht völlig befriedigend in dem Sinn, dass wir für den Beweis, dass man jedes Element von $SO(3)$ quaternionisch realisieren kann, die explizite Beschreibung von $SO(3)$ aus 9.4 gebraucht haben. Das ist aber etwas irreführend, weil es, wenn man etwas mehr Theorie zur Verfügung hat, auch einen viel einfacheren Beweisweg gibt. Dazu benötigt man einige fundamentale Resultate über Matrizengruppen, die algebraische Methoden mit mehrdimensionaler Analysis kombinieren. Hat man diese Methoden zur Hand, dann muss man im wesentlichen nur beweisen, dass $Sp(1)$ und $SO(3)$ beide Dimension 3 haben und kann dann die Surjektivität in Satz 9.17 einfach daraus folgern, dass $\varphi_q = \mathrm{id}$ nur für $q = \pm 1$ gilt. (Natürlich braucht man dazu, dass diese Gruppen tatsächlich eine wohldefinierte Dimension haben, aber intuitiv ist das verständlich.)

Diese Argumentationslinie funktioniert in beliebigen Dimensionen. Überlegt wir kurz, wie es mit den Dimensionen der Gruppen $O(n)$ aussieht. Für A in $O(n)$ bilden die Spaltenvektoren eine Orthonormalbasis von $\mathbb{R}^n$, also kann man $O(n)$ mit der Menge der Orthonormalbasen von $\mathbb{R}^n$ identifizieren. Um eine Orthonormalbasis für $\mathbb{R}^2$ festzulegen, muss man zunächst einen Einheitsvektor festlegen. Die Menge dieser Einheitsvektoren ist der Einheitskreis, hat also Dimension 1. Dann gibt es nur zwei Vektoren, die den gegebenen Einheitsvektor zu einer Orthonormalbasis ergänzen, also hat $O(2)$ Dimension 1. Für $O(3)$ wählt man erst einen Einheitsvektor $v \in \mathbb{R}^3$, was eine zweidimensionale Freiheit bedeutet und dann eine Orthonormalbasis für $v^\perp \cong \mathbb{R}^2$. Das liefert eine weitere Dimension, also hat $O(3)$ Dimension 3 (wie auch $Sp(1) \cong S^3$). Induktiv folgt, dass die Dimension von $O(n)$ gerade $n-1$ plus die Dimension von $O(n-1)$ ist, insbesondere hat $O(4)$ Dimension 6. Daraus schließt man leicht, dass die Dimension von $O(n)$ gerade $n(n-1)/2$ ist. Da $SO(n)$ gerade "die Hälfte" von $O(n)$ ist, haben beide Gruppen die gleiche Dimension.

Insbesondere hat $SO(4)$ die gleiche Dimension wie $Sp(1) \times Sp(1)$, also ein Produkt von zwei Kopien der Einheitsquaternionen. Dieses Produkt ist eine Gruppe unter der Operation $(p_1, p_2)(q_1, q_2) = (p_1q_1, p_2q_2)$. Nun können wir analog zu 9.16 einem Element $(q_1, q_2) \in Sp(1) \times Sp(1)$ eine Funktion $\psi_{(q_1,q_2)} : \mathbb{H} \to \mathbb{H}$ zuordnen, indem wir $\psi_{(q_1,q_2)}(p) := q_1 p \overline{q}_2$ definieren.

Satz 9.17. *(1) Für jedes Element $(q_1, q_2) \in Sp(1) \times Sp(1)$ ist $\psi_{(q_1,q_2)}$ eine orthogonale lineare Abbildung $\mathbb{H} \to \mathbb{H}$.*

(2) Die Abbildung $(q_1, q_2) \mapsto \psi_{(q_1,q_2)}$ *definiert einen Homomorphismus von Gruppen* $Sp(1) \times Sp(1) \to O(4)$ *und es gilt* $\psi_{(q_1,q_2)} = \psi_{(p_1,p_2)}$ *genau dann, wenn* $(p_1, p_2) = (q_1, q_2)$ *oder* $(p_1, p_2) = (-q_1, -q_2)$ *gilt.*

Beweis. (1) Die Linearität von $\psi_{(q_1,q_2)}$ folgt sofort aus dem Distributivgesetz. Für die Orthogonalität genügt es wieder $|\psi_{(q_1,q_2)}(p)| = |p|$ zu beweisen. Das folgt aber sofort aus $|q_1 p\overline{q}_2| = |q_1||p||\overline{q}_2|$.

(2) Es gilt $\psi_{(p_1q_1,p_2q_2)}(p) = (p_1q_1)p(\overline{p_2q_2}) = p_1q_1p\overline{q}_2\overline{p}_2 = \psi_{(p_1,p_2)}(\psi_{(q_1,q_2)}(p))$, also $\psi_{(p_1q_1,p_2q_2)} = \psi_{(p_1,p_2)} \circ \psi_{(q_1,q_2)}$. Für den zweiten Teil genügt es dann wie in Satz 9.16 den Fall $\psi_{(q_1,q_2)} = \mathrm{id}$ zu analysieren. Dann gilt aber $q_1p\overline{q}_2 = p$ für alle $p \in \mathbb{H}$. Für $p = 1$ erhält man $q_1\overline{q}_2 = 1$, also $\overline{q}_2 = (q_1)^{-1} = \overline{q}_1$. Damit folgt $q_1 = q_2$, und in Satz 9.16 haben wir schon bewiesen, dass aus $\psi_{(q,q)}|_{\mathrm{Im}(\mathbb{H})} = \mathrm{id}$ schon $q = \pm 1$ folgt. □

Man kann dann noch relativ einfach zeigen, dass $\psi_{(q_1,q_2)}$ immer Determinante 1 hat, also in $SO(4)$ liegt. Die oben angesprochenen allgemeinen Resultate zeigen dann, dass jedes Element von $SO(4)$ in der Form $\psi_{(q_1,q_2)}$ geschrieben werden kann, sodass man eine vollständige Beschreibung von $SO(4)$ erhält.

Die Tatsache, dass die Gruppe $SO(4)$ "beinahe als Produkt von zwei Gruppen geschrieben werden kann", stimmt nur für diese eine Dimension, es gibt weder in niedrigeren noch in höheren Dimensionen ein ähnliches Phänomen. Das spielt eine wichtige Rolle in Teilen der Differentialgeometrie und der theoretischen Physik, in denen die Dimension 4 eine besondere Rolle spielt.

Andererseits sind die "Erweiterungen" $Sp(1) \to SO(3)$ und $Sp(1) \times Sp(1) \to SO(4)$ die sogenannten *Spin–Gruppen* in Dimension 3 und 4. Das führt zum Konzept von Spinoren, die ebenfalls eine wichtige Rolle in Teilen der Differentialgeometrie und der theoretischen Physik spielen.

KAPITEL 10

Normalformen

In diesem Kapitel kehren wir zu dem Problem zurück, für einen endlichdimensionalen Vektorraum V über einem Körper $\mathbb{K}$ und eine lineare Abbildung $f : V \to V$ eine möglichst einfache Matrixdarstellung zu finden. Wir kehren also wieder in das Setting der allgemeinen linearen Algebra (ohne innere Produkte oder Normen) und (zumindest für den Anfang) über allgemeinen Körpern zurück. Tatsächlich werden wir Normalformen für beliebige lineare Abbildungen nur im Fall eines algebraisch abgeschlossenen Körpers (also insbesondere für $\mathbb{K} = \mathbb{C}$) finden können und daraus auch ein Resultat für $\mathbb{K} = \mathbb{R}$ ableiten. Wir werden aber auch sehen, was man "verstehen muss" um das Problem für andere Körper zu lösen.

Technisch gesehen ist der Schwerpunkt dieses Kapitels der Zusammenhang zwischen der algebraischen Theorie des Polynomringes $\mathbb{K}[x]$ und der Theorie der linearen Abbildungen. Insbesondere werden wir die Analogie von $\mathbb{K}[x]$ zum Ring $\mathbb{Z}$ der ganzen Zahlen, die wir schon in Kapitel 7 angesprochen haben, noch viel tiefer analysieren.

Invariante Teilräume

Wir werden unsere Normalform für eine lineare Abbildung $f : V \to V$ in zwei Schritten konstruieren. Man zerlegt einerseits V in Teilräume, auf die man f einschränken kann. Dann muss man analysieren, wie die einzelnen Einschränkungen aussehen können.

10.1. Invariante Teilräume und charakteristisches Polynom. Sei V ein Vektorraum über $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. In Definition 7.13 haben wir einen f–invariant Teilraum von V als einen Teilraum $W \subset V$ definiert, sodass $f(w) \in W$ für alle $w \in W$ gilt. Dort haben wir auch bemerkt, dass wir dann einerseits die Einschränkung $f|_W : W \to W$ bilden können. Andererseits erhalten wir die induzierte Abbildung $\underline{f} : V/W \to V/W$ auf dem Quotientenraum V/W, die durch $\underline{f} \circ \pi = \pi \circ f$ charakterisiert ist, wobei $\pi : V \to V/W$ die kanonische Surjektion bezeichnet.

Proposition 10.1. *Für einen n–dimensionalen $\mathbb{K}$–Vektorraum V und eine lineare Abbildung $f : V \to V$ gilt:*

Es gibt genau dann einen k–dimensionalen f–invarianten Teilraum $W \subset V$, wenn bezüglich einer geeigneten Basis $\mathcal{B}$ von V die Matrixdarstellung von f die Blockform $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ für $A \in M_k(\mathbb{K})$, $B \in M_{k,n-k}(\mathbb{K})$ und $C \in M_{n-k}(\mathbb{K})$ besitzt.

Ist das der Fall, dann ist A eine Matrixdarstellung der Einschränkung $f|_W$ und C eine Matrixdarstellung der induzierten Abbildung $\underline{f} : V/W \to V/W$.

Beweis. Ist $W \subset V$ ein k–dimensionaler, f–invarianter Teilraum, dann wählen wir eine Basis $\{v_1, \dots, v_k\}$ für W und ergänzen sie zu einer Basis $\mathcal{B} = \{v_1, \dots, v_n\}$ von V. Da W f–invariant ist, gilt insbesondere $f(v_i) \in W$ für $i = 1, \dots, k$, also haben die ersten k Spalten $[f(v_1)]_{\mathcal{B}}, \dots, [f(v_k)]_{\mathcal{B}}$ nur in den ersten k Zeilen Eintragungen ungleich Null. Das bedeutet aber gerade, dass $[f]_{\mathcal{B}}$ die verlangte Blockform besitzt. Klarerweise ist der

Block $A \in M_k(\mathbb{K})$ genau die Matrixdarstellung von $f|_W$ bezüglich der Basis $\{v_1, \dots, v_k\}$ von W.

Die Abbildung $\pi : V \to V/W$ ist surjektiv, und ihr Kern ist W. Aus dem Beweis des Dimensionssatzes für lineare Abbildungen (Satz 4.11) sehen wir damit sofort, dass $\mathcal{B}' := \{\pi(v_{k+1}), \dots, \pi(v_n)\}$ eine Basis für V/W ist. Setzen wir $B = (b_{ij})$ und $C = (c_{ij})$, dann gilt für $j = 1, \dots, n-k$ nach Konstruktion

$$f(v_{k+j}) = b_{1,j}v_1 + \dots + b_{k,j}v_k + c_{1,j}v_{k+1} + \dots + c_{n-k,j}v_n.$$

Damit ist aber $\underline{f}(\pi(v_{k+j})) = \pi(f(v_{k+j})) = c_{1,j}\pi(v_{k+1}) + \dots + c_{n-k,j}\pi(v_n)$, also ist $C = [\underline{f}]_{\mathcal{B}'}$.

Ist umgekehrt $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis von V, sodass die Matrixdarstellung $[f]_{\mathcal{B}}$ die angegebene Blockform hat, dann sei $W := \langle v_1, \dots, v_k \rangle$ der von den ersten k Basisvektoren erzeugte Teilraum. Wegen der Blockform der Matrix ist $f(v_i) \in W$ für $i = 1, \dots, k$, also $f(W) \subset W$. □

Lemma 10.1. *Sei R ein kommutativer Ring mit Einselement, seien $n, k \geq 1$ und betrachte eine Blockmatrix $M := \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ mit $A \in M_k(R)$, $B \in M_{k,n-k}(R)$ und $C \in M_{n-k}(R)$ wie in Proposition 10.1. Dann ist* $\det(M) = \det(A)\det(C)$.

Beweis. Wir beweisen die Behauptung durch Induktion nach k (für beliebiges n). Ist $k = 1$, dann folgt die Behauptung sofort durch Entwickeln nach der ersten Spalte.

Nehmen wir also an, dass $k \geq 2$ gilt und der Satz für Blöcke der Größe $k-1$ bereits bewiesen ist. Wie in 6.9 bezeichnen wir mit A_{ij} die Matrix, die aus $A = (a_{ij})$ durch Streichen der $i$ten Zeile und der $j$ten Spalte entsteht. Weiters sei B_j die Matrix die aus B durch Streichen der $j$ten Zeile entsteht. Entwickeln wir nun die Determinante nach der ersten Spalte, dann erhalten wir

$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \sum_{j=1}^{k} (-1)^{1+j} a_{j1} \det \begin{pmatrix} A_{j1} & B_j \\ 0 & C \end{pmatrix}.$$

Nach Induktionsvoraussetzung ist die letzte Determinante gleich $\det(A_{j1})\det(C)$, und wir erhalten $\left(\sum_{j=1}^{k} (-1)^{1+j} a_{j1} \det(A_{j1})\right) \det(C) = \det(A)\det(C)$. □

Korollar 10.1. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$ und sei $f : V \to V$ eine lineare Abbildung. Für einen f–invarianten Teilraum $W \subset V$ sei $f|_W : W \to W$ die Einschränkung von f und $\underline{f} : V/W \to V/W$ die induzierte Abbildung. Dann gilt für die charakteristischen Polynome $p_f = (p_{f|_W})(p_{\underline{f}})$.*

Beweis. Nach Proposition 10.1 finden wir eine Basis $\mathcal{B}$ für V, sodass $[f]_{\mathcal{B}}$ Blockform hat. Das charakteristische Polynom von f ist dann

$$\det\left(\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} - x\mathbb{I}\right) = \det \begin{pmatrix} A - x\mathbb{I} & B \\ 0 & C - x\mathbb{I} \end{pmatrix}.$$

Damit folgt die Behauptung aus Lemma 10.1. □

Bemerkung 10.1. Mit diesem Resultat können wir leicht die Resultate über komplexe Triangulierbarkeit aus Proposition 7.13 auf allgemeine Körper erweitern. Sei V ein $\mathbb{K}$–Vektorraum und $f : V \to V$ eine lineare Abbildung. Falls es eine Basis $\mathcal{B}$ von V gibt, sodass $[f]_{\mathcal{B}} =: (a_{ij})$ eine obere Dreiecksmatrix ist, dann kann man das charakteristische Polynom von f aus dieser Matrixdarstellung ausrechnen. Man erhält natürlich $p_f = (a_{11} - x) \cdots (a_{nn} - x)$, also zerfällt p_f in ein Produkt von Polynomen ersten Grades (und die a_{ii} sind die Eigenwerte von f).

Nehmen wir umgekehrt an, dass p_f als Produkt von Polynomen ersten Grades geschrieben werden kann. Dann kann man wie im Beweis von Proposition 7.13 induktiv zeigen, dass f eine Matrixdarstellung als obere Dreiecksmatrix besitzt. Nach Voraussetzung hat nämlich f einen Eigenwert λ_1, und ein Eigenvektor v_1 zu diesem Eigenwert spannt einen eindimensionalen, f–invarianten Teilraum $W \subset V$ auf. Natürlich ist $p_{f|_W} = (\lambda_1 - x)$ und $p_f = (\lambda_1 - x)q$, wobei q nach Voraussetzung in ein Produkt von Polynomen ersten Grades zerfällt. Ist $\underline{f} : V/W \to V/W$ die von f induzierte Abbildung, dann gilt nach Lemma 10.1 $(\lambda_1 - x)p_{\underline{f}} = p_f$ und damit $p_{\underline{f}} = q$. Damit zerfällt aber $p_{\underline{f}}$ in ein Produkt von Polynomen ersten Grades, also kann man die Induktionsvoraussetzung anwenden und den Beweis wie in 7.13 abschließen.

Man bemerke auch, dass dieses Resultat impliziert, dass für eine lineare Abbildung $f : V \to V$ sodass p_f in ein Produkt von Polynomen ersten Grades zerfällt, die Spur $\mathrm{tr}(f)$ die Summe aller Eigenwerte von f und die Determinante $\det(f)$ das Produkt aller Eigenwerte von f ist.

10.2. Invariante Komplemente und invariante Summenzerlegungen. Sei $f : V \to V$ eine lineare Abbildung auf einem $\mathbb{K}$–Vektorraum und $W \subset V$ ein f–invarianter Teilraum. Dann ist aus Proposition 10.1 offensichtlich, dass die Abbildungen $f|_W : W \to W$ und $\underline{f} : V/W \to V/W$ nicht die vollständige Information über f enthalten. In Termen einer Matrixdarstellung wie in der Proposition "verliert" man ja den Block B. Deshalb führt die Betrachtung eines einzelnen invarianten Teilraumes auch nur zu einer Darstellung durch eine obere Dreiecksmatrix wie in Bemerkung 10.1 beschrieben, die noch recht weit von einer endgültigen Normalform entfernt ist.

Um die Betrachtung einer linearen Abbildung f vollständig auf die Betrachtung von Abbildungen auf Teilräumen zu reduzieren, müssen wir Zerlegungen von V in eine direkte Summe von f–invarianten Teilräumen finden. In der Sprache von 4.9 bedeutet das gerade, dass wir invariante Teilräume finden müssen, die ein invariantes Komplement besitzen. Um gleich feinere Zerlegungen betrachten zu können, verallgemeinern wir die Resultate über direkte Summen, die wir bisher bewiesen haben, auf den Fall von mehr als zwei Summanden.

Beim Begriff der Summe von Teilräumen aus 4.8 ist die Verallgemeinerung auf endlich viele Summanden einfach: Für einen Vektorraum V und Teilräume $W_1, \dots, W_k \subset V$ definiert man $W_1 + \dots + W_k \subset V$ als den von $W_1 \cup \dots \cup W_k$ erzeugten Teilraum. Ein Element $v \in V$ liegt genau dann in diesem Teilraum, wenn es in der Form $w_1 + \dots + w_k$ für $w_j \in W_j$ geschrieben werden kann.

Definition 10.2. Sei V ein $\mathbb{K}$–Vektorraum und für $i = 1, \dots, k$ sei $W_i \subset V$ ein Teilraum. Man sagt, V ist die *direkte Summe* der Teilräume W_j, wenn einerseits $V = W_1 + \dots + W_k$ gilt, und andererseits für jedes $j = 1, \dots, k$ auch

$$W_j \cap (W_1 + \dots + W_{j-1} + W_{j+1} + \dots + W_k) = \{0\}$$

erfüllt ist. Ist das der Fall, dann schreibt man $V = W_1 \oplus \dots \oplus W_k$.

Die folgende Charakterisierung von direkten Summen verallgemeinert Teile von Proposition 4.9 und die Beschreibung von direkten Summen durch Projektionen aus Lemma 8.9.

Proposition 10.2. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$ und seien $W_1, \dots, W_k \subset V$ Teilräume. Dann sind äquivalent:*

(i) $V = W_1 \oplus \dots \oplus W_k$
(ii) *Jedes $v \in V$ kann eindeutig in der Form $v = w_1 + \dots + w_k$ mit $w_j \in W_j$ für $j = 1, \dots, k$ geschrieben werden.*

(iii) *Ist für jedes* $j = 1, \dots, k$ $\mathcal{B}_j$ *eine Basis für* W_j*, dann bilden die in* $\mathcal{B}_1, \dots, \mathcal{B}_k$ *enthaltenen Vektoren gemeinsam eine Basis für* V*.*
(iv) *Es gibt zu jedem* $j = 1, \dots, k$ *eine Basis* $\mathcal{B}_j$ *von* W_j*, sodass die in* $\mathcal{B}_1, \dots, \mathcal{B}_k$ *enthaltenen Vektoren gemeinsam eine Basis für* V *bilden.*
(v) *Für* $j = 1, \dots, k$ *gibt es lineare Abbildungen* $\pi_j : V \to V$*, die* $\pi_i \circ \pi_j = 0$ *für* $i \neq j$ *erfüllen, sodass* $\operatorname{Im}(\pi_j) = W_j$ *und* $\pi_1 + \dots + \pi_k = \mathrm{id}_V$ *ist.*

Beweis. (i) $\Rightarrow$ (ii): Da $V = W_1 + \dots + W_k$ gilt, kann jedes $v \in V$ in der Form $v = w_1 + \dots + w_k$ geschrieben werden. Sind $w_j, w_j' \in W_j$ für $j = 1, \dots, k$, sodass $w_1 + \dots + w_k = w_1' + \dots + w_k'$ gilt, dann folgt

$$w_j - w_j' = (w_1' - w_1) + \dots + (w_{j-1}' - w_{j-1}) + (w_{j+1}' - w_{j+1}) + \dots + (w_k' - w_k)$$

für alle $j = 1, \dots, k$. Nach Konstruktion liegt die linke Seite dieser Gleichung in W_j und die rechte Seite in $W_1 + \dots + W_{j-1} + W_{j+1} + \dots + W_k$. Nach Voraussetzung besteht der Durchschnitt dieser beiden Teilräume nur aus der Null, also folgt die Eindeutigkeit der Darstellung sofort.

(ii) $\Rightarrow$ (iii): Seien die Basen $\mathcal{B}_j$ für $j = 1, \dots, k$ gegeben, und sei $\mathcal{B}$ ihre Vereinigung. Nach (ii) kann jedes Element $v \in V$ in der Form $v = w_1 + \dots + w_k$ mit $w_j \in W_j$ geschrieben werden. Andererseits besitzt jedes der Elemente w_j eine Darstellung als Linearkombination von Elementen von $\mathcal{B}_j$. Damit kann aber v als Linearkombination von Elementen von $\mathcal{B}$ geschrieben werden, also ist $\mathcal{B}$ ein Erzeugendensystem für V.

Um die lineare Unabhängigkeit von $\mathcal{B}$ zu zeigen, schreiben wir $\mathcal{B}_j = \{w_1^{(j)}, \dots, w_{n_j}^{(j)}\}$ für $j = 1, \dots, k$. Seien nun $\lambda_\ell^{(j)} \in \mathbb{K}$ Skalare, sodass $0 = \sum_{j,\ell} \lambda_\ell^{(j)} w_\ell^{(j)}$ gilt. Für $j = 1, \dots, k$ setzen wir dann $w_j := \lambda_1^{(j)} w_1^{(j)} + \dots + \lambda_{n_j}^{(j)} w_{n_j}^{(j)} \in W_j$, und erhalten $w_1 + \dots + w_k = 0$. Nach der Eindeutigkeit in (ii) ist das nur möglich, wenn $w_j = 0$ für alle $j = 1, \dots, k$ gilt. Wegen der linearen Unabhängigkeit der $w_\ell^{(j)}$ ist das aber wieder nur möglich, wenn $\lambda_1^{(j)} = \dots = \lambda_{n_j}^{(j)} = 0$ gilt. Somit ist $\mathcal{B}$ linear unabhängig und damit eine Basis für V.

Da (iii) $\Rightarrow$ (iv) offensichtlich gilt, betrachten wir als nächstes (iv) $\Rightarrow$ (v): Sei $\mathcal{B}_j = \{v_1^{(j)}, \dots, v_{n_j}^{(j)}\}$ für jedes $j = 1, \dots, k$ eine Basis von W_j, sodass $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ eine Basis für V ist. Für jedes $j = 1, \dots, k$ sei $\pi_j : V \to V$ die eindeutige lineare Abbildung, die $\pi_j(v_\ell^{(j)}) = v_\ell^{(j)}$ für alle $\ell = 1, \dots, n_j$ und $\pi_j(v_\ell^{(i)}) = 0$ für alle $i \neq j$ und alle ℓ erfüllt. Aus dieser Definition folgt sofort, dass für $i \neq j$ die Abbildung $\pi_i \circ \pi_j$ alle Elemente von $\mathcal{B}$ auf 0 abbildet, während $\pi_j \circ \pi_j$ und π_j beziehungsweise $\pi_1 + \dots + \pi_k$ und id_V auf allen Elementen von $\mathcal{B}$ übereinstimmen. Damit sind die in (v) geforderten Relationen erfüllt und $\operatorname{Im}(\pi_j) = W_j$ folgt sofort aus der Definition.

(v) $\Rightarrow$ (i): Nach Voraussetzung haben wir die linearen Abbildung $\pi_j : V \to V$, und wir können π_j auch als surjektive Abbildung $V \to W_j$ betrachten. Nach Voraussetzung gilt für $v \in V$ die Gleichung $v = \pi_1(v) + \dots + \pi_k(v)$, also ist $V = W_1 + \dots + W_k$. Für $w \in W_j$ gibt es nach Voraussetzung ein Element $v \in V$, sodass $w = \pi_j(v)$ gilt. Schreibt man $w = \sum_i \pi_i(w) = \sum_i \pi_i \circ \pi_j(v)$, dann folgt $w = \pi_j \circ \pi_j(v) = \pi_j(w)$ und $0 = \pi_i(w)$ für $i \neq j$.

Ist nun $v \in W_j \cap (W_1 + \dots + W_{j-1} + W_{j+1} + \dots + W_k)$, dann gilt einerseits $\pi_j(v) = v$. Andererseits kann man v als $w_1 + \dots + w_{j-1} + w_{j+1} + \dots + w_k$ schreiben und wendet man darauf π_j an, dann erhält man 0. $\square$

Im Weiteren wird besonders die Charakterisierung durch Projektionen aus Punkt (v) wichtig sein. Man bemerke, dass aus dem letzten Teil des Beweises folgt, dass die eindeutige Darstellung von $v \in V$ als Summe von Elementen der W_j in diesem Bild durch $v = \pi_1(v) + \dots \pi_k(v)$ gegeben ist. Der große Vorteil dieser Charakterisierung ist, dass man in diesem Bild ganz leicht beschreiben kann, wann die Teilräume W_j alle invariant unter einer linearen Abbildung $f : V \to V$ sind.

SATZ 10.2. *Sei $V = W_1 \oplus \dots \oplus W_k$ eine Zerlegung eines endlichdimensionalen $\mathbb{K}$–Vektorraumes V in eine direkte Summe von Teilräumen W_j mit zugehörigen Projektionen $\pi_j : V \to V$, und sei $f : V \to V$ eine lineare Abbildung. Dann gilt:*

(1) Die Teilräume W_j sind genau dann alle f–invariant, wenn $\pi_j \circ f = f \circ \pi_j$ für alle $j = 1, \dots, k$ gilt.

(2) Ist die Bedingung aus (1) erfüllt, dann besitzt f eine Matrixdarstellung in Blockdiagonalform

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & A_k \end{pmatrix}.$$

Dabei ist für $j = 1, \dots, k$ die Matrix $A_j \in M_{n_j}(\mathbb{K})$ (mit $n_j = \dim(W_j)$) eine Matrixdarstellung der Einschränkung $f|_{W_j} : W_j \to W_j$.

BEWEIS. (1) Nehmen wir zunächst an, dass alle W_j f–invariant sind. Weil für jedes $v \in V$ das Element $\pi_i(v)$ in W_i liegt, ist dann auch $f(\pi_i(v)) \in W_i$. Wegen $v = \sum_j \pi_j(v)$ erhalten wir $f(v) = \sum_j f(\pi_j(v))$, aber auch $f(v) = \sum_j \pi_j(f(v))$. Da $V = W_1 \oplus \dots \oplus W_k$ nach Voraussetzung gilt, müssen diese beiden Darstellungen nach Teil (ii) von Proposition 10.2 übereinstimmen, also gilt $f(\pi_j(v)) = \pi_j(f(v))$ für alle j und alle $v \in V$.

Nehmen wir umgekehrt an, dass $\pi_j \circ f = f \circ \pi_j$ gilt. Für $v \in W_j$ ist dann $v = \pi_j(v)$, also $f(v) = f(\pi_j(v)) = \pi_j(f(v)) \in W_j$. Damit ist W_j ein f–invarianter Teilraum.

(2) Wir wählen für jedes $j = 1, \dots k$ eine Basis $\mathcal{B}_j$ für W_j. Nach Teil (iii) von Proposition 10.2 bildet die Vereinigung dieser Basen eine Basis $\mathcal{B}$ für V. Aus $f(W_j) \subset W_j$ für alle j folgt nun sofort, dass die Matrixdarstellung $[f]_{\mathcal{B}}$ die angegebene Blockform hat, wobei der Block A_j durch $[f|_{W_j}]_{\mathcal{B}_j}$ gegeben ist. □

BEMERKUNG 10.2. (1) Wenn wir für gegebenes $f : V \to V$ den Raum V in eine direkte Summe von f–invarianten Teilräumen W_j zerlegen können, dann genügt es also, eine schöne Matrixdarstellung für $f|_{W_j}$ für alle j zu finden. So kann man also das Problem des Findens schöner Matrixdarstellungen tatsächlich auf Teilräume reduzieren.

(2) Es gibt eine offensichtliche Umkehrung zu Teil (2) des Satzes. Angenommen, wir finden eine Basis von V bezüglich derer f eine Matrixdarstellung in Blockdiagonalform wie in Teil (2) des Satzes hat. Dann können wir die Basis als Vereinigung von Mengen $\mathcal{B}_j$ schrieben, die jeweils den Blöcken entsprechen, also aus n_j Basiselementen bestehen. Definiert man dann W_j als den von $\mathcal{B}_j$ aufgespannten Teilraum, dann folgt f–Invarianz von W_j sofort aus der Form der Matrixdarstellung und nach Proposition 10.2 ist $V = W_1 \oplus \dots \oplus W_k$.

10.3. Der Fall von diagonalisierbaren Abbildungen. Im Moment wissen wir noch nicht, wie wir für eine gegebene lineare Abbildung $f : V \to V$ eine Zerlegung von V in eine direkte Summe von f–invarianten Teilräumen finden können. Einen Hinweis darauf können wir aber erhalten, indem wir die Resultate über diagonalisierbare lineare Abbildungen aus Kapitel 7 unter dem Gesichtspunkt der invarianten Summenzerlegungen analysieren.

Nehmen wir an, dass $f : V \to V$ diagonalisierbar ist. Seien $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von f und für jedes $j = 1, \dots, k$ sei n_j die geometrische Vielfachheit von λ_j. Ist nun W_j der Eigenraum von f zum Eigenwert λ_j, dann ist $V = W_1 \oplus \dots \oplus W_k$ und jedes W_j ist f–invariant, weil ja $f(v) = \lambda_j v$ für alle $v \in W_j$ gilt.

Das charakteristische Polynom von f ist klarerweise durch

$$p_f = (\lambda_1 - x)^{n_1}(\lambda_2 - x)^{n_2} \cdots (\lambda_k - x)^{n_k}$$

gegeben. Andererseits kann der Eigenraum W_j als Kern der linearen Abbildung $f - \lambda_j \,\mathrm{id}$ realisiert werden. Betrachten wir nun zwei Abbildungen dieser Form, etwa $(f - \lambda_i \,\mathrm{id})$ und $(f - \lambda_j \,\mathrm{id})$, dann gilt

$$(f - \lambda_i \,\mathrm{id}) \circ (f - \lambda_j \,\mathrm{id}) = f \circ f - (\lambda_i + \lambda_j)f + \lambda_i\lambda_j \,\mathrm{id} = (f - \lambda_j \,\mathrm{id}) \circ (f - \lambda_i \,\mathrm{id}).$$

Schreiben wir wie in 8.4 $f \circ f =: f^2$ und $\mathrm{id}_V =: f^0$, dann ist das ein Polynom (zweiten Grades) in f. Analog hängt ein Produkt von mehreren Faktoren dieser Art nicht von der Reihenfolge ab und liefert ein Polynom in f. Betrachten wir insbesondere die Komposition $g := (f - \lambda_1 \,\mathrm{id}) \circ \ldots \circ (f - \lambda_k \,\mathrm{id}) : V \to V$, die ein Polynom (vom Grad k) in f ist. Dann gilt offensichtlich $g|_{W_k} = 0$, aber da wir die Reihenfolge der Komposition beliebig vertauschen können, folgt analog $g|_{W_j} = 0$ für alle $j = 1, \ldots, k$. Da jedes Element $v \in V$ als Summe von Elementen der W_j geschrieben werden kann, folgt $(f - \lambda_1 \,\mathrm{id}) \circ \ldots \circ (f - \lambda_k \,\mathrm{id}) = 0$. Die lineare Abbildung f erfüllt also eine Polynomgleichung!

Damit können wir aber nun relativ leicht "erraten", wie eine explizite Beschreibung der Projektionen auf die Eigenräume aussieht. Betrachten wir nämlich für ein gegebenes j die lineare Abbildung

$$g_j := (f - \lambda_1 \,\mathrm{id}) \circ \ldots \circ (f - \lambda_{j-1} \,\mathrm{id}) \circ (f - \lambda_{j+1} \,\mathrm{id}) \circ \ldots \circ (f - \lambda_k \,\mathrm{id}),$$

also die Komposition aller Faktoren außer $(f - \lambda_j \,\mathrm{id})$. Wie oben folgt dann $g_j|_{W_i} = 0$ für alle $i \neq j$. Auf W_j wirkt f durch Multiplikation mit λ_j, also sehen wir sofort, dass g_j auf W_j durch Multiplikation mit einem Skalar wirkt. Es gilt nämlich $g_j(w) = \prod_{i \neq j}(\lambda_j - \lambda_i)w$ für $w \in W_j$. Nach Voraussetzung sind die λ's alle verschieden, also können wir durch diesen Faktor dividieren und $\pi_j := (\prod_{i \neq j}(\lambda_j - \lambda_i)^{-1}) \cdot g_j$ definieren. Für $w_i \in W_i$ mit $i = 1, \ldots, k$ folgt daher $\pi_j(w_1 + \cdots + w_k) = w_j$, also ist π_j genau die Projektion auf W_j zur Summenzerlegung $W_1 \oplus \cdots \oplus W_k$.

Man kann also die Projektionen auf die Eigenräume als Polynome in der ursprünglichen linearen Abbildung f schreiben. Jedes Polynom in f kommutiert aber automatisch mit der Abbildung f. Konstruiert man also Projektionen als Polynome in f, dann erhält man automatisch f–invariante Summenzerlegungen.

Polynome von linearen Abbildungen und Matrizen

Wir werden nun die Methode, Polynome einer gegebenen linearen Abbildung $f : V \to V$ zu bilden, allgemein studieren. Es stellt sich heraus, dass man meist am besten so vorgeht, dass man hauptsächlich mit Polynomen rechnet und dann die Resultate auf lineare Abbildungen überträgt.

10.4. Grundlagen. Wir haben bereits in 8.4 diskutiert, wie man Polynome von linearen Abbildungen und Matrizen bildet. Für eine lineare Abbildung $f : V \to V$ auf einem $\mathbb{K}$–Vektorraum V definiert man $f^0 = \mathrm{id}_V$, $f^1 = f$, $f^2 = f \circ f$ und rekursiv $f^k = f^{k-1} \circ f$. Ein Polynom $p \in \mathbb{K}[x]$ kann dann als Summe $p = \sum_{i=0}^N a_i x^i$ geschrieben werden, und man definiert $p(f) := \sum_{i=0}^N a_i f^i$. Die Operationen in dieser Gleichung sind einfach die üblichen (punktweisen) Operationen auf $L(V, V)$, also ist $p(f) \in L(V, V)$ gegeben durch $p(f)(v) = \sum_i a_i f^i(v)$. Das zeigt sofort, dass die Abbildung $p(f)$ in engem Zusammenhang mit f steht. Ist etwa $v \in V$ ein Eigenvektor für f zum Eigenwert λ, dann gilt natürlich $f^k(v) = \lambda^k v$ für alle k. Daraus folgt aber sofort, dass v ein Eigenvektor für $p(f)$ zum Eigenwert $p(\lambda)$ ist (wobei wir hier die Polynomfunktion zu p auf $\lambda \in \mathbb{K}$ anwenden).

Noch einfacher ist das Anwenden von Polynomen im Bild von Matrizen zu verstehen. Für $A \in M_n(\mathbb{K})$ definiert man einfach $A^0 := \mathbb{I}$, $A^1 = A$ und A^k als die $k$te Potenz von A bezüglich der Matrizenmultiplikation. Für ein Polynom $p = \sum_{i=0}^N a_i x^i$ bildet man dann einfach $p(A) := \sum_{i=0}^N a_i A^i \in M_n(\mathbb{K})$.

Da die Matrizenmultiplikation der Komposition linearer Abbildungen entspricht, ist das Anwenden von Polynomen mit dem Bilden von Matrixdarstellungen verträglich. Ist $A = [f]_{\mathcal{B}}$ für eine Basis $\mathcal{B}$ von V, dann folgt sofort $A^k = [f^k]_{\mathcal{B}}$ (siehe Satz 4.16) und daraus wiederum $p(A) = [p(f)]_{\mathcal{B}}$ für jedes Polynom $p \in \mathbb{K}[x]$. Das impliziert auch Verträglichkeit mit Ähnlichkeit von Matrizen, die man aber auch leicht direkt einsehen kann. Sind $A, B \in M_n(\mathbb{K})$ ähnlich, dann gibt es eine invertierbare Matrix $T \in M_n(\mathbb{K})$, sodass $B = TAT^{-1}$ gilt. Daraus folgt sofort $B^2 = TA^2T^{-1}$ und induktiv $B^k = TA^kT^{-1}$ für alle $k \in \mathbb{N}$, was wiederum $p(B) = Tp(A)T^{-1}$ für jedes $p \in \mathbb{K}[x]$ liefert.

Den Schlüssel zum Verständnis der Konstruktion liefert das Studium der Funktion $p \mapsto p(f)$ (für fixes f), vergleiche mit 6.2.

Definition 10.4. Sei V ein Vektorraum über $\mathbb{K}$, $f : V \to V$ eine lineare Abbildung und $\mathbb{K}[x]$ der Ring der Polynome mit Koeffizienten in $\mathbb{K}$.

(1) Die *Auswertung auf f* ist die Funktion $\mathrm{ev}_f : \mathbb{K}[x] \to L(V, V)$, die definiert ist durch $\mathrm{ev}_f(p) := p(f)$.

(2) Mit $I_f \subset \mathbb{K}[x]$ bezeichnen wir die Teilmenge $\{p \in \mathbb{K}[x] : p(f) = 0\}$.

(3) Das Bild $\{p(f) : p \in \mathbb{K}[x]\}$ von ev_f wird mit $\mathbb{K}[f] \subset L(V, V)$ bezeichnet.

(4) Analog definieren wir für eine Matrix $A \in M_n(\mathbb{K})$ die Abbildung $\mathrm{ev}_A : \mathbb{K}[x] \to M_n(\mathbb{K})$, und die Teilmengen $I_A \subset \mathbb{K}[x]$ und $\mathbb{K}[A] \subset M_n(\mathbb{K})$.

Die Bezeichnung $\mathbb{K}[f]$ für die Menge aller linearen Abbildungen, die als Polynom in f geschrieben werden können, ist mit etwas Vorsicht zu genießen. Man muss sich klar machen, dass es sich *nicht* um einen Polynomring handelt, in dem die Variable mit f statt mit x bezeichnet wird, sondern um eine Teilmenge in $L(V, V)$ (die, wie wir bald beweisen werden, ein Teilraum und damit für endlichdimensionales V selbst endlichdimensional ist).

Die Menge $\mathbb{K}[x]$ ist ein (unendlichdimensionaler) $\mathbb{K}$–Vektorraum und ein kommutativer Ring mit Einselement unter den üblichen Operationen. Analog wissen wir, dass für einen endlichdimensionalen $\mathbb{K}$–Vektorraum V, die Menge $L(V, V)$ einen endlichdimensionalen Vektorraum und mit der punktweisen Addition und der Komposition als Multiplikation einen nicht–kommutativen Ring mit Einselement bildet. Das fundamentale Resultat ist nun, dass die Funktion ev_f mit beiden Strukturen verträglich ist.

Satz 10.4. *Sei V ein endlichdimensionaler $\mathbb{K}$–Vektorraum und $f : V \to V$ eine lineare Abbildung. Dann gilt:*

(1) Die Funktion $\mathrm{ev}_f : \mathbb{K}[x] \to L(V, V)$ ist eine lineare Abbildung und ein Ringhomomorphismus, d.h. es gilt $(p + \lambda q)(f) = p(f) + \lambda q(f)$ und $(pq)(f) = p(f) \circ q(f)$ für alle $p, q \in \mathbb{K}[x]$ und alle $\lambda \in \mathbb{K}$.

(2) Die Teilmenge $I_f \subset \mathbb{K}[x]$ ist ein Teilraum und für $p \in I_f$ und $q \in \mathbb{K}[x]$ gilt $pq \in I_f$.

(3) Das Teilmenge $\mathbb{K}[f] \subset L(V, V)$ ist ein Teilraum und ein kommutativer Teilring.

(4) Analoge Resultate gelten für $\mathrm{ev}_A : \mathbb{K}[x] \to M_n(\mathbb{K})$ sowie $I_A \subset \mathbb{K}[x]$ und $\mathbb{K}[A] \subset M_n(\mathbb{K})$.

Beweis. Die meisten dieser Eigenschaften sind ganz offensichtlich, sobald man sie sich einmal bewusst gemacht hat.

(1) Seien a_i die Koeffizienten von p und b_j die Koeffizienten von q. Dann ist $p(f) = \sum_{i=0}^N a_i f^i$ und $q(f) = \sum_{i=0}^N b_i f^i$, wobei wir N als das Maximum der Grade von p und q definieren und Koeffizienten über dem jeweiligen Grad Null setzen. Damit ist aber offensichtlich $p(f) + \lambda q(f) = \sum_{i=0}^N (a_i + \lambda b_i) f^i = (p + \lambda q)(f)$, also ist ev_f linear. Nach Definition ist $x^i x^j = x^{i+j}$ in $\mathbb{K}[x]$ und $f^i \circ f^j = f^{i+j}$ in $L(V, V)$. Da beide Multiplikationen distributiv bezüglich der Addition sind, impliziert das sofort $(pq)(f) = p(f) \circ q(f)$.

(2) Nach (1) ist ev_f eine lineare Abbildung und damit ist I_f nach Definition der Kern von ev_f und somit ein Teilraum von $\mathbb{K}[x]$. Ist $p \in I_f$ und $q \in \mathbb{K}[x]$, dann ist nach Teil (1) $(pq)(f) = p(f) \circ q(f) = 0 \circ q(f) = 0$, also $pq \in I_f$.

(3) Als Bild einer linearen Abbildung ist $\mathbb{K}[f]$ ein Teilraum von $L(V, V)$. Sind $g, h \in \mathbb{K}[f]$, dann gibt es Polynome $p, q \in \mathbb{K}[x]$, sodass $g = p(f)$ und $h = q(f)$ gilt. Damit ist aber $h \circ g = q(f) \circ p(f) = (qp)(f) \in \mathbb{K}[f]$, also ist $\mathbb{K}[f]$ ein Teilring von $L(V, V)$. Andererseits ist $qp = pq$ in $\mathbb{K}[x]$ und damit $h \circ g = (qp)(f) = (pq)(f) = g \circ h$, also ist $\mathbb{K}[f]$ ein kommutativer Ring.

Die Beweise der Behauptung in (4) sind völlig analog zu den Beweisen von (1)–(3). □

10.5. Der Hauptidealsatz. Interessanterweise führt der Weg zum Verständnis des endlichdimensionalen Raumes $\mathbb{K}[f]$ über den Teilraum $I_f \subset \mathbb{K}[x]$, der überraschend einfach zu verstehen ist. Wesentlich ist hier die in Teil (2) von Satz 10.4 beobachtete Eigenschaft von I_f bezüglich der Multiplikation, die zu einem der wichtigsten Begriffe der Ringtheorie führt.

Definition 10.5. Sei $(R, +, \cdot)$ ein kommutativer Ring mit Einselement. Ein *Ideal* in R ist eine Teilmenge $I \subset R$, sodass I eine Untergruppe der kommutativen Gruppe $(R, +)$ ist und für beliebige Elemente $s \in I$ und $r \in R$ immer $rs \in I$ gilt.

Ein Ideal $I \subset R$ ist also insbesondere ein Teilring. In Teil (2) von Satz 10.4 haben wir gezeigt, dass $I_f \subset \mathbb{K}[x]$ ein Ideal ist. Analog zum Beweis dort sieht man sofort, dass der Kern eines Homomorphismus zwischen kommutativen Ringen immer ein Ideal ist. Man bemerke, dass für ein Ideal $I \subset R$, das das Einselement von R enthält, automatisch $I = R$ gelten muss. Allgemeiner gilt, dass ein Ideal, das ein invertierbares Element von R enthält, schon mit R übereinstimmt.

Mit Hilfe der Division mit Rest aus 7.6 können wir nun Ideale in $\mathbb{K}[x]$ vollständig beschreiben. Dazu erinnern wir uns noch, dass wir in 7.6 ein Polynom *monisch* genannt haben, wenn sein führender Koeffizient (also der Koeffizient der höchsten Potenz, der ungleich Null ist) gleich Eins ist.

Satz 10.5 (Hauptidealsatz). *Sei $\mathbb{K}$ ein Körper. Ist $p \in \mathbb{K}[x]$ ein beliebiges Polynom, dann ist die Teilmenge $p\mathbb{K}[x] := \{pq : q \in \mathbb{K}[x]\}$ ein Ideal in $\mathbb{K}[x]$.*

Ist umgekehrt $\{0\} \neq I \subset \mathbb{K}[x]$ ein beliebiges Ideal, dann gibt es ein eindeutig bestimmtes monisches Polynom $p \in \mathbb{K}[x]$, sodass $I = p\mathbb{K}[x]$ gilt.

Beweis. Da $pq_1 + pq_2 = p(q_1 + q_2)$, $0 = p \cdot 0$, $-(pq) = p(-q)$ und $(pq_1)q_2 = p(q_1 q_2)$ gelten, ist $p\mathbb{K}[x]$ ein Ideal. Sei umgekehrt $I \subset \mathbb{K}[x]$ ein Ideal. Falls I ein konstantes Polynom ungleich Null enthält, dann ist $I = \mathbb{K}[x] = 1\mathbb{K}[x]$, weil so ein Polynom in $\mathbb{K}[x]$ invertierbar ist. Ansonsten haben Polynome ungleich Null, die in I liegen, positiven Grad und wir wählen ein Polynom $0 \neq \tilde{p} \in I$ für das der Grad minimal ist. Ist a der führende Koeffizient von $\tilde{p}$, dann liegt auch $p = a^{-1}\tilde{p}$ in I, und nach Konstruktion ist p monisch.

Da $p \in I$ gilt und I ein Ideal ist, ist offensichtlich $p\mathbb{K}[x] \subset I$. Ist andererseits $q \in I$, dann dividieren wir q mit Rest durch p. Nach Lemma 7.6 finden wir Polynome $\tilde{q}, r \in \mathbb{K}[x]$, sodass $q = p\tilde{q} + r$ und $r = 0$ oder $\deg(r) < \deg(p)$ gilt. Da $q \in I$ und wegen $p \in I$ auch $p\tilde{q} \in I$ gilt, ist auch $q - p\tilde{q} = r \in I$. Wäre nun $r \neq 0$, dann wäre das (wegen $\deg(r) < \deg(p)$) ein Widerspruch zur Konstruktion von p. Damit ist $q \in p\mathbb{K}[x]$, also $I = p\mathbb{K}[x]$.

Um die Eindeutigkeit von p zu beweisen, nehmen wir an, dass $p_1, p_2 \in \mathbb{K}[x]$ monische Polynome sind, sodass $p_1\mathbb{K}[x] = p_2\mathbb{K}[x]$ gilt. Dann gibt es Polynome $q_1, q_2 \in \mathbb{K}[x]$, sodass $p_2 = p_1 q_1$ und $p_1 = p_2 q_2$ gilt. Insbesondere folgt daraus $\deg(p_2) \geq \deg(p_1)$ und $\deg(p_1) \geq \deg(p_2)$, also haben p_1 und p_2 gleichen Grad. Damit müssen aber q_1 und q_2 Grad Null haben, also in $\mathbb{K} \subset \mathbb{K}[x]$ liegen. Da aber beide Polynome monisch sind, folgt $q_1 = q_2 = 1$, also $p_1 = p_2$. □

Bemerkung 10.5. (1) Der erste Teil des Satzes ist natürlich gar nicht spezifisch für den Ring $\mathbb{K}[x]$. Ist R ein beliebiger kommutativer Ring mit Eins, dann ist für jedes Element $r \in R$ die Teilmenge $rR = \{rs : s \in R\}$ ein Ideal in R. Ideale dieser Form nennt man in der Algebra *Hauptideale*. Unser Satz sagt also, dass jedes Ideal im Ring $\mathbb{K}[x]$ ein Hauptideal ist, und Ringe, in denen das gilt, nennt man *Hauptidealbereiche*.

(2) Analog zum Beweis des Satzes kann man die Division in $\mathbb{Z}$ benutzen um zu zeigen, dass jedes Ideal in $\mathbb{Z}$ ein Hauptideal, also von der Form $m\mathbb{Z}$ für $m \in \mathbb{N}$ ist. Man kann auch leicht sehen, dass jede Untergruppe von $(\mathbb{Z}, +)$ automatisch ein Ideal ist, also liefert das eine Beschreibung aller Untergruppen von $\mathbb{Z}$.

(3) Der Hauptidealsatz liefert auch sofort die algebraische Charakterisierung von Nullstellen in Satz 7.6. Für eine Zahl $\lambda \in \mathbb{K}$ ist natürlich $I := \{p \in \mathbb{K}[x] : p(\lambda) = 0\}$ ein Ideal in $\mathbb{K}[x]$, und das monische Polynom $x - \lambda$ liegt in diesem Ideal. Damit ist aber $I = (x - \lambda)\mathbb{K}[x]$.

10.6. Minimalpolynom und Struktur von $\mathbb{K}[f]$. Wir können nun den Hauptidealsatz auf das Ideal $I_f \subset \mathbb{K}[x]$ anwenden, das durch eine lineare Abbildung $f : V \to V$ definiert wird. Dazu bemerken wir einerseits, dass I_f kein konstantes Polynom ungleich Null enthalten kann, da so ein Polynom durch ev_f auf ein nichttriviales Vielfaches der Identitätsabbildung id_V abgebildet wird. Andererseits hat der Raum $L(V, V)$ Dimension n^2, wobei $n = \dim(V)$ ist. Damit müssen die Elemente $\mathrm{id}_v, f, f^2, \dots, f^{n^2}$ auf jeden Fall linear abhängig sein. Schreibt man 0 als nichttriviale Linearkombination dieser Abbildungen, dann erhält man ein Polynom $p \in I_f$ mit $\deg(p) \leq n^2$. (Wir werden in 10.7 sehen, dass das charakteristische Polynom p_f von f in I_f liegt, also findet man immer Polynome vom Grad $\leq n$ in I_f.)

Definition 10.6. Sei $f : V \to V$ eine lineare Abbildung. Das eindeutig bestimmte monische Polynom $p \in \mathbb{K}[x]$ sodass $I_f = p\mathbb{K}[x]$ gilt, heißt das *Minimalpolynom* von f und wird mit m_f bezeichnet.

Nach Definition gilt also $m_f(f) = 0$ und ein Polynom $p \in \mathbb{K}[x]$ erfüllt genau dann $p(f) = 0$, wenn es ein Polynom $q \in \mathbb{K}[x]$ gibt, sodass $p = qm_f$ gilt.

Damit können wir nun auch den Teilraum $\mathbb{K}[f] \subset L(V, V)$ genauer beschreiben. Diese Beschreibung wird für die folgenden theoretischen Überlegungen kaum benötigt, sie dient hauptsächlich dazu, dass man sich besser vorstellen kann, wie "die Dinge aussehen".

Proposition 10.6. *Sei $f : V \to V$ eine lineare Abbildung mit Minimalpolynom $m_f := x^k + a_{k-1}x^{k-1} + \dots + a_1 x + a_0$. Dann bilden die linearen Abbildungen*

$\mathrm{id}_V, f, f^2, \dots, f^{k-1}$ *eine Basis* $\mathcal{B}$ *für den Teilraum* $\mathbb{K}[f] \subset L(V,V)$. *Die Komposition auf diesem Teilraum ist vollständig bestimmt durch* $f^i \circ f^j = f^{i+j}$ *und* $f^k = -\sum_{\ell=0}^{k-1} a_\ell f^\ell$.

Beweis. Nach Definition bilden die Elemente f^i für $i \in \mathbb{N}$ ein Erzeugendensystem für $\mathbb{K}[f]$. Wegen $m_f(f) = 0$ folgt, dass $f^k = -\sum_{\ell=0}^{k-1} a_\ell f^\ell$ gilt, also liegt f^k im Erzeugnis von $\mathcal{B}$. Komponiert man diese Gleichung mit f, dann erhält man eine Darstellung von f^{k+1} als Linearkombination von $f, f^2, \dots, f^k$, also liegt auch das im Erzeugnis von $\mathcal{B}$. Induktiv folgt nun, dass alle f^i im Erzeugnis von $\mathcal{B}$ liegen, also ist $\mathcal{B}$ ein Erzeugendensystem für $\mathbb{K}[f]$.

Sind $\lambda_0, \dots, \lambda_{k-1} \in \mathbb{K}$ Skalare, sodass $0 = \sum_{i=0}^{k-1} \lambda_i f^i$ gilt, dann bedeutet das $p(f) = 0$, wobei man $p = \sum_{i=0}^{k-1} \lambda_i x^i$ setzt. Damit gilt aber $p \in I_f = m_f \mathbb{K}[x]$ und wäre $p \neq 0$, dann wäre $\deg(p) < \deg(m_f)$ ein Widerspruch. Damit muss $p = 0$ gelten, also sind $\mathrm{id}_V, f, \dots, f^{k-1}$ linear unabhängig und somit eine Basis für $\mathbb{K}[f]$.

Wie oben beschrieben, kann man aus $f^k = -\sum_{\ell=0}^{k-1} a_\ell f^\ell$ eine Darstellung von f^i als Linearkombination von Elementen von $\mathcal{B}$ für alle $i > k$ ableiten, also folgt die letzte Behauptung. □

10.7. Der Satz von Cayley–Hamilton. Der Schlüssel zum Verständnis des Minimalpolynoms ist der Satz von Cayley–Hamilton, der sagt, dass das Minimalpolynom m_f einer linearen Abbildung f ein Teiler des charakteristischen Polynoms p_f ist.

Satz 10.7 (Cayley–Hamilton). *Ist* $\mathbb{K}$ *ein Körper und* $A \in M_n(\mathbb{K})$ *eine* $n \times n$*–Matrix mit charakteristischem Polynom* $p_A \in \mathbb{K}[x]$, *dann ist* $p_A(A) = 0$. *Analog gilt für jede lineare Abbildung* $f : V \to V$ *auf einem endlichdimensionalen* $\mathbb{K}$*–Vektorraum* V *mit charakteristischem Polynom* p_f *die Gleichung* $p_f(f) = 0$.

Beweis. Ist $\mathcal{B}$ eine Basis für V und $A = [f]_\mathcal{B}$, dann ist $p_A = p_f$, und aus 10.4 wissen wir, dass dann $p_A(A) = p_f(A) = [p_f(f)]_\mathcal{B}$ gilt. Damit genügt es, die Matrizenversion des Resultats zu beweisen. Weiters können wir $A \in M_n(\mathbb{K})$ auch als Matrix über dem algebraischen Abschluss $\tilde{\mathbb{K}}$ von $\mathbb{K}$ betrachten. Über $\tilde{\mathbb{K}}$ ist A aber nach Proposition 7.13 ähnlich zu einer oberen Dreiecksmatrix B und natürlich ist $p_B = p_A$. Wiederum aus 10.4 wissen wir, dass dann $p_A(A)$ ähnlich zu $p_B(B)$ ist. Somit genügt es, das Resultat für obere Dreiecksmatrizen zu beweisen, also dürfen wir ohne Beschränkung der Allgemeinheit annehmen, dass $A = (a_{ij})$ selbst eine obere Dreiecksmatrix ist.

Unter dieser Annahme ist aber $p_A = (a_{11} - x) \cdots (a_{nn} - x)$ und wir müssen zeigen, dass $(a_{11}\mathbb{I} - A) \cdots (a_{nn}\mathbb{I} - A) = 0$ gilt. Wir behaupten, dass $(a_{11}\mathbb{I} - A) \cdots (a_{ii}\mathbb{I} - A)e_j = 0$ für $j = 1, \dots, i$ gilt, wobei $\{e_1, \dots, e_n\}$ die Standardbasis von $\mathbb{K}^n$ bezeichnet und beweisen das mit Induktion nach i. Für $i = 1$ bemerken wir, dass nach Konstruktion die erste Spalte von $a_{11}\mathbb{I} - A$ nur aus Nullen besteht, also bildet diese Matrix e_1 auf 0 ab.

Nehmen wir induktiv an, dass $(a_{11}\mathbb{I} - A) \cdots (a_{i-1,i-1}\mathbb{I} - A)e_j = 0$ für $j = 1, \dots, i-1$ gilt. Dann ist aber $Ae_i = a_{1i}e_1 + \cdots + a_{ii}e_i$ und damit $(a_{ii}\mathbb{I} - A)e_i$ eine Linearkombination von $e_1, \dots, e_{i-1}$, also ist $(a_{11}\mathbb{I} - A) \cdots (a_{ii}\mathbb{I} - A)e_i$ nach Induktionsvoraussetzung gleich Null. Andererseits ist

$$(a_{11}\mathbb{I} - A) \cdots (a_{i-1,i-1}\mathbb{I} - A)(a_{ii}\mathbb{I} - A) = (a_{ii}\mathbb{I} - A)(a_{11}\mathbb{I} - A) \cdots (a_{i-1,i-1}\mathbb{I} - A),$$

also werden auch die e_j für $j < i$ auf Null abgebildet. Nach Induktion ist somit

$$(a_{11}\mathbb{I} - A) \cdots (a_{nn}\mathbb{I} - A)e_j = 0$$

für alle $j = 1, \dots, n$ und somit ist das Produkt die Nullmatrix. □

Bemerkung 10.7. Der obige Beweis ist etwas unbefriedigend, weil er die Existenz des algebraischen Abschlusses eines Körpers voraussetzt, die ein ziemlich tief liegendes algebraisches Resultat ist.

Es gibt einen alternativen Beweis, der die Existenz des algebraischen Abschlusses nicht benötigt, dafür aber eher raffinierte Interpretationen braucht. Er folgt in gewissem Sinn der naiven (und falschen) Idee, dass ja $p_A = \det(A - x\mathbb{I})$ ist und die rechte Seite dieser Gleichung verschwindet, wenn man formal $x = A$ setzt.

Dieser alternative Beweis beruht auf einer Interpretation von $p_A(A)$ als Determinante. Dazu muss man über dem kommutativen Ring $\mathbb{K}[A]$ mit Eins arbeiten. Betrachten wir die Matrix $C = (c_{ij}) \in M_n(\mathbb{K}[A])$ die definiert ist durch $c_{ii} = a_{ii}\mathbb{I} - A$ und $c_{ij} = a_{ij}\mathbb{I}$ für $i \neq j$. Berechnen wir nun $\det(C)$ nach der Leibniz Formel, also $\det(C) = \sum_{\sigma\in\mathfrak{S}_n} \operatorname{sgn}(\sigma) c_{\sigma(1)1} \cdots c_{\sigma(n)n}$, dann folgt aus der Definition von C sofort, dass $\det(C) = p_A(A) \in \mathbb{K}[A]$ gilt. Um den Satz von Cayley–Hamilton zu beweisen, müssen wir also $\det(C) = 0$ zeigen.

Da $c_{ij} \in \mathbb{K}[A] \subset M_n(\mathbb{K})$ gilt, können wir $\sum_i c_{ij}e_i \in \mathbb{K}^n$ betrachten. Nach Definition ist das gleich $(\sum_i a_{ij}e_i) - Ae_j = 0$, also gilt $\sum_i c_{ij}e_i = 0$ für alle $j = 1, \dots, n$. Bezeichnen wir nun mit $C_{ij} \in M_{n-1}(\mathbb{K}[A])$ die Matrix, die entsteht, wenn man in C die $i$te Zeile und die $j$te Spalte streicht. Dann gilt nach Satz 6.10, dass $\sum_\ell (-1)^{j+\ell} c_{i\ell} \det(C_{j\ell})$ für $j = i$ gerade $\det(C)$ und für $i \neq j$ gleich Null ist. Wegen der Kommutativität von $\mathbb{K}[A]$ erhalten wir damit

$$\begin{aligned}(p_A(A))(e_j) &= \textstyle\sum_\ell (-1)^{j+\ell} \det(C_{j\ell})(c_{j\ell}(e_j))\\ 0 = 0(e_i) &= \textstyle\sum_\ell (-1)^{j+\ell} \det(C_{j\ell})(c_{i\ell}(e_i)) \quad \text{für } i \neq j.\end{aligned}$$

Addieren wir diese n Gleichungen, dann erhalten wir aber

$$(p_A(A))(e_j) = \textstyle\sum_i \sum_\ell (-1)^{j+\ell} \det(C_{j\ell})(c_{i\ell}(e_i)) = \sum_\ell (-1)^{j+\ell} \det(C_{j\ell})(\sum_i c_{i\ell}(e_i)) = 0.$$

Da das für alle $j = 1, \dots, n$ gilt, ist $p_A(A) = 0$.

Primfaktorzerlegung und Primärzerlegung

Für eine lineare Abbildung $f : V \to V$ können wir nun eine Zerlegung von V in eine direkte Summe f–invarianter Teilräume finden, indem wir ein Analogon der Primfaktorzerlegung für das Minimalpolynom m_f (oder das charakteristische Polynom p_f) finden.

10.8. Primfaktorzerlegung von Polynomen. Die bekannten Begriffe von Teilbarkeit, Primzahlen und dem größten gemeinsamen Teiler lassen sich nun relativ leicht auf Polynome übertragen.

Definition 10.8. Sei $\mathbb{K}$ ein Körper und seien $p, p_1, p_2 \in \mathbb{K}[x]$ Polynome.

(1) Man sagt p_1 teilt p_2, wenn es ein Polynom $q \in \mathbb{K}[x]$ gibt, sodass $p_2 = p_1q$ gilt, oder äquivalent wenn p_2 im Hauptideal $p_1\mathbb{K}[x]$ liegt.

(2) Man sagt, das Polynom p ist *irreduzibel*, wenn es keine Polynome q_1 und q_2 gibt, die beide echt kleineren Grad als p haben, sodass $p = q_1q_2$ gilt.

Beispiel 10.8. (1) Polynome ersten Grades sind immer irreduzibel, weil wegen $\deg(q_1q_2) = \deg(q_1) + \deg(q_2)$ ein Produkt nur dann Grad eins haben kann, wenn einer der beiden Faktoren Grad eins hat.

Über einem algebraisch abgeschlossenen Körper hat jedes Polynom $p \in \mathbb{K}[x]$ von positivem Grad eine Nullstelle. Nach Satz 7.7 hat p damit einen Teiler mit Grad 1 und kann somit nur dann irreduzibel sein, wenn $\deg(p) = 1$ gilt.

(2) Das Polynom $p = x^2 + 1 \in \mathbb{R}[x]$ ist irreduzibel. Wäre das nicht der Fall, dann wäre $p = q_1 q_2$ mit $\deg(q_1) = \deg(q_2) = 1$. Natürlich würde das zwei Nullstellen von p liefern, die es in $\mathbb{R}$ nicht gibt.

Allgemeiner kann man Korollar 7.12 so interpretieren, dass ein Polynom über $\mathbb{R}$ genau dann irreduzibel ist, wenn es entweder Grad eins hat, oder Grad zwei hat, aber keine reelle Nullstelle besitzt.

(3) Für allgemeine Körper gibt es irreduzible Polynome höheren Grades. So ist etwa $x^3 - 2 \in \mathbb{Q}[x]$ irreduzibel. Könnte man dieses Polynom nämlich in ein Produkt zerlegen, dann müsste einer der Faktoren Grad eins haben, was zur Existenz einer Nullstelle in $\mathbb{Q}$ führen würde. Irreduzible Polynome über $\mathbb{Q}$ allgemein zu verstehen, ist ein schwieriges Problem.

Für den Begriff des größten gemeinsamen Teilers zweier Polynome $0 \neq p_1, p_2 \in \mathbb{K}[x]$ betrachten wir die Teilmenge $I \subset \mathbb{K}[x]$, all jener Polynome, die man in der Form $q_1 p_1 + q_2 p_2$ schreiben kann. Man verifiziert sofort, dass I ein Ideal in $\mathbb{K}[x]$ ist, also gibt es nach dem Hauptidealsatz (Satz 10.5) ein eindeutig bestimmtes monisches Polynom $q \in \mathbb{K}[x]$ sodass $I = q\mathbb{K}[x]$ gilt. Nun ist aber nach Definition $p_1, p_2 \in I$, also q ein gemeinsamer Teiler von p_1 und p_2. Ist umgekehrt r ein gemeinsamer Teiler von p_1 und p_2, dann ist $p_1 = s_1 r$ und $p_2 = s_2 r$ für passendes $s_1, s_2 \in \mathbb{K}[x]$. Für $q_1, q_2 \in \mathbb{K}[x]$ ist damit $q_1 p_1 + q_2 p_2 = (q_1 s_1 + q_2 s_2) r$, also teilt r jedes Element von I und damit insbesondere q. Somit ist q tatsächlich der größte gemeinsame Teiler von p_1 und p_2, und man schreibt $q = \operatorname{ggT}(p_1, p_2)$.

Wie bei ganzen Zahlen, nennt man $p_1, p_2 \in \mathbb{K}[x]$ *relativ prim*, wenn $\operatorname{ggT}(p_1, p_2) = 1$ gilt. In diesem Fall gibt es somit Polynome $q_1, q_2 \in \mathbb{K}[x]$, sodass $q_1 p_1 + q_2 p_2 = 1$ gilt. Wir können nun einige grundlegende Resultate von den ganzen Zahlen auf Polynome übertragen.

Lemma 10.8. *Sei $\mathbb{K}$ ein Körper und $p, q_1, q_2 \in \mathbb{K}[x]$ Polynome. Ist p irreduzibel und teilt das Produkt $q_1 q_2$, dann teilt p einen der beiden Faktoren.*

Beweis. Nehmen wir an, dass p das Produkt $q_1 q_2$, aber nicht den Faktor q_1 teilt. Dann betrachten wir $s := \operatorname{ggT}(p, q_1)$ und bemerken, dass $\deg(s) \leq \deg(p)$ gilt. Wäre $\deg(s) = \deg(p)$, dann wäre $s = a^{-1} p$, wobei a der führende Koeffizient von p ist. Nach Konstruktion ist aber s ein Teiler von q_1 und damit wäre auch $p = as$ ein Teiler von q_1, ein Widerspruch. Also ist $\deg(s) < \deg(p)$ und weil p irreduzibel ist, folgt $\deg(s) = 0$, also $s = 1$. Damit gibt es aber Polynome $\tilde{p}$ und $\tilde{q}$, sodass $p\tilde{p} + q_1 \tilde{q} = 1$ und damit $p\tilde{p}q_2 + \tilde{q}q_1 q_2 = q_2$ gilt. Nun ist aber nach Voraussetzung p ein Teiler des Produktes $q_1 q_2$, also gibt es ein Polynom r, sodass $q_1 q_2 = pr$ gilt. Damit liefert aber die obige Gleichung $q_2 = p(\tilde{p}q_2 + \tilde{q}r)$, also teilt p den Faktor q_2. □

Induktiv folgt aus diesem Lemma sofort, dass ein irreduzibles Polynom, das ein Produkt von endlich vielen Polynomen teilt, mindestens eines dieser Polynome teilen muss.

Nun haben wir alle Zutaten gesammelt, um die eindeutige Primfaktorzerlegung für Polynome zu beweisen. Das Problem, dass man skalare Faktoren beliebig auf die einzelnen Primfaktoren verteilen kann, löst man dadurch, dass man sich auf monische Primfaktoren beschränkt.

Satz 10.8 (Eindeutige Primfaktorzerlegung). *Sei $\mathbb{K}$ ein Körper und sei $p \in \mathbb{K}[x]$ ein Polynom mit $\deg(p) > 0$. Dann gibt es paarweise verschiedene, irreduzible, monische Polynome $p_1, \dots, p_k \in \mathbb{K}[x]$, Elemente $n_1, \dots, n_k \in \mathbb{N} \setminus \{0\}$ und ein Element $c \in \mathbb{K} \setminus \{0\}$,* *sodass $p = cp_1^{n_1} \cdots p_k^{n_k}$ gilt. Diese Darstellung ist bis auf die Reihenfolge der Faktoren eindeutig bestimmt.*

Beweis. Zunächst beweisen wir die Existenz der Zerlegung durch Induktion nach $n = \deg(p)$. Ist $n = 1$, dann ist p irreduzibel. Ist c der führende Koeffizient von p, dann ist $p_1 := c^{-1}p$ irreduzibel und monisch, und $p = cp_1$ ist die gesuchte Zerlegung.

Nehmen wir also an, dass $n \geq 2$ gilt, und die Existenz der Zerlegung für alle $k < n$ schon bewiesen wurde. Ist p irreduzibel, dann liefert wieder $p = c(c^{-1}p)$ die gesuchte Zerlegung, wobei c der führende Koeffizient von p ist. Ist p nicht irreduzibel, dann gibt es Polynome $q_1, q_2 \in \mathbb{K}[x]$ mit $\deg(q_1), \deg(q_2) < \deg(p)$, sodass $p = q_1q_2$ gilt. Nach Induktionsvoraussetzung können wir q_1 und q_2 in der gewünschten Form darstellen, und durch Multiplizieren und Zusammenfassen gemeinsamer Faktoren erhalten wir die gesuchte Darstellung für p.

Zur Eindeutigkeit: Da der Faktor c offensichtlich gerade der führende Koeffizient des Produktes $cp_1^{n_1} \dots p_k^{n_k}$ ist, ist dieser Faktor eindeutig bestimmt. Es genügt also zu zeigen, dass für (nicht notwendigerweise verschiedene) monische irreduzible Polynome $p_1, \dots, p_n, q_1, \dots, q_m$ die Gleichung $p_1 \cdots p_n = q_1 \cdots q_m$ nur dann gelten kann, wenn $n = m$ gilt und sich die Produkte nur durch die Reihenfolge der Faktoren unterscheiden. Wir dürfen ohne Beschränkung der Allgemeinheit $n \leq m$ annehmen und führen den Beweis nun durch Induktion nach n. Für $n = 1$ ist das Resultat offensichtlich, weil alle auftretenden Polynome monisch und irreduzibel, also vom Grad ≥ 1, sind.

Für $n > 1$ sehen wir, dass p_1 das Produkt $q_1 \cdots q_m$ teilt, also gibt es nach dem Lemma ein i, sodass p_1 den Faktor q_i teilt. Weil q_i irreduzibel und $\deg(p_1) > 0$ ist, muss $p_1 = q_i$ gelten. Damit erhalten wir aber $p_1(p_2 \cdots p_n - q_1 \cdots q_{i-1}q_{i+1} \cdots q_m) = 0$. Das kann aber nur sein, wenn die Klammer verschwindet, also erhalten wir $p_2 \cdots p_n = q_1 \cdots q_{i-1}q_{i+1} \cdots q_m$ und die Behauptung folgt nach Induktion. □

10.9. Die Primärzerlegung. Für eine lineare Abbildung $f : V \to V$ auf einem endlichdimensionalen $\mathbb{K}$–Vektorraum V können wir nun parallel zur Primfaktorzerlegung des Minimalpolynoms m_f den Vektorraum V in eine direkte Summe f–invarianter Teilräume zerlegen. Der technische Schlüssel dazu ist folgendes Resultat.

Lemma 10.9. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung. Seien $p_1, p_2 \in \mathbb{K}[x]$ Polynome, die relativ prim sind und $(p_1p_2)(f) = 0$ erfüllen.*

Dann definiert $W_i := \operatorname{Ker}(p_i(f))$ für $i = 1, 2$ einen f–invarianten Teilraum von V und $V = W_1 \oplus W_2$.

Beweis. Da p_1 und p_2 relativ prim sind, finden wir nach 10.8 Polynome $q_1, q_2 \in \mathbb{K}[x]$, sodass $q_1p_1 + q_2p_2 = 1$ gilt. Setzen wir nun $\pi_i = (q_ip_i)(f) \in L(V, V)$ für $i = 1, 2$, dann bedeutet diese Gleichung gerade $\pi_1 + \pi_2 = \mathrm{id}_V$. Außerdem ist

$$\pi_2 \circ \pi_1 = (q_2p_2)(f) \circ (q_1p_1)(f) = (q_1q_2)(f) \circ (p_1p_2)(f) = 0.$$

Weil $\pi_1, \pi_2, f \in \mathbb{K}[f] \subset L(V, V)$ und das ein kommutativer Teilring ist (siehe Proposition 10.6), gilt auch $\pi_1 \circ \pi_2 = 0$ sowie $\pi_i \circ f = f \circ \pi_i$ für $i = 1, 2$. Damit ist nach Proposition 10.2 $V = \operatorname{Im}(\pi_1) \oplus \operatorname{Im}(\pi_2)$ und nach Satz 10.2 ist das eine Zerlegung in f–invariante Teilräume.

Wegen $p_2(f) \circ \pi_1 = (p_2q_1p_1)(f) = 0$ gilt $\operatorname{Im}(\pi_1) \subset W_2 = \operatorname{Ker}(p_2(f))$. Andererseits ist $W_2 = \operatorname{Ker}(p_2(f)) \subset \operatorname{Ker}(q_2(f) \circ p_2(f)) = \operatorname{Ker}(\pi_2)$. Für $v \in W_2$ ist damit aber $v = \pi_1(v) + \pi_2(v) = \pi_1(v)$, also $W_2 \subset \operatorname{Im}(\pi_1)$. Damit gilt aber $W_2 = \operatorname{Im}(\pi_1)$ und analog folgt $W_1 = \operatorname{Im}(\pi_2)$. □

Damit können wir nun die allgemeine Form der Primärzerlegung beweisen.

SATZ 10.9. *Sei V ein endlichdimensionaler $\mathbb{K}$–Vektorraum, $f : V \to V$ eine lineare Abbildung und $p \in \mathbb{K}[x]$ ein Polynom, sodass $p(f) = 0$ gilt. Betrachte die eindeutige Primfaktorzerlegung von p, d.h. die eindeutige Darstellung in der Form $p = cp_1^{n_1} \cdots p_k^{n_k}$ für paarweise verschiedene monische irreduzible Polynome $p_j \in \mathbb{K}[x]$, Exponenten $n_j \in \mathbb{N} \setminus \{0\}$ und ein Element $c \in \mathbb{K} \setminus \{0\}$.*

Setzt man $W_i := \operatorname{Ker}(p_i(f)^{n_i})$ für $i = 1, \dots, k$, dann ist jedes W_i ein f–invarianter Teilraum von V und $V = W_1 \oplus \cdots \oplus W_k$.

BEWEIS. Durch Induktion nach der Anzahl k der verschiedenen Primfaktoren von p. Für $k = 1$ ist $p = cp_1^{n_1}$, wobei c der führende Koeffizient von p ist. Nach Voraussetzung ist $p(f) = 0$, also $V = \operatorname{Ker}(p(f))$. Alle anderen Aussagen sind trivial.

Ist $k \geq 2$, dann setzen wir $q_1 = p_1^{n_1}$ und $q_2 = p_2^{n_2} \cdots p_k^{n_k}$. Dann ist $(q_1q_2)(f) = p(f) = 0$. Außerdem sind q_1 und q_2 relativ prim, weil die p_i paarweise verschiedene irreduzible Polynome sind. Ein irreduzibler Teiler von q_1 muss nämlich nach Lemma 10.8 p_1 teilen, also gleich p_1 sein. Analog muss ein irreduzibler Teiler von q_2 eines der p_i für $i = 2, \dots, k$ teilen, also gleich diesem Polynom sein.

Damit liefert Lemma 10.9 eine Zerlegung $V = W_1 \oplus \tilde{W}_2$ in f–invariante Teilräume. Hier ist $W_1 = \operatorname{Ker}(p_1(f)^{n_1})$ und $\tilde{W}_2 = \operatorname{Ker}((p_2^{n_2} \cdots p_k^{n_k})(f))$. Seien π_1 und $\tilde{\pi}_2$ die entsprechenden Projektionen.

Ist nun $\tilde{f}$ die Einschränkung von f auf $\tilde{W}_2$, dann gilt $(p_2^{n_2} \cdots p_k^{n_k})(\tilde{f}) = 0$ nach Konstruktion, also ist nach Induktionsvoraussetzung $\tilde{W}_2 = W_2 \oplus \cdots \oplus W_k$, wobei $W_i = \operatorname{Ker}(p_i^{n_i}(\tilde{f}))$. Sind $\bar{\pi}_2, \dots, \bar{\pi}_k$ die entsprechenden Projektionen, dann definieren wir für $i = 2, \dots, k$ die linearen Abbildungen $\pi_i := \bar{\pi}_i \circ \tilde{\pi}_2$. Betrachten wir nun die Familie $\{\pi_1, \dots, \pi_k\}$, dann ist nach Konstruktion

$$\pi_1 + \cdots + \pi_k = \pi_1 + ((\bar{\pi}_2 + \cdots + \bar{\pi}_k) \circ \tilde{\pi}_2) = \pi_1 + \mathrm{id}_{\tilde{W}_2} \circ \tilde{\pi}_2 = \pi_1 + \tilde{\pi}_2 = \mathrm{id}_V .$$

Für $i = 2, \dots, k$ ist $\pi_i \circ \pi_1 = \bar{\pi}_i \circ \tilde{\pi}_2 \circ \pi_1 = \bar{\pi}_i \circ 0 = 0$ und $\pi_1 \circ \pi_i = 0$, weil π_i nach Konstruktion Werte in $\tilde{W}_2$ hat. Für $i, j \geq 2$, $i \neq j$ ist $\pi_i \circ \pi_j = \bar{\pi}_i \circ \tilde{\pi}_2 \circ \bar{\pi}_j \circ \tilde{\pi}_2$, und das verschwindet, weil $\bar{\pi}_j$ Werte in $\tilde{W}_2$ hat und daher $\tilde{\pi}_2 \circ \bar{\pi}_j = \bar{\pi}_j$ und nach Induktion $\bar{\pi}_i \circ \bar{\pi}_j = 0$ gilt. Aus Lemma 10.9 folgt $\pi_1 \circ f = f \circ \pi_1$. Für $i \geq 2$ ist

$$\pi_i \circ f = \bar{\pi}_i \circ \tilde{\pi}_2 \circ f = \bar{\pi}_i \circ \tilde{f} \circ \tilde{\pi}_2 = \tilde{f} \circ \pi_i = f \circ \pi_i.$$

Damit ist $V = W_1 \oplus \cdots \oplus W_k$ eine Zerlegung in eine direkte Summe f–invarianter Teilräume. Nach Konstruktion ist $W_1 = \operatorname{Ker}(p_1(f)^{n_1})$ und $W_i = \tilde{W}_2 \cap \operatorname{Ker}(p_i(f)^{n_i})$ für $i \geq 2$. Da aber nach Konstruktion $\tilde{W}_2 = \operatorname{Ker}(p_2(f)^{n_2} \circ \ldots \circ p_k(f)^{n_k})$ gilt und alle diese Abbildungen kommutieren ist klarerweise $\operatorname{Ker}(p_i(f)^{n_i}) \subset \tilde{W}_2$ und somit $W_i = \operatorname{Ker}(p_i(f)^{n_i})$. □

Im Moment sieht es so aus, als könnte die Primärzerlegung stark von dem verwendeten Polynom p abhängen. Das ist aber nicht wirklich der Fall. Die natürlichste Wahl für p ist das Minimalpolynom m_f von f. Sei $m_f = p_1^{m_1} \cdots p_k^{m_k}$ die Primfaktorzerlegung und sei $V = W_1 \oplus \cdots \oplus W_k$ die entsprechende Primärzerlegung. Dann ist $W_i = \operatorname{Ker}(p_i(f)^{m_i})$ für alle $i = 1, \dots, k$. Nun ist jedes Polynom $p \in \mathbb{K}[x]$, das $p(f) = 0$ erfüllt, von der Form $m_f q$ für ein Polynom $q \in \mathbb{K}[x]$. In Termen der Primfaktorzerlegung heißt das aber einfach, dass entweder die Potenzen von in m_f auftretenden Primfaktoren vergrößert werden, oder zusätzliche Primfaktoren hinzukommen.

Wir wollen zeigen, dass beides die Primärzerlegung nicht wirklich verändert. Ist $m \geq m_i$, dann behaupten wir, dass $\operatorname{Ker}(p_i(f)^m) = W_i$ gilt. Setzen wir $\tilde{W}_i := \operatorname{Ker}(p_i(f)^m)$, dann gilt natürlich $W_i \subset \tilde{W}_i$. Bilden wir andererseits die Primärzerlegung von V bezüglich des Polynoms $p = p_i^{m-n_i} m_f$, dann sehen wir, dass auch

$$V = W_1 \oplus \cdots \oplus W_{i-1} \oplus \tilde{W}_i \oplus W_{i+1} \oplus \cdots \oplus W_k$$

gilt. Damit ist aber $\dim(\tilde{W}_i) = \dim(V) - \sum_{j \neq i} \dim(W_j) = \dim(W_i)$, also $\tilde{W}_i = W_i$.

Sei andererseits q ein irreduzibles Polynom, das verschieden von allen p_i ist. Dann können wir die Primärzerlegung von V bezüglich $m_f q^\ell$ für beliebiges $\ell > 0$ bilden. Diese hat die Form $V = W_1 \oplus \cdots \oplus W_k \oplus W$, wobei $W = \mathrm{Ker}(q(f)^\ell)$ ist. Aus $V = W_1 \oplus \cdots \oplus W_k$ folgt dann aber sofort $\dim(W) = 0$, also $W = \{0\}$, also erhalten wir wieder die gleiche Primärzerlegung.

Jordan Zerlegung und Jordan'sche Normalform

Wir können jetzt eine allgemeine Normalform für lineare Abbildungen finden, deren charakteristisches Polynom in ein Produkt von Polynomen ersten Grades zerfällt. Insbesondere ist diese Voraussetzung über algebraisch abgeschlossenen Körpern und damit über $\mathbb{C}$ immer erfüllt.

10.10. Verallgemeinerte Eigenräume und Jordan–Zerlegung. Wir analysieren zunächst, wie die Summanden in der Primärzerlegung aussehen, die zu Primfaktoren ersten Grades gehören. Nehmen wir an, dass in der Primfaktorzerlegung von m_f ein Faktor $p_1^{m_1}$ mit $p_1 = x - \lambda$ und $m_1 \in \mathbb{N} \setminus \{0\}$ vorkommt. Dann ist $x - \lambda$ ein Teiler von p_f, also ist λ ein Eigenwert von f. In der Primärzerlegung erhalten wir einen entsprechenden Summenden $W_1 = \mathrm{Ker}((f - \lambda\,\mathrm{id})^{m_1}) \subset V$. Ist $m_1 = 1$, dann ist W_1 einfach der Eigenraum V_λ^f von f zum Eigenwert λ. Aus 10.9 wissen wir auch schon, dass $W_1 = \mathrm{Ker}((f - \lambda\,\mathrm{id})^n)$ für alle $n \geq m_1$ gilt. Damit können wir W_1 auch als $\{v \in V : \exists n \in \mathbb{N} : (f - \lambda\,\mathrm{id})^n(v) = 0\}$ beschreiben.

Definition 10.10. Sei V ein $\mathbb{K}$–Vektorraum.

(1) Für eine lineare Abbildung $f : V \to V$ und einen Eigenwert $\lambda \in \mathbb{K}$ von f definieren wir den *verallgemeinerten Eigenraum* $V_{(\lambda)}^f$ von f zum Eigenwert λ als

$$\{v \in V : \exists n \in \mathbb{N} : (f - \lambda\,\mathrm{id})^n(v) = 0\}.$$

(2) Eine lineare Abbildung $\varphi : V \to V$ heißt *nilpotent*, wenn es einen Index $N \in \mathbb{N}$ gibt, sodass $\varphi^N = 0$ gilt. Die kleinste Zahl N, für die das erfüllt ist, heißt der *Nilpotenzindex* von φ.

Man kann den verallgemeinerten Eigenraum $V_{(\lambda)}^f$ von f zum Eigenwert λ also als den größten Teilraum von V definieren, auf dem $(f - \lambda\,\mathrm{id})$ nilpotent ist. (Analog ist der Eigenraum V_λ^f der größte Teilraum von V auf dem $f - \lambda\,\mathrm{id} = 0$ gilt.)

Zunächst beweisen wir ein einfaches Resultat über nilpotente lineare Abbildungen, das wir in Kürze noch stark verfeinern werden.

Lemma 10.10. *Sei V ein n–dimensionaler Vektorraum über einem Körper $\mathbb{K}$ und sei $\varphi : V \to V$ eine nilpotente lineare Abbildung mit Nilpotenzindex k. Dann kann φ durch eine obere Dreiecksmatrix mit Nullen auf der Hauptdiagonale dargestellt werden. Daher ist $p_\varphi = (-1)^n x^n$ und $m_\varphi = x^k$, also insbesondere $k \leq n$.*

Beweis. Für $i = 1, \dots, k$ setzte $V_i := \mathrm{Ker}(\varphi^i) \subset V$. Dann gilt natürlich einerseits $V_1 \subset V_2 \subset \cdots \subset V_{k-1} \subset V_k = V$ und andererseits $\varphi(V_i) \subset V_{i-1}$. Ausserdem muss V_{i-1} jeweils ein echter Teilraum von V_i sein, sonst wäre $\mathrm{Ker}(\varphi^i) = \mathrm{Ker}(\varphi^{i-1})$ und damit $\mathrm{Ker}(\varphi^{i+1}) = \mathrm{Ker}(\varphi^i) = \mathrm{Ker}(\varphi^{i-1})$ und so weiter, was nur für $i - 1 \geq k$ möglich ist.

Wähle nun eine Basis von V_1 erweitere sie zu einer Basis von V_2, diese zu einer Basis von V_3 und so weiter, bis man eine Basis $\mathcal{B} = \{w_1, \dots, w_n\}$ für $V_k = V$ erhält. Für ein Basiselement $w_j \in \mathcal{B}$ gibt es einen kleinsten Index i, sodass $w_j \in V_i$ gilt. Damit ist aber $\varphi(w_j) \in V_{i-1}$ und kann damit als Linearkombination von Basiselementen w_ℓ mit $\ell < j$ geschrieben werden, also hat $[\varphi]_{\mathcal{B}}$ die verlangte Form.

Daraus folgt aber offensichtlich die Aussage über das charakteristische Polynom p_φ. Nach dem Satz von Cayley–Hamilton folgt $\varphi^n = 0$, also $k \leq n$ und damit folgt die Aussage über m_φ aus der Definition des Nilpotenzindex k. □

Nehmen wir nun an, dass $f : V \to V$ eine lineare Abbildung auf einem endlichdimensionalen $\mathbb{K}$–Vektorraum V ist, deren charakteristisches Polynom p_f in ein Produkt von Polynomen ersten Grades zerfällt. Ist $n = \dim(V)$, dann ist die Primfaktorzerlegung von p_f gerade $p_f = (-1)^n (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}$, wobei $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von f sind und n_i die algebraische Vielfachheit von λ_i bezeichnet.

Satz 10.10. *Sei $f : V \to V$ eine lineare Abbildung auf einem n–dimensionalen $\mathbb{K}$–Vektorraum V mit charakteristischem Polynom $p_f = (-1)^n (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}$. Dann gilt:*

(1) Für die verallgemeinerten Eigenräume $V^f_{(\lambda_i)}$ gilt $\dim(V^f_{(\lambda_i)}) = n_i$ und V ist die direkte Summe $V = V^f_{(\lambda_1)} \oplus \cdots \oplus V^f_{(\lambda_k)}$.

(2) Es gibt natürliche Zahlen $m_1, \dots, m_k$, sodass $1 \leq m_i \leq n_i$ für alle i gilt, sodass $m_f = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$ gilt. Die Abbildung f ist genau dann diagonalisierbar wenn $m_i = 1$ für alle i gilt.

(3) ("Jordan Zerlegung") Es gibt eindeutig bestimmte lineare Abbildungen $f_D, f_N : V \to V$, sodass f_D diagonalisierbar und f_N nilpotent ist, f_D und f_N miteinander und mit f kommutieren und $f = f_D + f_N$ gilt.

Beweis. (1) Die Zerlegung $V = V^f_{(\lambda_1)} \oplus \cdots \oplus V^f_{(\lambda_k)}$ folgt unmittelbar aus der Primärzerlegung aus Satz 10.9 und der Beschreibung der entsprechenden Summanden von oben. Nach Definition ist $(f - \lambda_i \operatorname{id})$ nilpotent auf $V^f_{(\lambda_i)}$. Wählen wir eine Basis für diesen Raum wie in Lemma 10.10 dann erhalten wir eine Matrixdarstellung für die Einschränkung von f auf diesen Raum als obere Dreiecksmatrix mit Hauptdiagonaleinträgen gleich λ_i. Daher ist das charakteristische Polynom dieser Einschränkung $(\lambda_i - x)^{r_i}$ wobei $r_i = \dim(V^f_{(\lambda_i)})$. Das Produkt aller dieser Polynome muss natürlich p_f ergeben, also folgt $r_i = n_i$.

(2) Nach dem Satz von Cayley–Hamilton (Satz 10.7) ist m_f ein Teiler von p_f. Daher muss es eine Darstellung in der angegebenen Form mit $0 \leq m_i \leq n_i$ für alle i geben und es bleibt nur noch $m_i > 0$ für alle i zu zeigen. Das folgt aber leicht, weil es zum Eigenwert λ_i mindestens einen Eigenvektor v_i gibt. Für diesen und beliebige m_j gilt aber

$$(f-\lambda_1)^{m_1} \circ \dots \circ (f-\lambda_{i-1})^{m_{i-1}} \circ (f-\lambda_{i+1})^{m_{i+1}} \circ \dots \circ (f-\lambda_k)^{m_k}(v_i) = \left(\prod_{j \neq i} (\lambda_i - \lambda_j)^{m_j}\right) v_i$$

und das ist $\neq 0$.

Für jedes i ist der Eigenraum $V^f_{\lambda_i}$ ein Teilraum von $V^f_{(\lambda_i)} = \operatorname{Ker}((f - \lambda_i)^{m_i})$. Offensichtlich sind die beiden Räume genau dann gleich, wenn $m_i = 1$ gilt. Ist das für alle i erfüllt, dann ist V die direkte Summe der Eigenräume von f, also f diagonalisierbar. Ist umgekehrt $m_i > 1$ für ein i, dann ist die Summe aller Eigenräume ein echter Teilraum von V, also kann f nicht diagonalisierbar sein.

(3) Wir beweisen erst die Existenz der Zerlegung. Nach Satz 10.9 finden wir für jedes i die Projektion π_i auf $V^f_{(\lambda_i)}$ und diese liegt in $\mathbb{K}[f]$. Wir setzen nun $f_D := \sum_i \lambda_i \pi_i \in \mathbb{K}[f]$

und $f_N := f - f_D \in \mathbb{K}[f]$. Dann kommutieren nach Konstruktion alle drei Abbildungen und $f = f_D + f_N$. Für $v \in V^f_{(\lambda_j)}$ ist $f_D(v) = \lambda_j \pi_j(v) = \lambda_j v$. Wählen wir Basen für die $V^f_{(\lambda_j)}$, dann ist die Vereinigung eine Basis für V, die aus Eigenvektoren für f_D besteht, also ist f_D diagonalisierbar. Andererseits stimmt auf $V^f_{(\lambda_j)}$ die Abbildung $f_N = f - f_D$ mit $f - \lambda_j\,\mathrm{id}$ überein, also verschwindet $(f_N)^{m_j}$ auf diesem Teilraum. Wählt man K als das Maximum der m_j, dann verschwindet $(f_N)^K$ auf jedem $V^f_{(\lambda_j)}$, also ist $(f_N)^K = 0$, also f_N nilpotent.

Zur Eindeutigkeit nehmen wir an, dass $f = f_1 + f_2$ noch eine Zerlegung in einen diagonalisierbaren Teil f_1 und einen nilpotenten Teil f_2 ist, sodass $f_1 \circ f_2 = f_2 \circ f_1$ gilt. Sei μ_i ein Eigenwert von f_1 und $v \in V^{f_1}_{\mu_i}$. Dann ist $f_1(f_2(v)) = f_2(f_1(v)) = \mu_i f_2(v)$, also gilt $f_2(V^{f_1}_{\mu_i}) \subset V^{f_1}_{\mu_i}$ und damit ist $V^{f_1}_{\mu_i}$ auch f–invariant. Da f_2 nilpotent ist, können wir wieder Lemma 10.10 verwenden, um eine Basis für $V^{f_1}_{\mu_i}$ zu finden, bezüglich derer die Einschränkung von f durch eine obere Dreiecksmatrix mit Hauptdiagonaleintragungen μ_i dargestellt wird. Damit ist aber μ_i ein Eigenwert von f und $V^{f_1}_{\mu_i}$ ist im zugehörigen verallgemeinerten Eigenraum enthalten. Daraus folgt leicht, dass $f_1 = f_D$ gilt und damit ist auch $f_2 = f - f_1 = f - f_D = f_N$. □

10.11. Die Normalform für nilpotente lineare Abbildungen. Zunächst benötigen wir noch eine Definition:

Definition 10.11. Für $\lambda \in \mathbb{K}$ und $m \geq 1$ definiert man den *Jordan–Block* $J_m(\lambda)$ der Größe m mit Eigenwert λ als die $m \times m$–Matrix (a_{ij}) mit $a_{i,i} = \lambda$ für alle i und $a_{i,i+1} = 1$ für $1 \leq i < m$ und alle anderen Eintragungen gleich Null.

Nach Definition sind also alle Eintragungen auf der Hauptdiagonale von $J_m(\lambda)$ gleich λ, die Eintragungen unmittelbar über der Hauptdiagonale gleich 1 und alle anderen Eintragungen gleich 0. Somit hat $J_m(\lambda)$ offensichtlich charakteristisches Polynom $(\lambda - x)^m$. Andererseits erfüllt $J_m(0) = J_m(\lambda) - \lambda\mathbb{I}_m$ gerade $J_m(e_1) = 0$ und $J_m(0)e_i = e_{i-1}$ für alle $i = 2, \dots, m$, ist also nilpotent mit Nilpotenzindex m. Daher stimmt das Minimalpolynom von $J_m(\lambda)$ mit $(x - \lambda)^m$ überein, und $J_m(\lambda)$ ist die einfachste Matrix mit diesem Minimalpolynom.

Wir kommen nun zur angekündigten Verfeinerung von Lemma 10.10. Dort haben wir für eine nilpotente lineare Abbildung φ schon die Kerne der Potenzen φ^i betrachtet. Andererseits können wir natürlich auch die Bilder $\mathrm{Im}(\varphi^i)$ für alle Potenzen von φ betrachten. Ist φ nilpotent mit Nilpotenzindex k, dann gilt natürlich $\mathrm{Im}(\varphi^i) \subset \mathrm{Ker}(\varphi^{k-i})$. Der wesentliche Schritt zum Verständnis nilpotenter linearer Abbildungen ist folgendes Lemma.

Lemma 10.11. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, $\varphi : V \to V$ eine lineare Abbildung und $k \in \mathbb{N}$. Angenommen, $\tilde{v}_1, \dots, \tilde{v}_m$ sind linear unabhängige Vektoren in $\mathrm{Im}(\varphi^{k-1}) \cap \mathrm{Ker}(\varphi)$, dann gilt:*

Sind $v_1, \dots, v_m \in V$ so, dass $\varphi^{k-1}(v_i) = \tilde{v}_i$ für $i = 1, \dots, m$ gilt, dann sind die Vektoren $\varphi^i(v_j)$ für $0 \leq i \leq k-1$ und $1 \leq j \leq m$ linear unabhängig in V und spannen einen φ–invarianten Teilraum W von V auf. Somit bilden diese Vektoren eine Basis $\mathcal{B}$ für W.

Bezüglich einer geeigneten Anordnung der Basis $\mathcal{B}$ ist die Matrixdarstellung der Einschränkung von φ auf W durch eine Blockdiagonalmatrix mit m Jordan–Blöcken $J_k(0)$ entlang der Hauptdiagonale und ansonsten lauter Nullen gegeben.

Beweis. Seien $a^i_j \in \mathbb{K}$ so, dass $\sum_{i,j} a^i_j \varphi^i(v_j) = 0$ gilt. Wenden wir auf diese Gleichung φ^{k-1} an, dann gilt $\varphi^{k-1}(\varphi^i(v_j)) = \varphi^{k+i-1}(v_j) = \varphi^i(\varphi^{k-1}(v_j)) = \varphi^i(\tilde{v}_j)$. Für $i = 0$

ist das natürlich $\tilde{v}_j$, für $i > 0$ erhält man 0, weil $\tilde{v}_j \in \mathrm{Ker}(\varphi)$ nach Voraussetzung gilt. Damit erhalten wir aber $0 = \sum_{j=1}^m a_j^0 \tilde{v}_j$, was wegen der linearen Unabhängigkeit der $\tilde{v}_j$ schon $a_j^0 = 0$ für alle $j = 1, \dots, m$ impliziert.

Nehmen wir nun induktiv an, dass wir $a_j^i = 0$ für alle $i < \ell$ und alle $j = 1, \dots, m$ bereits gezeigt haben. Dann hat unsere Gleichung die Form $\sum_{i=\ell}^{k-1} \sum_{j=1}^m a_j^i \varphi^i(v_j) = 0$. Wenden wir auf diese Gleichung $\varphi^{k-\ell-1}$ an, dann verschwinden wieder alle Terme mit $i > \ell$, und wir erhalten $\sum_{j=1}^n a_j^\ell \tilde{v}_j = 0$. Wegen der linearen Unabhängigkeit der $\tilde{v}_j$ impliziert das wieder $a_j^\ell = 0$ für alle $j = 1, \dots, m$. Nach Induktion folgt nun $a_j^i = 0$ für alle i und j, also ist die Menge $\mathcal{B}$ linear unabhängig.

Ist W der von dieser Menge erzeugte Teilraum, dann ist $\varphi(\varphi^i(v_j)) = \varphi^{i+1}(v_j) \in W$ für $i < k-1$ und $\varphi(\varphi^{k-1}(v_j)) = \varphi(\tilde{v}_j) = 0 \in W$. Da die Elemente $\varphi^i(v_j)$ eine Basis für W bilden, folgt damit $\varphi(W) \subset W$, also ist W ein φ–invarianter Teilraum.

Natürlich können wir W aber noch feiner zerlegen. Für $i = 1, \dots, m$ setze $\mathcal{B}_i = \{\varphi^{k-1}(v_i), \varphi^{k-2}(v_i), \dots, \varphi(v_i), v_i\}$. Als Teilmenge der linear unabhängigen Menge $\mathcal{B}$ ist das linear unabhängig, und φ bildet das erste Element von $\mathcal{B}_i$ auf 0 und jedes weitere Element auf das vorhergehende Element von $\mathcal{B}_i$ ab. Damit erzeugt jedes $\mathcal{B}_i$ einen φ–invarianten Teilraum $W_i \subset W$ und die Matrixdarstellung der Einschränkung von φ bezüglich $\mathcal{B}_i$ ist $J_k(0)$. Da die Vereinigung der $\mathcal{B}_i$ die Basis $\mathcal{B}$ von W ist, folgt die letzte Behauptung des Satzes. □

Damit kann man nun eine Normalform für nilpotente lineare Abbildungen durch einen elementaren aber mühsamen Induktionsbeweis herleiten. Wir führen den Beweis nicht ganz genau aus, weil die verkürzte Version viel leichter verständlich ist.

Satz 10.11. *Sei V ein endlichdimensionaler Vektorraum über einem Körper $\mathbb{K}$ und $\varphi : V \to V$ eine nilpotente lineare Abbildung mit Nilpotenzindex N. Dann gibt es eine Basis $\mathcal{B}$ von V, bezüglich derer die Matrixdarstellung von φ Blockdiagonalform hat, wobei entlang der Hauptdiagonale nur Jordan–Blöcke der Form $J_r(0)$ mit $r \leq N$ auftreten, wobei mindestens ein Block Größe N hat.*

Beweisskizze. Da φ Nilpotenzindex N hat, ist $\mathrm{Im}(\varphi^{N-1}) \subset \mathrm{Ker}(\varphi)$. Damit können wir Lemma 10.11 auf eine Basis für $\mathrm{Im}(\varphi^{N-1})$ anwenden und erhalten einen φ–invarianten Teilraum $W \subset V$ mit einer Basis $\mathcal{B}_1$ die von Vektoren der Form $\varphi^i(v_j)$ mit $0 \leq i \leq N-1$ und $1 \leq j \leq m_1\}$ gebildet wird, wobei $m_1 = \dim(\mathrm{Im}(\varphi^{N-1}))$ gilt. Das liefert m_1 Jordan–Blöcke $J_N(0)$ für die Matrixdarstellung.

Ist $W = V$ dann sind wir fertig. Andernfalls betrachten wir die Teilräume $\varphi^k(W) \subset \varphi^k(V)$. Für $k = N-1$ gilt nach Konstruktion $\varphi^{N-1}(W) = \mathrm{Im}(\varphi^{N-1}) = \varphi^{N-1}(V)$. Andererseits ist $\varphi^0(W) = W$ eine echte Teilmenge von $V = \varphi^0(V)$. Damit finden wir einen eindeutigen Index $r < N$, sodass $\varphi^k(W) = \varphi^k(V)$ für alle $k \geq r$ gilt, aber $\varphi^{r-1}(W)$ ein echter Teilraum von $\varphi^{r-1}(V)$ ist. Dann gilt natürlich auch

$$U' := \mathrm{Ker}(\varphi) \cap \varphi^{r-1}(W) \subset \mathrm{Ker}(\varphi) \cap \varphi^{r-1}(V) =: U.$$

Schränkt man φ auf $\varphi^{r-1}(W)$ ein, dann erhält man eine surjektive lineare Abbildung $\varphi^{r-1}(W) \to \varphi^r(W)$ mit Kern U'. Damit ist $\dim(U') = \dim(\varphi^{r-1}(W)) - \dim(\varphi^r(W))$ und analog ist $\dim(U) = \dim(\varphi^{r-1}(V)) - \dim(\varphi^r(V))$. Nach Voraussetzung ist aber $\varphi^r(W) = \varphi^r(V)$, also folgt

$$\dim(U) - \dim(U') = \dim(\varphi^{r-1}(V)) - \dim(\varphi^{r-1}(W)).$$

Wählt man einen Teilraum $U'' \subset U$, sodass $U = U' \oplus U''$ gilt, dann impliziert das sofort, dass $\varphi^{r-1}(V) = \varphi^{r-1}(W) \oplus U''$ gilt.

Nach Konstruktion ist $U'' \subset \mathrm{Ker}(\varphi)$, also kann man Lemma 10.10 auf die Elemente $\tilde{u}_1, \dots, \tilde{u}_{m_2}$ einer Basis für U'' anwenden. Das liefert linear unabhängige Vektoren $\varphi^i(u_j)$ mit $0 \leq i \leq r-1$, $1 \leq j \leq$. Nun sind aber nach Konstruktion die Vektoren $\tilde{u}_1, \dots, \tilde{u}_{m_2}$ und $\varphi^i(v_j)$ für $r \leq i \leq N-1$ (die eine Basis für $\varphi^r(W)$ bilden) linear unabhängig. Analog zum Beweis von Lemma 10.10 schließt man daraus leicht, dass die Vektoren $\varphi^i(u_j)$ und $\varphi^\ell(v_k)$ gemeinsam linear unabhängig sind. Damit spannen sie einen φ–invarianten Teilraum $W_1 \subset V$ auf und für eine passende Anordnung erhält man eine Matrixdarstellung mit m_1 Jordan–Blöcken $J_N(0)$ und m_2 Jordan–Blöcken $J_r(0)$.

Ist $W_1 = V$, dann sind wir fertig. Sonst ersetzt man W durch W_1, betrachtet wieder die Bilder der Potenzen von φ. Das liefert eine weitere Vergrößerung der Basis und kleinere Jordan–Blöcke mit Eigenwert 0. Offensichtlich erreicht man in endlich vielen Schritten eine Basis von V. □

10.12. Die Jordan'sche Normalform. Damit haben wir alle Zutaten beisammen um eine Normalform für lineare Abbildungen anzugeben, deren charakteristisches Polynom in ein Produkt von Polynomen ersten Grades zerfällt. Über einem algebraisch abgeschlossenen Körper ist diese Bedingung natürlich immer erfüllt.

Satz 10.12. *Sei V ein n–dimensionaler $\mathbb{K}$–Vektorraum und sei $f : V \to V$ eine lineare Abbildung, deren charakteristisches Polynom in ein Produkt von Polynomen ersten Grades zerfällt. Seien $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von f und für jedes $i = 1, \dots, k$ sei n_i die algebraische Vielfachheit von λ_i, r_i die geometrische Vielfachheit von λ_i und m_i die Potenz, mit der $(x - \lambda_i)$ in der Primfaktorzerlegung des Minimalpolynoms m_f auftritt.*

Dann gibt es eine Basis $\mathcal{B}$ für V, sodass die Matrixdarstellung $A = [f]_{\mathcal{B}}$ von f folgende Form hat: A ist eine Blockdiagonalmatrix wie in Satz 10.2. Jeder der Blöcke auf der Hauptdiagonale ist ein Jordan Block der Form $J_{m_{ij}}(\lambda_i)$ für einen der Eigenwerte λ_i und für den Eigenwert λ_i treten genau r_i solche Blöcke auf. Die Größen dieser Blöcke erfüllen $m_{ij} \leq m_i$, für mindestens ein j gilt $m_{ij} = m_i$ und $\sum_j m_{ij} = n_i$. Die Matrixdarstellung dieser Form ist bis auf die Reihenfolge der Jordan Blöcke eindeutig bestimmt.

Analog ist jede quadratische Matrix B über $\mathbb{K}$, deren charakteristisches Polynom in ein Produkt von Polynomen ersten Grades zerfällt, ähnlich zu einer Matrix der obigen Form. Diese Matrix heißt die Jordan'sche Normalform von B und ist bis auf die Reihenfolge der Blöcke eindeutig bestimmt. Ähnlichkeit solcher Matrizen ist äquivalent dazu, dass ihre Jordan'schen Normalformen sich nur durch die Reihenfolge der Blöcke unterscheiden.

Beweis. Wir müssen eigentlich nur noch die Resultate zusammenfassen. Sei $V = V^f_{(\lambda_1)} \oplus \dots \oplus V^f_{(\lambda_k)}$ die Zerlegung von V in die verallgemeinerten Eigenräume von f und $f = f_D + f_N$ die Jordan–Zerlegung von f. Wie im Beweis von Satz 10.10 sehen wir, dass jedes $V^f_{(\lambda_i)}$ ein f_N–invarianter Teilraum ist. Auf diesem Teilraum stimmt f_N mit $f - \lambda_i \mathrm{id}$ überein, ist also nilpotent mit Nilpotenzindex m_i. Damit liefert aber Satz 10.11 eine Basis für $V^f_{(\lambda_i)}$ für die die Matrixdarstellung der Einschränkung von f aus Jordan–Blöcken mit Eigenwert λ_i besteht, wobei die angegebenen Eigenschaften der Blockgrößen erfüllt sind. Nun ist für einen Jordan–Block $J_{m_{ij}}(\lambda_i)$ der Rang von $J_{m_{ij}}(\lambda_i) - \lambda_i \mathbb{I} = J_{m_{ij}}(0)$ offensichtlich gleich $m_{ij} - 1$. Daraus folgt aber sofort, dass der Rang der Abbildung $f - \lambda_i \mathrm{id}$ auf $V^f_{(\lambda_i)}$ gerade n_i minus der Anzahl der Jordan–Blöcke zum Eigenwert λ_i ist. Damit folgt aber, dass diese Anzahl genau die geometrische Vielfachheit r_i von λ_i

ist. Die Aussage über die Existenz der Jordan'schen Normalform einer Matrix B folgt daraus natürlich sofort.

Es bleiben also nur noch die Aussagen über Eindeutigkeit der Matrixform bzw. Ähnlichkeit von Matrizen zu beweisen. Zunächst kann man eine Permutation der Blöcke in einer Blockmatrixdarstellung natürlich durch eine Permutation der Basiselemente (in Gruppen) erreichen, die wiederum durch einen linearen Isomorphismus realisiert wird. Andererseits folgt aus dem Beweis von Satz 10.11, dass man die Größen der auftretenden Jordan–Blöcke sowie die Anzahl der Blöcke jeder auftretenden Größe aus den Dimensionen der Bilder von Potenzen der Einschränkung von $f - \lambda_i \,\mathrm{id}$ auf den verallgemeinerten Eigenraum $V^f_{(\lambda_i)}$ ablesen kann, die natürlich für ähnliche Matrizen gleich sind. □

Beispiel 10.12. (1) Für diagonalisierbare Matrizen A ist die Diagonalform die Jordan'sche Normalform und in diesem Fall haben alle Jordan Blöcke die Größe Eins.

(2) Eine 2×2–Matrix mit zwei verschiedenen Eigenwerten ist diagonalisierbar. Allgemein kann man sich für verschiedene Eigenwerte immer auf die verallgemeinerten Eigenräume einschränken und damit die Frage auf kleinere Matrizen reduzieren.

Für $A \in M_2(\mathbb{K})$ mit $p_A = (x - \lambda)^2$ gibt es zwei Möglichkeiten: Hat λ geometrische Vielfachheit 2, bzw. ist $m_A = (x - \lambda)$, dann ist $A = \lambda\,\mathrm{id}$ und ist damit schon in Jordan'scher Normalform mit zwei Blöcken $J_1(\lambda)$. Ist andererseits $m_A = (x - \lambda)^2$, dann muss in der Jordan'schen Normalform ein Block $J_2(\lambda)$ vorkommen, also muss die Jordan'sche Normalform $J_2(\lambda)$ sein.

(3) Für $A \in M_3(\mathbb{K})$ mit $p_A = -(x - \lambda)^3$ ist die Situation ähnlich einfach und die Jordan'sche Normalform ist durch m_A eindeutig bestimmt. Ist $m_A = (x - \lambda)$, dann ist $A = \lambda\,\mathrm{id}$, also diagonal. Für $m_A = (x - \lambda)^2$ müssen wir einen Block $J_2(\lambda)$ erhalten, dürfen aber keine größeren Blöcke haben. Somit kommt für die Jordan'sche Normalform nur ein Block $J_2(\lambda)$ und ein Block $J_1(\lambda)$ in Frage. Ist schließlich $m_A = (x - \lambda)^3$, dann muss ein Block $J_3(\lambda)$ auftauchen, also muss die Jordan'sche Normalform $J_3(\lambda)$ sein. Alternativ kann man die möglichen Normalformen hier auch durch die geometrische Vielfachheit von λ unterscheiden.

(4) Betrachten wir $A \in M_4(\mathbb{K})$ mit $p_A = (x - \lambda)^4$, dann erhalten wir den ersten Fall, in dem das Minimalpolynom alleine nicht mehr ausreicht, um die verschiedenen möglichen Jordan'schen Normalformen zu unterscheiden. Für $m_A = (x-\lambda)$ ist $A = \lambda\,\mathrm{id}$, für $m_A = (x-\lambda)^4$ ist die Jordan'sche Normalform $J_4(\lambda)$, und für $m_A = (x-\lambda)^3$ muss ein Block $J_3(\lambda)$ auftauchen, also bleibt nur noch ein Block $J_1(\lambda)$ übrig. Im Fall $m_A = (x-\lambda)^2$ gibt es aber zwei Möglichkeiten, nämlich entweder zwei Blöcke $J_2(\lambda)$, oder ein Block $J_2(\lambda)$ und zwei Blöcke $J_1(\lambda)$. Diese beiden Möglichkeiten werden durch die geometrische Vielfachheit unterschieden, die im ersten Fall 2 und im zweiten Fall 3 beträgt.

(5) Im Allgemeinen reichen Minimalpolynom und geometrische Vielfachheit nicht aus, um die Jordan'sche Normalform eindeutig festzulegen. Betrachten wir man etwa $A \in M_7(\mathbb{K})$ mit $p_A = -(x - \lambda)^7$ und $m_A = (x - \lambda)^3$, sodass λ geometrische Vielfachheit 3 hat. Dann muss es 3 Blöcke geben der Größe ≤ 3 geben, von denen mindestens einer die Größe 3 hat und deren Größen sich auf 7 summieren. Dafür gibt es die Möglichkeiten $7 = 3 + 3 + 1$ und $7 = 3 + 2 + 2$. Um zwischen diesen Möglichkeiten zu unterscheiden, bemerkt man, dass der Rang von $(A - \lambda\mathbb{I})^2$ im ersten Fall 2 und im zweiten Fall 1 ist.

10.13. Bemerkung: Normalform über $\mathbb{R}$. Betrachten wir eine lineare Abbildung $f : V \to V$ auf einem n–dimensionalen $\mathbb{R}$–Vektorraum V mit charakteristischem Polynom p_f. Aus Beispiel (2) von 10.9 wissen wir, dass irreduzible Polynome über $\mathbb{R}$ entweder Grad eins oder Grad zwei haben und im letzteren Fall gibt es keine reelle

Nullstelle. Man erhält also eine eindeutige Primfaktorzerlegung der Form

$$p_f = (-1)^n (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k} (q_1)^{r_1} \dots (q_\ell)^{r_\ell},$$

für paarweise verschiedene $\lambda_i \in \mathbb{R}$ und monische Polynome $q_j \in \mathbb{R}[x]$ vom Grad 2 ohne reelle Nullstellen. Dabei ist $n_1 + \cdots + n_k + 2r_1 + \cdots + 2r_\ell = n$. In der entsprechenden Primärzerlegung erhält man einerseits die verallgemeinerten Eigenräume $V^f_{(\lambda_i)}$ die jeweils Dimension n_i haben, und die Räume $\operatorname{Ker}((q_j)^{r_j})$.

Um eine Normalform über $\mathbb{R}$ zu finden, kann man im wesentlichen wieder den Trick benutzen, dass man eine reelle Matrix als komplexe Matrix betrachtet und die Resultate über die Jordan'sche Normalform aus 10.12 über $\mathbb{C}$ benutzt. Dabei kann man sich auf einen der Summanden $\operatorname{Ker}((q_j)^{r_j})$ einschränken, erhält eine Matrix in $M_{2n}(\mathbb{C})$ mit zwei konjugiert komplexe Eigenwerten. Nun bestimmt man für einen der beiden verallgemeinerten Eigenräume eine komplexe Basis, bezüglich derer man eine Matrixdarstellung durch Jordan–Blöcke erhält. Dann zeigt man, dass man durch komplexe Konjugation dieser Basisvektoren eine Basis für den anderen Eigenraum erhält, bezüglich derer man genau die komplex konjugierten Jordan–Blöcke als Matrixdarstellung bekommt. Dann kann man aus den Realteilen und den Imaginärteilen dieser Basisvektoren eine reelle Basis für den Teilraum $\mathbb{R}^{2n} \subset \mathbb{C}^{2n}$ bilden, bezüglich derer man eine schöne Matrixdarstellung der ursprünglichen linearen Abbildung erhält.

Die "Bausteine" für diese schönen Matrixdarstellungen sind Blöcke, die man mit $\tilde{J}_k(\alpha, \beta) \in M_{2k}(\mathbb{R})$ bezeichnet. Am einfachsten kann man $\tilde{J}_k(\alpha, \beta)$ als $k \times k$–Matrix beschreiben, deren Eintragungen 2×2–Matrizen sind. Entlang der Hauptdiagonale haben diese Matrizen Eintragungen der Form $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, direkt über der Hauptdiagonale erhält man Eintragungen $\mathbb{I}_2$, sonst lauter Nullen. Die Parameter $\alpha, \beta \in \mathbb{R}$ mit $\beta > 0$ sind dadurch bestimmt, dass die komplexen Nullstellen des entsprechenden Polynoms q_i gerade $\alpha \pm i\beta$ sind. Wie im komplexen Fall ist die gesamte Darstellung bis auf die Reihenfolge der Blöcke eindeutig bestimmt und zwei Matrizen sind genau dann ähnlich, wenn sie die gleiche Normalform in diesem Sinne haben.

KAPITEL 11

Multilineare Algebra

Wir haben im Verlauf der Vorlesung schon einige Beispiele von bilinearen und multilinearen Abbildungen kennen gelernt, insbesondere Determinantenfunktionen und innere Produkte. In diesem letzten Kapitel werden wir diese Abbildungen systematischer studieren und insbesondere Tensorprodukte von Vektorräumen und symmetrische und äußere Potenzen eines Vektorraumes studieren. Diese Konzepte werden sowohl in der Algebra als auch in der Analysis angewandt. Andererseits werden wir uns in diesem Kapitel systematischer mit Fragen der Natürlichkeit befassen, womit auch ein erster Einblick in die kategorielle Denkweise vermittelt werden soll.

Natürlichkeit

11.1. Natürliche Isomorphismen. Aus Korollar 4.10 wissen wir, dass zwei endlichdimensionale Vektorräume über einem Körper $\mathbb{K}$ genau dann isomorph sind, wenn sie die gleiche Dimension haben. Daher sind Aussagen über Isomorphie von endlichdimensionalen Räumen grundsätzlich eher uninteressant. Wir haben allerdings schon Fälle kennen gelernt, in denen man stärkere Aussagen beweisen kann. Ein typisches Beispiel bilden die Dual– und Bidualräume, die wir in Abschnitt 5.12 und 5.13 besprochen haben.

Für einen $\mathbb{K}$–Vektorraum V ist der *Dualraum* V^* der Raum $L(V,\mathbb{K})$ aller linearen Funktionale auf V. Natürlich kann man dann auch den Dualraum $V^{**} := (V^*)^*$ von V^* bilden, den man als *Bidualraum* von V bezeichnet. Ist V endlichdimensional, dann ist $\dim(V^*) = \dim(V) \cdot \dim(\mathbb{K}) = \dim(V)$, also haben V, V^* und V^{**} alle die gleiche Dimension und sind daher isomorph. Aus Satz 5.12 wissen wir auch, dass man zu einer Basis von V die duale Basis von V^* erhält, womit man explizit einen Isomorphismus $V \to V^*$ angeben kann (der aber von der Wahl der Basis abhängt).

Im Fall des Bidualraumes sieht die Situation anders aus. Hier kann man "ganz allgemein" und ohne jede Wahl eine lineare Abbildung $i_V : V \to V^{**}$ angeben, siehe Abschnitt 5.13. Man ordnet nämlich jedem $v \in V$ die Funktion $\mathrm{ev}_v : V^* \to \mathbb{K}$ zu, die durch $\mathrm{ev}_v(\varphi) = \varphi(v)$ für eine lineare Abbildung $\varphi : V \to \mathbb{K}$ definiert ist. Man zeigt dann leicht (siehe Satz 5.13), dass für jedes $v \in V$ die Abbildung ev_v linear ist, also ein Element von V^{**} definiert. Weiters ist auch $i_V : V \to V^{**}$ linear und injektiv und damit für endlichdimensionales V automatisch ein linearer Isomorphismus.

Ein weiteres Beispiel liefern die Quotientenräume aus Abschnitt 5.10. Ist V ein $\mathbb{K}$–Vektorraum und $W \subset V$ ein Teilraum, dann kann man für jeden Vektorraum Z und jede lineare Abbildung $f : V \to Z$ die Einschränkung $f|_W : W \to Z$ bilden. Das definiert eine lineare Abbildung $L(V,Z) \to L(W,Z)$ und wir können ihren Kern betrachten, den wir vorübergehend mit $\mathcal{A}_Z$ bezeichnen. Ist V endlichdimensional, dann sieht man leicht, dass diese Abbildung surjektiv ist: Man wählt einfach eine Basis $\{w_1,\dots,w_k\}$ für W und erweitert sie durch $v_{k+1}, \dots v_n$ zu einer Basis für V. Hat man $\varphi : W \to Z$ gegeben, dann betrachtet man die eindeutige lineare Abbildung $f : V \to Z$, die $f(w_i) = \varphi(w_i)$ für $i = 1,\dots,k$ und $f(v_j) = 0$ für $j = k+1,\dots,n$ erfüllt. Diese erfüllt natürlich $f|_V = \varphi$ (und man hätte die $f(v_j)$ auch beliebig anders wählen können). Ist auch Z

endlichdimensional, dann erhält man $\dim(\mathcal{A}_Z) = (n-k)\dim(Z)$. Wegen $\dim(V/W) = n-k$ folgt dann, dass die Vektorräume $\mathcal{A}_Z$ und $L(V/W, Z)$ die gleiche Dimension haben, also isomorph sind.

Unsere Überlegungen über Quotientenräume legen aber eine viel "bessere" (oder "genauere") Beschreibung nahe. Betrachten wir die kanonische Surjektion $\pi : V \to V/W$, dann können wir eine Funktion $T : L(V/W, Z) \to L(V, Z)$ definieren, indem wir jeder linearen Abbildung $\varphi : V/W \to Z$ die Abbildung $T(\varphi) := \varphi \circ \pi : V \to Z$ zuordnen, die als Komposition linearer Abbildungen linear ist. Man verifiziert sofort, dass die Abbildung T linear ist. Weiters ist $W = \operatorname{Ker}(\pi)$ und daraus folgt sofort, dass $W \subset \operatorname{Ker}(T(\varphi))$ für alle $\varphi \in L(V/W, Z)$ gilt, also ist $\operatorname{Im}(T) \subset \mathcal{A}_Z$. Betrachten wir nun die universelle Eigenschaft von π aus Proposition 5.11. Diese sagt, dass es für eine lineare Abbildung $f : V \to Z$ mit $W \subset \operatorname{Ker}(f)$ (was genau $f \in \mathcal{A}_Z$ bedeutet) eine eindeutige lineare Abbildung $\underline{f} : V/W \to Z$ gibt, die $\underline{f} \circ \pi = f$, also $T(\underline{f}) = f$ erfüllt. Das bedeutet aber genau, dass T bijektiv ist, also einen linearen Isomorphismus $L(V/W, Z) \to \mathcal{A}_Z$ definiert.

Um den Eindruck, dass $i_V : V \to V^{**}$ und $T : L(V/W, Z) \to \mathcal{A}_Z$ besonders "schöne" oder "natürliche" Isomorphismen sind, in eine mathematische Aussage zu verwandeln, muss man bemerken, dass in beiden Fällen Konstruktionen vorliegen, die nicht nur auf Vektorräume wirken, sondern auch auf lineare Abbildungen. Man spricht von *funktoriellen* Konstruktionen, der genaue mathematische Rahmen dafür ist die Kategorientheorie. Ist $f : U \to V$ eine lineare Abbildung zwischen zwei Vektorräumen, dann erhält man eine induzierte Abbildung $f^* : V^* \to U^*$ durch $f^*(\varphi) = \varphi \circ f$ und weiters $f^{**} : U^{**} \to V^{**}$. Es stellt sich heraus, dass die wichtigsten Eigenschaften die Gleichung $(f \circ g)^* = g^* \circ f^*$ aus Satz 5.12 und die offensichtliche Tatsache $(\mathrm{id}_V)^* = \mathrm{id}_{V^*}$ sind. Daraus ergibt sich sofort $(f \circ g)^{**} = f^{**} \circ g^{**}$ und $\mathrm{id}^{**} = \mathrm{id}$.

In Lemma 5.13 haben wir direkt verifiziert, dass die Familie i_V von Abbildungen in folgendem Sinn natürlich ist: Hat man $f : U \to V$ gegeben, dann gibt es zwei Möglichkeiten "von U nach V^{**} zu kommen". Diese beiden Möglichkeiten liefern das gleiche Ergebnis, denn es gilt $i_V \circ f = f^{**} \circ i_U : U \to V^{**}$.

Für Dualräume gibt es kein analoges Resultat. Hier würden wir verlangen, dass wir lineare Abbildungen $j_U : U \to U^*$ und $j_V : V \to V^*$ gegeben haben und die einzige passende Verträglichkeitsbedingung wäre, dass $j_U = f^* \circ j_V \circ f$ für alle linearen Abbildungen $f : U \to V$ gelten müsste. Setzt man für f die Nullabbildung ein, dann sieht man sofort, dass $j_U = 0$ gelten muss. Man zeigt leicht, dass eine derartige Verträglichkeitsbedingung nicht einmal dann erfüllt sein kann, wenn man für f nur (beliebige) lineare Isomorphismen zulässt.

Im Fall von $\mathcal{A}_Z$ sieht das ähnlich aus. Für eine lineare Abbildung $g : Z \to \tilde{Z}$ erhält man eine lineare Abbildung $S_g : L(V, Z) \to L(V, \tilde{Z})$ durch $S_g(f) = g \circ f$. (Auch hier gilt wieder $S_{g_1 \circ g_2} = S_{g_1} \circ S_{g_2}$ und $S_{\mathrm{id}} = \mathrm{id}$.) Aus dieser Definition folgt sofort, dass $S_g(\mathcal{A}_Z) \subset \mathcal{A}_{\tilde{Z}}$ gilt, also kann man S_g auch als Abbildung zwischen diesen Räumen betrachten. Analog erhält man auch Abbildungen $\underline{S}_g : L(V/W, Z) \to L(V/W, \tilde{Z})$, also gibt es auch hier wieder zwei Möglichkeiten, von $L(V/W, Z)$ nach $\mathcal{A}_{\tilde{Z}}$ zu kommen. Diese stimmen wieder überein, denn für $f \in \mathcal{A}_Z$ und die eindeutige Abbildung $\underline{f} : V/W \to Z$, sodass $f = \underline{f} \circ \pi$ ist, ist $g \circ \underline{f} \circ \pi = g \circ f = (\underline{g \circ f}) \circ \pi$ und damit $\underline{g \circ f} = g \circ \underline{f}$ nach der Eindeutigkeit in der universellen Eigenschaft.

11.2. Der freie Vektorraum. Als Vorbereitung für die Konstruktion des Tensorprodukts besprechen wir eine Konstruktion die sehr kategoriell angehaucht ist und sich damit auch gut zur weiteren Eingewöhnung in die Denkweise dieses Abschnitts eignet.

Wir suchen für eine Menge X einen $\mathbb{K}$–Vektorraum, der X als Basis hat. Denkt man daran, dass man eine lineare Abbildung auf den Elementen einer Basis beliebig vorgeben kann, dann bietet es sich an, das durch eine universelle Eigenschaft zu beschreiben, in der keine Basis vorkommt. In dieser Formulierung suchen wir zu gegebenem X einen $\mathbb{K}$–Vektorraum $\mathcal{F}(X) = \mathcal{F}_{\mathbb{K}}(X)$ zusammen mit einer Funktion $i : X \to \mathcal{F}(X)$, sodass es für einen beliebigen $\mathbb{K}$–Vektorraum Z und eine beliebige Funktion $f : X \to Z$ eine eindeutige lineare Abbildung $\tilde{f} : \mathcal{F}(X) \to Z$ gibt, die $f = \tilde{f} \circ i$ erfüllt.

Der Vorteil einer derartigen Beschreibung ist, dass sie, wie wir gleich beweisen werden, das Paar $(\mathcal{F}_{\mathbb{K}}(X), i)$ bis auf natürliche Isomorphie festlegt. Hat man also die Existenz eines Vektorraumes mit dieser Eigenschaft einmal bewiesen, dann spielt die konkrete Realisierung keine Rolle mehr, es genügt die universelle Eigenschaft zu kennen.

Der freie Vektorraum wir oft etwas schwammig als Menge von "formalen Linearkombinationen" von Elementen von X beschrieben. Es ist aber leicht eine saubere Beschreibung zu geben. Zu der Menge X betrachten wir die Menge $\mathbb{K}^X$ aller Funktionen $\varphi : X \to \mathbb{K}$. Aus Beispiel 2.2 wissen wir, dass $\mathbb{K}^X$ ein Vektorraum unter den punktweisen Operationen $(\varphi + \psi)(x) = \varphi(x) + \psi(x)$ und $(\lambda\varphi)(x) = \lambda(\varphi(x))$ ist. Nun definieren wir $\mathcal{F}_{\mathbb{K}}(X) \subset \mathbb{K}^X$ als die Teilmenge all jener Funktionen, die nur auf endlich vielen Punkten einen Wert $\neq 0$ annehmen. Diese Teilmenge ist offensichtlich abgeschlossen unter den punktweisen Operationen, also ein Teilraum von $\mathbb{K}^X$ und damit ein $\mathbb{K}$–Vektorraum. Für ein Element $x \in X$ definieren wir die *charakteristische Funktion* $\chi_x : X \to \mathbb{K}$, als jene Funktion, die x auf 1 und alle anderen Punkte von X auf 0 abbildet. Dann setzen wir $i(x) := \chi_x$.

Proposition 11.2. *(1) Das Paar $(\mathcal{F}_{\mathbb{K}}(X), i)$ hat die universelle Eigenschaft, dass es zu einer beliebigen Funktion $f : X \to Z$ in einen $\mathbb{K}$–Vektorraum Z eine eindeutige lineare Abbildung $\tilde{f} : \mathcal{F}_{\mathbb{K}}(X) \to Z$ gibt, die $f = \tilde{f} \circ i$ erfüllt.*

(2) Ist V ein $\mathbb{K}$–Vektorraum und $j : X \to V$ eine Funktion, sodass das Paar (V, j) ebenfalls die universelle Eigenschaft aus (1) besitzt, dann gibt es einen eindeutigen linearen Isomorphismus $F : \mathcal{F}_{\mathbb{K}}(X) \to V$, sodass $F \circ i = j$ gilt.

Beweis. (1) Ist $f : X \to Z$ gegeben, dann setzt man für $\varphi \in \mathcal{F}(X)$ einfach $\tilde{f}(\varphi) := \sum_{x:\varphi(x)\neq 0} \varphi(x) \cdot f(x)$. Dann gilt natürlich $\tilde{f}(i(x)) = \tilde{f}(\chi_x) = 1 \cdot f(x) = f(x)$, also $\tilde{f} \circ i = f$. Für $\varphi \in \mathcal{F}(X)$ gilt natürlich $\varphi = \sum_{x:\varphi(x)\neq 0} \varphi(x)\chi_x$ also ist $\tilde{f}$ durch die Werte auf den Punkten $\chi_x = i(x)$ eindeutig bestimmt. Man rechnet sofort nach, dass $\tilde{f}$ eine lineare Abbildung ist.

(2) Das ist ein ganz allgemeines Prinzip, dass wir noch öfter benutzen werden. Wenden wir die universelle Eigenschaft von $(\mathcal{F}(X), i)$ auf die Funktion $j : X \to V$ an, dann erhalten wir eine lineare Abbildung $F := \tilde{j} : \mathcal{F}(X) \to V$, die $F \circ i = j$ erfüllt. Umgekehrt kann man die universelle Eigenschaft von (V, j) auf $i : X \to \mathcal{F}(X)$ anwenden und erhält eine lineare Abbildung $G : V \to \mathcal{F}(X)$ sodass $G \circ j = i$ gilt. Betrachten wir nun $G \circ F : \mathcal{F}(X) \to \mathcal{F}(X)$, dann gilt $G \circ F \circ i = G \circ j = i = \mathrm{id} \circ i$, also muss nach der Eindeutigkeit in der universellen Eigenschaft $G \circ F = \mathrm{id}$ gelten. Analog erhält man $F \circ G = \mathrm{id}$, also ist F ein linearer Isomorphismus und $G = F^{-1}$. □

Tensorprodukte

11.3. Bilineare Abbildungen und Tensorprodukt. Der Begriff der Bilinearität, den wir bisher hauptsächlich für Bilinearformen mit Werten in $\mathbb{R}$ oder $\mathbb{C}$ kennen gelernt haben, macht natürlich für beliebige Vektorräume über allgemeinen Körpern

Sinn. Für $\mathbb{K}$–Vektorräume V, W und Z nennt man eine Funktion $\varphi : V \times W \to Z$ bilinear, wenn $\varphi(v_1+\lambda v_2, w) = \varphi(v_1)+\lambda\varphi(v_2, w)$ und $\varphi(v, w_1+\lambda w_2) = \varphi(v, w_1)+\lambda\varphi(v, w_2)$ für alle $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$ und $\lambda \in \mathbb{K}$ gilt. Das bedeutet gerade, dass für jedes fixe $w_0 \in W$, die Zuordnung $v \mapsto \varphi(v, w_0)$ eine lineare Abbildung $V \to Z$ definiert und analog für fixes $v_0 \in V$ und die Abbildung $W \to Z$, die durch $w \mapsto \varphi(v_0, w)$ definiert ist.

Proposition 11.3. *Die Menge $B(V, W; Z)$ aller bilinearen Abbildungen $V \times W \to Z$ ist ein Vektorraum unter den punktweisen Operationen, der natürlich isomorph zu $L(V, L(W, Z))$ und $L(W, L(V, Z))$ ist. Sind V, W und Z endlichdimensional, dann gilt insbesondere* $\dim(B(V, W; Z)) = \dim(V) \dim(W) \dim(Z)$.

Beweis. Man verifiziert sofort, dass für bilineare Abbildungen φ und ψ auch $\varphi+\psi$ und $\lambda\varphi$ für $\lambda \in \mathbb{K}$ bilinear sind. Damit ist $B(V, W; Z)$ ein Teilraum im Vektorraum aller Funktionen $V \times W \to Z$. Ist $\varphi : V \times W \to Z$ bilinear und $v \in V$, dann bezeichne mit $\varphi_v : W \to Z$ die Funktion $\varphi_v(w) := \varphi(v, w)$. Dann gilt natürlich $\varphi_v \in L(W, Z)$. Die Gleichung $\varphi(v_1 + \lambda v_2, w) = \varphi(v_1, w) + \lambda\varphi(v_2, w)$ gilt für alle $w \in W$ und sagt dann gerade, dass $\varphi_{v_1+\lambda v_2} = \varphi_{v_1} + \lambda\varphi_{v_2}$ in $L(W, Z)$ gilt. Damit ist aber $v \mapsto \varphi_v$ eine lineare Abbildung $V \to L(W, Z)$ und indem wir φ diese Funktion zuordnen, erhalten wir eine Funktion $B(V, W; Z) \to L(V, L(W, Z))$. Analog sagt $(\varphi+\lambda\psi)(v, w) = \varphi(v, w)+\lambda\psi(v, w)$ gerade, dass $(\varphi + \lambda\psi)_v = \varphi_v + \lambda\psi_v$ für alle $\varphi, \psi \in B(V, W; Z)$ und $\lambda \in \mathbb{K}$ gilt. Da dies für alle $v \in V$ gilt, ist die Abbildung $B(V, W; Z) \to L(V, L(W, Z))$, die wir definiert haben, linear.

Ist umgekehrt $F : V \to L(W, Z)$ eine lineare Abbildung, dann definieren wir $\varphi : V \times W \to Z$ durch $\varphi(v, w) := F(v)(w)$. Man rechnet sofort nach, dass φ bilinear ist und dass die dadurch definierte Abbildung $L(V, L(W, Z)) \to B(V, W; Z)$ invers zu der oben konstruierten ist. Damit ist die oben konstruierte lineare Abbildung bijektiv, also ein linearer Isomorphismus. Einen Isomorphismus $B(V, W; Z) \to L(W, L(V, Z))$ konstruiert man analog, indem man erst die W–Komponente "herauszieht". Die Aussage über die Dimensionen ist dann klar nach Satz 4.15. □

Ist $\varphi : V \times W \to Z$ eine bilineare Abbildung und $f : Z \to \tilde{Z}$ linear, dann ist natürlich auch $f \circ \varphi : V \times W \to \tilde{Z}$ eine bilineare Abbildung. (Und diese Tatsache benötigt man, um die Aussage über die Natürlichkeit des Isomorphismus in der obigen Proposition exakt zu machen.) Man kann sich daher fragen, ob es eine "universelle" bilineare Abbildung von $V \times W$ in einen passenden Vektorraum gibt, aus der man jede andere bilineare Abbildung durch Dahinterschalten einer linearen Abbildung erzeugen kann. Das motiviert die folgende Definition.

Definition 11.3. Seien V und W Vektorräume über $\mathbb{K}$. Ein *Tensorprodukt* von V und W ist ein $\mathbb{K}$–Vektorraum T zusammen mit einer bilinearen Abbildung $b : V \times W \to T$, die folgende universelle Eigenschaft hat: Ist Z ein $\mathbb{K}$–Vektorraum und $\varphi : V \times W \to Z$ eine bilineare Abbildung, dann gibt es eine eindeutige lineare Abbildung $\tilde{\varphi} : T \to Z$, sodass $\varphi = \tilde{\varphi} \circ b$ gilt.

Wie im Fall des freien Vektorraumes spielt es keine Rolle, wie ein Tensorprodukt konstruiert wird, es ist durch die universelle Eigenschaft vollständig bestimmt.

Lemma 11.3. *Sind T_1 und T_2 Tensorprodukte von V und W mit zugehörigen bilinearen Abbildungen $b_1 : V \times W \to T_1$ und $b_2 : V \times W \to T_2$, dann gibt es einen eindeutigen linearen Isomorphismus $f : T_1 \to T_2$, der $f \circ b_1 = b_2$ erfüllt.*

Beweis. Das ist völlig analog zu Teil (2) von Proposition 11.2. Die universellen Eigenschaften liefern lineare Abbildungen $f := \tilde{b}_2 : T_1 \to T_2$ mit $f \circ b_1 = b_2$ und $g : T_2 \to T_1$ mit $g \circ b_2 = b_1$. Damit ist $g \circ f \circ b_1 = g \circ b_2 = b_1 = \mathrm{id} \circ b_1$, also $g \circ f = \mathrm{id}$ und analog folgt $f \circ g = \mathrm{id}$. □

Wenn man gezeigt hat, dass ein Tensorprodukt existiert, dann kann man von *dem* Tensorprodukt von V und W sprechen, und muss sich nicht weiter darum kümmern, wie es konstruiert wurde. Man bezeichnet das Tensorprodukt von V und W dann mit $V \otimes_{\mathbb{K}} W$ oder, wenn der Grundkörper klar ist mit $V \otimes W$ und schreibt die zugehörige bilineare Abbildung als $(v, w) \mapsto v \otimes w$. Die "universelle" Konstruktion des Tensorprodukts ist etwas wild, kann aber dafür auch in anderen Teilen der Mathematik analog angewandt werden. Für endlichdimensionale Vektorräume werden wir bald eine viel einfachere Beschreibungen des Tensorprodukts erhalten.

Satz 11.3. *Für beliebige Vektorräume V und W über einem Körper $\mathbb{K}$ gibt es ein Tensorprodukt $V \otimes W$.*

Beweis. Betrachte die Menge $V \times W$, den freien Vektorraum $\mathcal{F}(V \times W)$ und die zugehörige Funktion $i : V \times W \to \mathcal{F}(V \times W)$. Sei $\mathcal{G} \subset \mathcal{F}(V \times W)$ der Teilraum der von allen Elementen der Form $i(v_1 + \lambda v_2, w) - i(v_1, w) - \lambda i(v_2, w)$ mit $v_1, v_2 \in V$, $w \in W$ und $\lambda \in \mathbb{K}$ sowie allen Elementen der Form $i(v, w_1 + \lambda w_2) - i(v, w_1) - \lambda i(v, w_2)$ mit $v \in V$, $w_1, w_2 \in W$ und $\lambda \in \mathbb{K}$ erzeugt wird. Dann betrachte den Quotientenraum $V \otimes W := \mathcal{F}(V \times W)/\mathcal{G}$ und die natürliche Surjektion $\pi : \mathcal{F}(V \times W) \to \mathcal{F}(V \times W)/\mathcal{G}$ und für $v \in V$ und $w \in W$ definiere $v \otimes w := \pi(i(v, w)) \in V \otimes W$.

Nach Konstruktion ist

$$0 = \pi(i(v_1 + \lambda v_2, w) - i(v_1, w) - \lambda i(v_2, w)) = (v_1 + \lambda v_2) \otimes w - v_1 \otimes w - \lambda(v_2 \otimes w)$$

und analog folgt $v \otimes (w_1 + \lambda w_2) = v \otimes w_1 + \lambda(v \otimes w_2)$. Damit ist die Funktion $(v, w) \mapsto v \otimes w$ bilinear und wir müssen nur noch zeigen, dass die universelle Eigenschaft erfüllt ist.

Sei also $\varphi : V \times W \to Z$ eine bilineare Abbildung in einen beliebigen $\mathbb{K}$–Vektorraum Z. Wenden wir die universelle Eigenschaft des freien Vektorraumes auf die Funktion φ an, dann erhalten wir eine lineare Abbildung $\hat{\varphi} : \mathcal{F}(V \times W) \to Z$, die $\hat{\varphi} \circ i = \varphi$ erfüllt. Nun rechnen wir einfach

$$\hat{\varphi}(i(v_1 + \lambda v_2, w) - i(v_1, w) - \lambda i(v_2, w)) = \varphi(v_1 + \lambda v_2, w) - \varphi(v_1, w) - \lambda\varphi(v_2, w) = 0,$$

weil φ bilinear ist. Analog liegt auch jedes Element der Form $i(v, w_1 + \lambda w_2) - i(v, w_1) - \lambda i(v, w_2)$ im Kern von $\hat{\varphi}$. Damit liegt aber der Teilraum $\mathcal{G}$, der von allen diesen Elementen erzeugt wird im Kern von $\hat{\varphi}$. Somit können wir die universelle Eigenschaft des Quotienten aus Proposition 5.11 anwenden, die eine lineare Abbildung

$$\tilde{\varphi} = \underline{\hat{\varphi}} : \mathcal{F}(V \times W)/\mathcal{G} = V \otimes W \to Z$$

liefert, die $\tilde{\varphi} \circ \pi = \hat{\varphi}$ erfüllt. Damit gilt aber $\tilde{\varphi}(v \otimes w) = \hat{\varphi}(i(v, w)) = \varphi(v, w)$.

Aus den universellen Eigenschaften wissen wir auch, dass eine lineare Abbildung $f : \mathcal{F}(V \times W)/\mathcal{G} \to Z$ eindeutig durch die lineare Abbildung $f \circ \pi : \mathcal{F}(V \times W) \to Z$ bestimmt ist, die wiederum eindeutig durch $f \circ \pi \circ i$ festgelegt wird. Damit ist aber eine lineare Abbildung $f : V \otimes W$ durch die Werte auf den Elementen der Form $v \otimes w$ vollständig bestimmt. □

11.4. Grundlegende Eigenschaften. Aus der abstrakten Konstruktion in Satz 11.3 ist nicht offensichtlich, wie das Tensorprodukt tatsächlich aussieht. Zunächst ist nur klar, dass man für Elemente $v \in V$ und $w \in W$ immer ein Element $v \otimes w \in V \otimes W$ erhält. Aus der Bilinearität ergeben sich natürlich Eigenschaften wie $\lambda(v \otimes w) = (\lambda v) \otimes w =$

$v \otimes (\lambda w)$ für $\lambda \in \mathbb{K}$ oder $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$. Um einige weitere Eigenschaften abklären zu können, benötigen wir noch ein Resultat.

Lemma 11.4. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$ mit Dualraum $V^* = L(V, \mathbb{K})$.*

(1) Sind $v_1, \dots, v_k \in V$ linear unabhängig, dann findet man lineare Funktionale $\lambda_1, \dots, \lambda_k \in V^*$, sodass $\lambda_i(v_i) = 1$ für alle i und $\lambda_i(v_j) = 0$ für alle $j \neq i$ gilt.

(2) Sind $\lambda_1, \dots, \lambda_k \in V^*$ linear unabhängig, dann findet man Vektoren $v_1, \dots, v_k \in V$, sodass $\lambda_i(v_i) = 1$ für alle i und $\lambda_i(v_j) = 0$ für alle $j \neq i$ gilt.

Beweis. (1) Man erweitert die linear unabhängige Menge $\{v_1, \dots, v_k\}$ zu einer Basis von V (Korollar 4.5 im endlichdimensionalen Fall, sonst Satz 4.6). Dann definiert man λ_i als die eindeutige lineare Abbildung $V \to \mathbb{K}$, die v_i auf 1 und alle anderen Basiselemente auf 0 abbildet.

(2) Betrachte die (offensichtlich lineare) Abbildung $f : V \to \mathbb{K}^k$, die definiert ist durch $f(v) := (\lambda_1(v), \dots, \lambda_k(v))$. Wir behaupten, dass f surjektiv ist. Wäre das nicht der Fall, dann finden wir eine lineare Abbildung $0 \neq \mu : \mathbb{K}^k \to \mathbb{K}$, sodass $\mu \circ f = 0$ gilt. Man kann nämlich eine Basis von $\mathrm{Im}(f)$ zu einer Basis von $\mathbb{K}^k$ erweitern und dann μ so definieren, dass einer der hinzugekommenen Basisvektoren auf 1 und alle anderen auf 0 abgebildet werden. Nun gibt es Elemente $a_1, \dots, a_k \in \mathbb{K}$, von denen mindestens eines ungleich Null ist, sodass $\mu(r_1, \dots, r_k) = a_1 r_1 + \dots + a_k r_k$ gilt. Die Gleichung $\mu \circ f = 0$, sagt dann gerade, dass

$$0 = a_1\lambda_1(v) + \dots + a_k\lambda_k(v) = (a_1\lambda_1 + \dots + a_k\lambda_k)(v)$$

für alle $v \in V$ gilt. Damit ist aber $a_1\lambda_1 + \dots + a_k\lambda_k = 0$, ein Widerspruch zur linearen Unabhängigkeit der λ_i. Damit ist f surjektiv und wir können für jedes i ein Urbild v_i von e_i wählen. □

Damit können wir nun mehr über die Struktur von $V \otimes W$ herausfinden.

Proposition 11.4. *Seien V und W Vektorräume über $\mathbb{K}$.*

(1) Die Menge $\{v \otimes w : v \in V, w \in W\}$ ist ein Erzeugendensystem für $V \otimes W$.

(2) Jedes Element von $V \otimes W$ kann in der Form $\sum_{i=1}^k v_i \otimes w_i$ für ein $k \in \mathbb{N}$, linear unabhängige Vektoren $v_1, \dots, v_k \in V$ und beliebige Vektoren $w_1, \dots, w_k \in W$ geschrieben werden.

(3) Ein Ausdruck der Form $\sum_{i=1}^k v_i \otimes w_i$ mit linear unabhängigen Vektoren $v_1, \dots, v_k$ stellt nur dann den Nullvektor in $V \otimes W$ dar, wenn alle w_i gleich Null sind.

(4) Sind V und W endlichdimensional und sind $\{v_1, \dots, v_n\}$ und $\{w_1, \dots, w_m\}$ Basen für V und W, dann ist $\mathcal{B} := \{v_i \otimes w_j : 1 \leq i \leq n, 1 \leq j \leq m\}$ eine Basis für $V \otimes W$. Insbesondere ist $\dim(V \otimes W) = \dim(V) \dim(W)$.

Beweis. (1) Sei $A \subset V \otimes W$ der von allen Elementen der Form $v \otimes w$ erzeugte Teilraum. Betrachte den Quotientenraum $Z := (V \otimes W)/A$ und die kanonische Surjektion $\pi : V \otimes W \to Z$. Dann gilt nach Konstruktion $\pi(v \otimes w) = 0$ für alle $v \in V$ und $w \in W$. Da aber die Nullabbildung die gleiche Eigenschaft hat, muss nach der Eindeutigkeit in der universellen Eigenschaft $\pi = 0$ gelten. Damit ist aber $A = \mathrm{Ker}(\pi) = V \otimes W$.

(2) Wir bemerken zunächst, dass nach der Bilinearität $v \otimes 0 = 0 \otimes w = 0$ für alle $v \in V$ und $w \in W$ gilt. Damit kann man den Nullvektor als $v \otimes 0$ mit $v \neq 0$ (also linear unabhängig) darstellen.

Nach Teil (1) gibt es für ein Element $\xi \in V \otimes W$ endlich viele Vektoren $v_1, \dots, v_\ell \in V$ und $w_1, \dots, w_\ell \in W$, sodass $\xi = \sum_{i=1}^\ell v_i \otimes w_i$ gilt. Nachdem man Terme mit gleicher v–Komponente offensichtlich zusammenfassen kann, dürfen wir ohne Beschränkung der

Allgemeinheit annehmen, dass die v_i alle verschieden und $\neq 0$ sind. Wenn die v_i linear unabhängig sind, dann sind wir fertig.

Wenn nicht, dann behaupten wir, dass wir nach eventuellem Umnummerieren eine Darstellung der Form $\xi = \sum_{i=1}^{\ell-1} v_i \otimes \tilde{w}_i$ erhalten. Nach Satz 4.2 können wir nämlich, eventuell nach Umnummerieren, v_ℓ als Linearkombination von $v_1, \dots, v_{\ell-1}$ schreiben, erhalten also $v_\ell = \sum_{i=1}^{\ell-1} a_i v_i$ mit $a_i \in \mathbb{K}$. Damit ist aber $v_\ell \otimes w_\ell = \sum_{i=1}^{\ell-1} v_i \otimes (a_i w_\ell)$ und man muss nur noch $\tilde{w}_i = w_i + a_i w_\ell$ für alle i setzen. In endlich vielen Schritten führt das zu einer Darstellung der gewünschten Form oder zu $\xi = 0$, wofür wir oben schon eine passende Darstellung gefunden haben.

(3) Sei zunächst $\lambda : V \to \mathbb{K}$ ein lineares Funktional. Betrachte die Abbildung $\varphi_\lambda : V \times W \to W$, die durch $\varphi_\lambda(v, w) := \lambda(v)w$ gegeben ist. Diese Abbildung ist offensichtlich bilinear, also gibt es eine eindeutige lineare Abbildung $\tilde{\varphi}_\lambda : V \otimes W \to W$, die $v \otimes w$ auf $\lambda(v)w$ abbildet. Haben wir linear unabhängige Vektoren $v_1, \dots, v_k \in V$ gegeben, dann finden wir nach Lemma 11.4 lineare Funktionale $\lambda_1, \dots, \lambda_k \in V^*$, sodass $\lambda_i(v_i) = 1$ und $\lambda_i(v_j) = 0$ für $i \neq j$ ist. Ist nun $0 = \sum_{j=1}^k v_j \otimes w_j$, dann folgt

$$0 = \tilde{\varphi}_{\lambda_i}(\textstyle\sum_j v_j \otimes w_j) = \sum_j \lambda_i(v_j) w_j = w_i.$$

(4) Sind $v \in V$ und $w \in W$, dann finden wir Skalare a_i und b_j, sodass $v = \sum_i a_i v_i$ und $w = \sum_j b_j w_j$. Damit ist aber $v \otimes w = \sum_{i,j} a_i b_j v_i \otimes v_j$. Damit liegt $v \otimes w$ im Erzeugnis von $\mathcal{B}$ und nach (1) ist $\mathcal{B}$ ein Erzeugendensystem für $V \otimes W$.

Seien andererseits $a_{ij} \in \mathbb{K}$ Skalare, sodass $\sum_{i,j} a_{ij} v_i \otimes w_j = 0$ gilt. Dann schreiben wir das als $0 = \sum_i v_i \otimes (\sum_j a_{ij} w_j)$. Da die v_i linear unabhängig sind, folgt $\sum_j a_{ij} w_j = 0$ für alle $i = 1, \dots, n$ aus (3). Wegen der linearen Unabhängigkeit der w_j ist das nur möglich, wenn $a_{ij} = 0$ für alle $j = 1, \dots, m$ gilt. Damit ist $\mathcal{B}$ linear unabhängig, also eine Basis. Die Aussage über die Dimension folgt offensichtlich. $\square$

Aus diesen Resultaten folgt auch, dass allgemeine Elemente von $V \otimes W$ *nicht* in der Form $v \otimes w$ geschrieben werden können. Seien etwa v_1, v_2 linear unabhängig in V und w_1, w_2 linear unabhängig in W. Dann gilt $v_1 \otimes w_1 + v_2 \otimes w_2 \neq v \otimes w$ für alle $v \in V$ und $w \in W$. Sonst könnte man nämlich $0 = v_1 \otimes w_1 + v_2 \otimes w_2 + v \otimes (-w)$ schreiben. Nach Teil (3) des Lemmas können v_1, v_2, v nicht linear unabhängig sein, also folgt leicht, dass $v = a_1 v_1 + a_2 v_2$ für passendes $a_1, a_2 \in \mathbb{K}$ gelten muss. Dann ist aber $v \otimes (-w) = v_1 \otimes (-a_1 w) + v_2 \otimes (-a_2 w)$ und setzt man das ein, dann erhält man $0 = v_1 \otimes (w_1 - a_1 w) + v_2 \otimes (w_2 - a_2 w)$. Nochmals nach Teil (3) der Proposition folgt daraus $w_1 = a_1 w$ und $w_2 = a_2 w$, ein Widerspruch zur linearen Unabhängigkeit.

Man nennt Elemente der Form $v \otimes w$ in $V \otimes W$ *zerlegbare Tensoren.* Wir werden in Kürze ein Beispiel besprechen, in dem man sehr handlich verstehen kann, welche Elemente zerlegbar sind.

11.5. Anwendung: Komplexifizierungen. Wir haben schon öfters benutzt, dass man eine Matrix mit reellen Eintragungen auch als komplexe Matrix ansehen kann oder allgemeiner eine Matrix mit Eintragungen in $\mathbb{K}$ als Matrix über dem algebraischen Abschluss von $\mathbb{K}$. Dahinter steckt ein Zusammenhang von $\mathbb{R}^n$ und $\mathbb{C}^n$ der für diese beiden Räume ziemlich offensichtlich ist. Mit Hilfe von Tensorprodukten kann man ein Analogon dieser Konstruktion für allgemeine Vektorräume angeben.

Aus Teil (2) von Beispiel 2.2 wissen wir, dass die Einschränkung von Addition und Multiplikation $\mathbb{C}$ zu einem Vektorraum über $\mathbb{R}$ macht. Für einen beliebigen $\mathbb{R}$–Vektorraum V definieren wir nun die *Komplexifizierung* $V_\mathbb{C}$ von V durch $V_\mathbb{C} := \mathbb{C} \otimes_\mathbb{R} V$. Das ist ein Vektorraum über $\mathbb{R}$, aber man kann leicht eine Skalarmultiplikation mit komplexen Zahlen definieren. Für fixes $\lambda \in \mathbb{C}$ betrachtet man die Abbildung $\mathbb{C} \times V \to V_\mathbb{C}$, die durch $(\mu, v) \mapsto \lambda\mu \otimes v$ gegeben ist. Diese ist offensichtlich bilinear über $\mathbb{R}$, also erhält man eine eindeutige $\mathbb{R}$–lineare Abbildung $V_{\mathbb{C}} \to V_{\mathbb{C}}$, die ein Element der Form $\mu \otimes v$ auf $(\lambda\mu) \otimes v$ abbildet. Definiert man dadurch die Multiplikation mit λ, dann erhält man eine Funktion $\mathbb{C} \times V_{\mathbb{C}} \to V_{\mathbb{C}}$, die durch $\lambda \cdot (\mu \otimes v) = (\lambda\mu) \otimes v$ charakterisiert ist.

Proposition 11.5. *Sei V ein $\mathbb{R}$–Vektorraum mit Komplexifizierung $\mathbb{V}_{\mathbb{C}}$ und sei $j : V \to V_{\mathbb{C}}$ definiert durch $j(v) := 1 \otimes v$.*

(1) Die oben definierte Skalarmultiplikation macht $V_{\mathbb{C}}$ zu einem komplexen Vektorraum. Ist V endlichdimensional, dann auch $V_{\mathbb{C}}$ und $\dim_{\mathbb{C}}(V_{\mathbb{C}}) = \dim_{\mathbb{R}}(V)$. Ist $\mathcal{B} = \{v_1, \dots, v_n\}$ eine Basis für V, dann ist $\{j(v_1), \dots, j(v_n)\}$ eine $\mathbb{C}$–Basis für $V_{\mathbb{C}}$.

(2) Für einen beliebigen $\mathbb{C}$–Vektorraum W und eine $\mathbb{R}$–lineare Abbildung $f : V \to W$ gibt es eine eindeutige $\mathbb{C}$–lineare Abbildung $\hat{f} : V_{\mathbb{C}} \to W$, die $\hat{f} \circ j = f$ erfüllt.

Beweis. (1) Die verlangten Eigenschaften der Skalarmultiplikation folgen leicht aus der definierenden Gleichung. Für die Basis $\mathcal{B}$ bilden nach Proposition 11.4 die Elemente $1 \otimes v_k = j(v_k)$ und $i \otimes v_k = i \cdot j(v_k)$ eine $\mathbb{R}$–Basis für $V_{\mathbb{C}}$. Damit ist einerseits $\dim_{\mathbb{R}}(V_{\mathbb{C}}) = 2n$ also $\dim_{\mathbb{C}}(V_{\mathbb{C}}) = n$. Andererseits folgt auch, dass die Elemente $j(v_1), \dots, j(v_n)$ ein Erzeugendensystem für den $\mathbb{C}$–Vektorraum $V_{\mathbb{C}}$ und damit eine Basis bilden.

(2) Haben wir $f : V \to W$ gegeben, dann definieren wir $\varphi : \mathbb{C} \times V \to W$ durch $\varphi(\lambda, v) := \lambda \cdot f(v)$. Diese Abbildung ist offensichtlich $\mathbb{R}$–bilinear, also erhalten wir eine $\mathbb{R}$–lineare Abbildung $\hat{f} = \tilde{\varphi} : V_{\mathbb{C}} \to W$. Diese ist durch $\hat{f}(\lambda \otimes v) = \lambda \cdot f(v)$ charakterisiert, erfüllt also $\hat{f} \circ j = f$. Andererseits folgt aber

$$\hat{f}(\lambda \cdot (\mu \otimes V)) = \hat{f}((\lambda\mu) \otimes v) = \lambda\mu \cdot f(v) = \lambda \cdot \hat{f}(\mu \otimes v),$$

und damit ist $\hat{f}$ $\mathbb{C}$–linear. Nachdem $j(V) \subset V_{\mathbb{C}}$ ein Erzeugendensystem über $\mathbb{C}$ ist, ist jede $\mathbb{C}$–lineare Abbildung auf $V_{\mathbb{C}}$ durch ihre Komposition mit j eindeutig bestimmt. □

Für die Komplexifizierung gibt es auch direktere Konstruktionen. Beispielsweise kann man $V \times V$ mit der komponentenweisen Addition betrachten und die Skalarmultiplikation durch $(a + ib)(v_1, v_2) := (av_1 - bv_2, bv_1 + av_2)$ definieren und dann alle Eigenschaften direkt verifizieren.

Die Konstruktion mit dem Tensorprodukt ist aber viel allgemeiner. Sie funktioniert völlig analog für beliebige Körpererweiterungen. Ist $\mathbb{K} \subset \mathbb{L}$ ein Teilkörper, dann bezeichnet man $\mathbb{L}$ als *Körpererweiterung* von $\mathbb{K}$. Aus den Körpereigenschaften folgt sofort, dass die Einschränkung der Multiplikation zusammen mit der Addition $\mathbb{L}$ zu einem $\mathbb{K}$–Vektorraum macht. Dann kann man einem $\mathbb{K}$–Vektorraum V den Raum $V_{\mathbb{L}} := \mathbb{L} \otimes_{\mathbb{K}} V$ zuordnen, der in natürlicher Weise ein $\mathbb{L}$–Vektorraum ist. Die Abbildung $j(v) := 1 \otimes V$ definiert eine $\mathbb{K}$–lineare Abbildung $V \to V_{\mathbb{L}}$, die eine universelle Eigenschaft analog zu Teil (2) von Proposition 11.5 besitzt. Analoge Konstruktionen funktionieren auch über Ringen.

11.6. Das Tensorprodukt als Operation. Zunächst bemerken wir, dass auch das Tensorprodukt nicht nur eine Konstruktion für Vektorräume, sondern auch für lineare Abbildungen ist. Seien V_1, V_2, W_1 und W_2 Vektorräume über $\mathbb{K}$ und $f : V_1 \to V_2$ und $g : W_1 \to W_2$ lineare Abbildungen. Betrachte die Funktion $V_1 \times W_1 \to V_2 \otimes W_2$, die definiert ist durch $(v, w) \mapsto f(v) \otimes g(w)$. Man verifiziert sofort, dass diese Abbildung bilinear ist, also gibt es eine eindeutige lineare Abbildung $f \otimes g : V_1 \otimes W_1 \to V_2 \otimes W_2$, die $(f \otimes g)(v \otimes w) = f(v) \otimes g(w)$ erfüllt. Man nennt diese Abbildung das *Tensorprodukt der linearen Abbildungen* f und g. Auch hier verifiziert man leicht die funktoriellen Eigenschaften $(f_1 \circ f_2) \otimes (g_1 \circ g_2) = (f_1 \otimes g_1) \circ (f_2 \otimes g_2)$ und $\mathrm{id} \otimes \mathrm{id} = \mathrm{id}$.

Damit kann man im weiteren die Aussagen über Natürlichkeit präzise machen, wie wir das in 11.1 besprochen haben. Wir werden diese Verifikationen aber meist auslassen und uns damit begnügen festzustellen, dass man Abbildungen "auf natürliche Weise angeben" kann. Wir verifizieren nun Eigenschaften des Tensorprodukts, die analog zu Assoziativität, Kommutativität und der Existenz eines neutralen Elements sind.

Proposition 11.6. *Seien U, V und W Vektorräume über einem Körper $\mathbb{K}$. Dann gibt es für die Tensorprodukte über $\mathbb{K}$ folgende natürliche Isomorphismen: $U\otimes(V\otimes W)\cong (U\otimes V)\otimes W$, $V\otimes W\cong W\otimes V$ und $\mathbb{K}\otimes V\cong V$.*

Beweis. Sei $\varphi:\mathbb{K}\times V\to V$ die Skalarmultiplikation, also $\varphi(r,v):=r\cdot v$. Dann sagen die entsprechenden Vektorraumaxiome, dass φ bilinear ist, also eine eindeutige lineare Abbildung $\tilde{\varphi}:\mathbb{K}\otimes V\to V$ induziert. Umgekehrt definiert $v\mapsto 1\otimes v$ eine lineare Abbildung $V\to\mathbb{K}\otimes V$, und man verifiziert sofort, dass diese invers zu $\tilde{\varphi}$ ist. Damit folgt der letzte der Isomorphismen.

Analog definiert $\varphi(v,w):=w\otimes v$ eine bilineare Abbildung $V\times W\to W\otimes V$, die eine lineare Abbildung $\tilde{\varphi}:V\otimes W\to W\otimes V$ induziert. Analog konstruiert man eine inverse lineare Abbildung $W\otimes V\to V\otimes W$.

Für den ersten der behaupteten Isomorphismen konstruiert man analog in zwei Schritten eine lineare Abbildung $U\otimes(V\otimes W)\to(U\otimes V)\otimes W$, die $u\otimes(v\otimes w)$ auf $(u\otimes v)\otimes w$ abbildet und ebenso eine Inverse dazu. □

Durch dieses Resultat kann man das Tensorprodukt von drei Vektorräumen einfach als $U\otimes V\otimes W$ schreiben und erhält darin Elemente der Form $u\otimes v\otimes w$. Man sieht leicht, dass dieses Tensorprodukt eine universelle Eigenschaft besitzt. Ist nämlich Z ein beliebiger $\mathbb{K}$–Vektorraum und $\varphi:U\times V\times W\to Z$ eine trilineare Abbildung (also linear in der dritten Variable, sofern man zwei der drei Variablen festhält). Dann gibt es eine eindeutige lineare Abbildung $\tilde{\varphi}:U\otimes V\otimes W\to Z$, sodass $\tilde{\varphi}(u\otimes v\otimes w)=\varphi(u,v,w)$ gilt.

Allgemeiner kann man für Vektorräume $V_1,\dots,V_k$ das Tensorprodukt $V_1\otimes\dots\otimes V_k$ bilden, das eine universelle Eigenschaft für k–lineare Abbildungen $V_1\times\ldots\times V_k\to Z$ besitzt. Hat man für $i=1,\dots,k$ weitere Vektorräume W_i und lineare Abbildungen $f_i:V_i\to W_i$ gegeben, dann erhält man

$$f_1\otimes\dots\otimes f_k:V_1\otimes\dots\otimes V_k\to W_1\otimes\dots\otimes W_k,$$

und wieder sind funktorielle Eigenschaften erfüllt. Setzt man in einem k–fachen Tensorprodukt $V_i=V$ für einen fixen Vektorraum V und alle $i=1,\dots,k$, dann erhält man die *kte Tensorpotenz* von V, die mit $\otimes^k V$ bezeichnet wird und die eine universelle Eigenschaft für k–lineare Abbildungen $V^k\to Z$ in beliebige $\mathbb{K}$–Vektorräume Z besitzt. Eine lineare Abbildung $f:V\to W$ liefert dann $\otimes^k f:\otimes^k V\to\otimes^k W$.

An dieser Stelle ist etwas Vorsicht geboten. Der natürliche Isomorphismus $V\otimes W\to W\otimes V$ aus Proposition 11.6, der durch die Vertauschung der Elemente induziert wird, bedeutet nicht, dass es für linear unabhängige Vektoren v_1 und v_2 in V einen Zusammenhang zwischen den Elementen $v_1\otimes v_2$ und $v_2\otimes v_1$ von $V\otimes V=\otimes^2 V$ gibt. Das sind vollkommen unabhängige Elemente. Analog kommt es auch in höheren Tensorpotenzen stark auf die Reihenfolge der Elemente an.

11.7. Einige weitere natürliche Abbildungen und Isomorphismen.

Proposition 11.7. *Seien V und W Vektorräume über $\mathbb{K}$. Dann gilt:*

(1) Für jeden $\mathbb{K}$–Vektorraum Z Sind die Räume $L(V\otimes W,Z)$, $L(V,L(W,Z))$ und $L(W,L(V,Z))$ natürlich isomorph. Insbesondere gilt $(V\otimes W)^*\cong L(V,W^*)\cong L(W,V^*)$.

(2) Es gibt eine natürliche injektive Abbildung $V^* \otimes W^* \to (V \otimes W)^*$, die ein Isomorphismus ist, falls V und W endlichdimensional sind.

(3) Es gibt eine injektive natürliche lineare Abbildung $V^* \otimes W \to L(V,W)$, deren Bild die Menge all jener linearen Abbildung $f : V \to W$ ist, für die $\mathrm{Im}(f)$ endlichdimensional ist. Insbesondere ist $V^* \otimes W \cong L(V,W)$, falls V oder W endlichdimensional ist.

(4) Es gibt eine eindeutige natürliche Surjektion $\mathcal{C} : V^* \otimes V \to \mathbb{K}$, die $\mathcal{C}(\lambda \otimes v) = \lambda(v)$ erfüllt. Ist V endlichdimensional, dann findet man auch eine natürliche Injektion $j : \mathbb{K} \to V^* \otimes V$, sodass $\mathcal{C} \circ j$ die Multiplikation mit $\dim(V)$ ist.

Beweis. (1) Für eine bilineare Abbildung $\varphi : V \times W \to Z$ erhält man nach Definition des Tensorprodukts eine eindeutige lineare Abbildung $\tilde{\varphi} : V \otimes W \to Z$ sodass $\tilde{\varphi}(v \otimes w) = \varphi(v,w)$ gilt. Das definiert eine Funktion $B(V,W;Z) \to L(V \otimes W, Z)$. Für $\varphi, \psi \in B(V,W;Z)$ und λ in $\mathbb{K}$ gilt aber

$$(\tilde{\varphi} + \lambda\tilde{\psi})(v \otimes w) = \varphi(v,w) + \lambda\psi(v,w) = (\varphi + \lambda\psi)(v,w).$$

Damit folgt $\widetilde{\varphi + \lambda\psi} = \tilde{\varphi} + \lambda\tilde{\psi}$ also ist unsere Abbildung linear. Die Bijektivität der Abbildung folgt unmittelbar aus der universellen Eigenschaft, also ist $B(V,W;Z) \cong L(V \otimes W, Z)$ und damit folgen die restlichen Isomorphismen aus Proposition 11.3. Die Natürlichkeit überlegt man leicht, siehe Bemerkung 11.7. Für $Z = \mathbb{K}$ erhält man sofort die Aussage über die Dualräume.

(2) Für $\lambda \in V^*$, also $\lambda : V \to \mathbb{K}$, und $\mu \in W^*$ definiert das Tensorprodukt $\lambda \otimes \mu$ von linearen Abbildungen eine lineare Abbildung $V \otimes W \to \mathbb{K} \otimes \mathbb{K}$. Benutzt man den Isomorphismus $\mathbb{K} \otimes \mathbb{K} \to \mathbb{K}$ aus Proposition 11.6, dann kann man $\lambda \otimes \mu$ als lineares Funktional auf $V \otimes W$ betrachten, das durch $(\lambda \otimes \mu)(v \otimes w) = \lambda(v)\mu(w)$ charakterisiert ist. Man verifiziert sofort, dass die Abbildung $V^* \times W^* \to L(V \otimes W, \mathbb{K})$, die (λ, μ) auf $\lambda \otimes \mu$ abbildet, bilinear ist, also erhält man eine lineare Abbildung $V^* \otimes W^* \to L(V \otimes W, \mathbb{K})$.

Um zu zeigen, dass diese natürliche Abbildung injektiv ist, genügt es nach Proposition 11.4 zu zeigen, dass für linear unabhängige Funktionale $\lambda_1, \dots, \lambda_k \in V^*$ und beliebige $\mu_1, \dots, \mu_k \in W^*$ die Abbildung $\sum \lambda_i \otimes \mu_i : V \otimes W \to \mathbb{K}$ nur dann gleich Null ist, wenn alle μ_i gleich 0 sind. Nach Lemma 11.4 finden wir $v_1, \dots, v_k \in V$, sodass $\lambda_i(v_i) = 1$ für alle i und $\lambda_i(v_j) = 0$ für $i \neq j$ gilt. Ist $\sum_i \lambda_i \otimes \mu_i$ die Nullabbildung, dann folgt

$$0 = (\textstyle\sum_j \lambda_j \otimes \mu_j)(v_i \otimes w) = \sum_j \lambda_j(v_i)\mu_j(w) = \mu_i(w)$$

für beliebiges $w \in W$ und damit $\mu_i = 0$.

(3) Für $\lambda \in V^*$, also $\lambda : V \to \mathbb{K}$ linear und $w \in W$, betrachte die Abbildung $V \to W$, die v auf $\lambda(v)w$ abbildet. Diese Abbildung ist offensichtlich linear, also haben wir eine Abbildung $V^* \times W \to L(V,W)$ definiert. Man verifiziert sofort, dass diese Abbildung bilinear ist, also induziert sie nach der universellen Eigenschaft eine lineare Abbildung $V^* \otimes W \to L(V,W)$. Die Natürlichkeit dieser Abbildung ist leicht zu verifizieren.

Um zu sehen, dass diese Abbildung injektiv ist, müssen wir nach Proposition 11.4 wieder nur zeigen, dass für linear unabhängige Funktionale $\lambda_1, \dots, \lambda_k$ auf V und beliebige Elemente $w_1, \dots, w_k \in W$, die lineare Abbildung $V \to W$ zu $\sum_i \lambda_i \otimes w_i$ nur dann Null ist, wenn alle $w_i = 0$ sind. Nach Lemma 11.4 finden wir wieder $v_1, \dots, v_k \in V$, sodass $\lambda_i(v_j)$ gleich 1 für $i = j$ und gleich 0 für $i \neq j$ ist. Die lineare Abbildung $\sum_j \lambda_j \otimes w_j$ bildet aber dann genau v_i auf w_i ab, also folgt die Behauptung.

Ist f die lineare Abbildung zu $\sum_{i=1}^k \lambda_i \otimes w_i$, dann gilt $f(v) = \sum_i \lambda_i(v)w_i$, also liegt $f(v)$ immer im Erzeugnis von $\{w_1, \dots, w_k\}$. Damit ist aber $\mathrm{Im}(f)$ in diesem Erzeugnis enthalten und damit endlichdimensional.

Sei umgekehrt $f : V \to W$ eine lineare Abbildung mit endlichdimensionalem Bild $\mathrm{Im}(f)$. Wählt man eine Basis $\{w_1, \dots, w_k\}$ für $\mathrm{Im}(f)$, dann gibt es zu jedem Vektor $v \in V$ eindeutige Elemente $\lambda_1(v), \dots, \lambda_k(v) \in \mathbb{K}$, sodass $f(v) = \sum_{i=1}^k \lambda_i(v) w_i$ gilt. Das definiert Funktionen $\lambda_1, \dots, \lambda_k : V \to \mathbb{K}$, die klarerweise linear sind und offensichtlich ist f dann die lineare Abbildung zu $\sum_{i=1}^k \lambda_i \otimes w_i$.

Ist V oder W endlichdimensional, dann hat jede lineare Abbildung $f : V \to W$ endlichdimensionales Bild, also folgt die letzte Behauptung.

(4) Betrachte die Abbildung $V^* \times V \to \mathbb{K}$, die (λ, v) auf $\lambda(v)$ abbildet. Diese ist offensichtlich bilinear und liefert daher eine lineare Abbildung $\mathcal{C} : V^* \otimes V \to \mathbb{K}$ wie behauptet.

Für jeden Vektorraum V definiert $r \mapsto r\,\mathrm{id}_V$ eine natürliche, injektive lineare Abbildung $\mathbb{K} \to L(V, V)$. Ist V endlichdimensional, dann kann man nach (3) den Raum $L(V, V)$ mit $V^* \otimes V$ identifizieren. Dort sehen wir auch, wie man $j(r) \in V^* \otimes V$ explizit beschreiben kann: Man wählt eine Basis $\mathcal{B} := \{v_1, \dots, v_n\}$ für V. Dann benötigt man Funktionale λ_i, sodass $r\,\mathrm{id}(v) = \sum_i \lambda_i(v) v_i$ gilt. Die Elemente $\{\varphi_i\}$ der dualen Basis zu $\mathcal{B}$ sind erfüllen genau das für $r = 1$, also gilt $j(r) = r \sum_i \varphi_i \otimes v_i$. Daraus folgt aber $\mathcal{C}(j(r)) = \dim(V) r$ sofort. $\square$

Bemerkung 11.7. (1) Um die Natürlichkeit in Teil (1) der Proposition explizit zu verifizieren, muss man zunächst überlegen, was das alles eigentlich bedeutet. Unser Isomorphismus $L(V, L(W, Z)) \to L(V \otimes W, Z)$ ist dadurch charakterisiert, dass er aus einer Abbildung $\Phi : V \to L(W, Z)$ die eindeutige lineare Abbildung $\Psi : V \otimes W \to Z$ macht, die $\Psi(v \otimes w) = \Phi(v)(w)$ erfüllt. Nehmen wir nun an, dass wir lineare Abbildungen $f : \hat{V} \to V$, $g : \hat{W} \to W$ und $h : Z \to \hat{Z}$ gegeben haben. Dann erhalten wir Abbildungen $L(V, L(W, Z)) \to L(\hat{V}, L(\hat{W}, \hat{Z}))$ und $L(V \otimes W, Z) \to L(\hat{V} \otimes \hat{W}, \hat{Z})$. Das Bild von Φ unter der ersten Abbildung ist dadurch charakterisiert, dass man $\hat{v} \in \hat{V}$ die lineare Abbildung $h \circ \Phi(f(\hat{v})) \circ g$ zuordnet. Das Bild von Ψ unter der zweiten Abbildung ist einfach $h \circ \Psi \circ (f \otimes g)$. Wendet man das auf $\hat{v} \otimes \hat{w}$ an, dann erhält man

$$h(\Psi(f(\hat{v}) \otimes g(\hat{w}))) = h\big(\Phi(f(\hat{v}))(g(\hat{w}))\big).$$

Damit folgt aber, dass unser Isomorphismus mit den beiden Abbildungen verträglich ist.

(2) Der Beweis von Teil (3) der Proposition liefert auch die in 11.4 angekündigte schöne Interpretation von zerlegbaren Tensoren. Wir können $V^* \otimes W$ mit dem Raum aller linearen Abbildungen $f : V \to W$ identifizieren, die endlichdimensionales Bild, also endlichen Rang haben. Aus dem Beweis sieht man, dass zerlegbare Tensoren (ungleich Null) genau den linearen Abbildungen mit Rang 1 entsprechen.

(3) Wenn man bereit ist, sich auf Fälle zu beschränken, in denen zumindest ein Faktor endlichdimensional ist, dann kann man ein Tensorprodukt auch ohne Umweg über den freien Vektorraum direkt konstruieren. Ist etwa V endlichdimensional, dann kann man $V \otimes W$ einfach als $L(V^*, W)$ realisieren. Dazu definiert man $b : V \times W \to L(V^*, W)$ indem man (v, w) auf die Abbildung $\lambda \mapsto \lambda(v) w$ abbildet.

Sei nun $\Phi : V \times W \to Z$ eine bilineare Abbildung. Sei $\mathcal{B} := \{v_1, \dots, v_n\}$ eine Basis für V und $\mathcal{B}^* = \{\varphi_1, \dots, \varphi_n\}$ die duale Basis für V^*. Für $f \in L(V^*, W)$ betrachte dann $\tilde{\Phi}(f) := \sum_i \Phi(v_i, f(\varphi_i)) \in Z$. Man verifiziert leicht, dass dies eine lineare Abbildung $\tilde{\Phi} : L(V^*, W) \to Z$ definiert. Für $v \in V$ gilt $v = \sum_j \varphi_j(v) v_j$ und für $\varphi \in V^*$ erhält man $\varphi = \sum_j \varphi(v_j) \varphi_j$. Die lineare Abbildung $b(v, w) : V^* \to W$ ist daher durch $b(v, w)(\varphi) = \varphi(v) w = \sum_j \varphi(v_j) \varphi_j(v) w$ gegeben und erfüllt daher $b(v, w)(\varphi_i) = \varphi_i(v) w$. Damit ist aber

$$\tilde{\Phi}(b(v,w)) = \textstyle\sum_i \Phi(v_i, \varphi_i(v)w) = \Phi(\sum_i \varphi_i(v)v_i, w) = \Phi(v,w).$$

Nun zeigt man leicht, dass für eine lineare Abbildung $f : V^* \to W$ die Gleichung $f = \sum_i b(v_i, f(\varphi_i))$ gilt. Damit ist eine lineare Abbildung von $L(V^*, W)$ in einen Vektorraum durch ihre Werte auf den Elementen $b(v,w)$ eindeutig bestimmt. Also ist $(L(V^*, W), b)$ ein Tensorprodukt von V und W.

Etwas mühsamer kann man auch nachrechnen, dass $\tilde{\Phi}$ unabhängig von der Wahl der Basis v_i ist. Dazu benutzt man, dass es für eine andere Basis $\mathcal{C} := \{w_1, \dots, w_n\}$ mit dualer Basis $\{\psi_1, \dots, \psi_n\}$ und die Basiswechselmatrizen $A = [\mathrm{id}]^{\mathcal{C}}_{\mathcal{B}} = (a_{ij})$ und $A^{-1} = [\mathrm{id}]^{\mathcal{B}}_{\mathcal{C}} = (b_{ij})$ die Gleichungen $w_j = \sum_i a_{ij} v_i$ und $v_j = \sum_i b_{ij} w_i$ für alle j gelten. Wendet man auf die erste Gleichung φ_i an, dann folgt $\varphi_i(w_j) = a_{ij}$ und damit $\varphi_i = \sum_j a_{ij}\psi_j$. Das setzt man gemeinsam mit der Formel für v_j in die Definition von $\tilde{\Phi}$ ein und daraus folgt das Resultat durch direkte Rechnung.

(4) Sei V endlichdimensional und $f : V \to V$ linear. Aus dem Beweis von Teil (3) von Proposition 11.7 sehen wir, wie man f als Element von $V^* \otimes V$ realisiert. Für eine Basis $\mathcal{B} = \{v_1, \dots, v_n\}$ definiert man $\lambda_1, \dots, \lambda_n : V \to \mathbb{K}$ durch $f(v) = \sum_i \lambda_i(v) v_i$, also gilt $[f(v)]_{\mathcal{B}} = (\lambda_1(v), \dots, \lambda_n(v))$. Das bedeutet aber gerade, dass die Matrixdarstellung $[f]_{\mathcal{B}} = (a_{ij})$ durch $a_{ij} = \lambda_i(v_j)$ gegeben ist. Als Element von $V^* \otimes V$ ist f durch $\sum_i \lambda_i \otimes v_i$ gegeben. Die Abbildung $\mathcal{C}$ aus Teil (4) von Proposition 11.7 bildet dieses Element auf $\sum_i(\lambda_i(v_i)) = \sum_i a_{ii} = \mathrm{tr}([f]_{\mathcal{B}}) = \mathrm{tr}(f)$ ab. Deshalb wird $\mathcal{C}$ als *Spur* oder als *Kontraktion* bezeichnet. Das liefert eine Definition der Spur ohne den Umweg über Matrizen.

(5) Mit Hilfe der Kontraktion kann man viele natürliche Operationen in schöner Weise darstellen. Zunächst kann man die Kontraktion problemlos erweitern. Ist W ein beliebiger Vektorraum, dann erhält man $\mathcal{C} \otimes \mathrm{id}_W : V^* \otimes V \otimes W \to W$. Kommen in einem Tensorprodukt mehrere Faktoren V und V^* vor, dann kann man jeweils einen Faktor auswählen und darauf eine Kontraktion anwenden. So erhält man etwa eine Kontraktion $V^* \otimes W \otimes V \to W$. Fasst man den Raum als Teilraum von $L(V, W) \otimes V$ auf, dann ist diese Kontraktion durch $f \otimes v \mapsto f(v)$ gegeben. Analog können wir eine Kontraktion $V^* \otimes W \otimes W^* \to V^*$ definieren. Wendet man diese auf $f \otimes \mu$ an, dann erhält man ein Element $\hat{f}(\mu) \in V^*$ und hat so $f \in V^* \otimes W$ eine lineare Abbildung $\hat{f} : W^* \to V^*$ zugeordnet.

Betrachtet man nun $V^* \otimes W \otimes V \otimes W^*$, dann kann man zunächst nach $W \otimes W^*$ und dann nach $\mathbb{K}$ kontrahieren oder erst nach $V^* \otimes V$ und dann weiter nach $\mathbb{K}$. Diese beiden Kontraktionen müssen übereinstimmen, weil sie beide der 4–linearen Abbildung $(\lambda, w, v, \mu) \mapsto \lambda(v)\mu(w)$ entsprechen. Betrachten wir das wieder als Teilraum von $L(V, W) \otimes V \otimes W^*$, dann erhält man einerseits $f \otimes v \otimes \mu \mapsto f(v) \otimes \mu \mapsto \mu(f(v))$. Andererseits ergibt sich $f \otimes \mu \otimes v \mapsto \hat{f}(\mu) \otimes v \mapsto (\hat{f}(\mu))(v)$. Man erhält also gerade $\hat{f}(\mu) = \mu \circ f$, also liefert das genau die duale Abbildung zu f.

Symmetrische und äußere Potenzen

11.8. Symmetrische Potenzen. Seien V und Z Vektorräume über einem Körper $\mathbb{K}$ und sei $\varphi : V \times V \to Z$ eine bilineare Abbildung. Analog zu 8.5 nennt man φ *symmetrisch*, wenn $\varphi(v_1, v_2) = \varphi(v_2, v_1)$ für alle $v_1, v_2 \in V$ gilt. Das lässt sich problemlos auf mehrere Argumente, also auf k–lineare Abbildungen $V^k \to Z$ verallgemeinern. Hier verlangt man einfach, dass der Wert von φ unverändert bleibt, wenn man zwei der Argumente von φ vertauscht. Daraus folgt dann sofort, dass für jede Permutation $\sigma \in \mathfrak{S}_k$ die Gleichung $\varphi(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \varphi(v_1, \dots, v_k)$ gilt.

Nun kann man sich natürlich fragen, ob es analog zu den Tensorpotenzen auch Räume gibt, die eine universelle Eigenschaft für symmetrische k–lineare Abbildungen haben. Analog zu den bisher besprochenen Fällen ist so ein Raum, wenn er existiert, durch die universelle Eigenschaft bis auf natürliche Isomorphie eindeutig bestimmt, also muss man sich wieder nicht darum kümmern, wie der Raum konstruiert wurde.

PROPOSITION 11.8. *Sei V ein Vektorraum über $\mathbb{K}$ und sei $k \geq 2$. Dann gilt:*

(1) Es gibt einen Vektorraum S^kV und eine symmetrische, k–lineare Abbildung $V^k \to S^kV$, die man als $(v_1, \dots, v_k) \mapsto v_1 \vee \dots \vee v_k$ schreibt, sodass folgende universelle Eigenschaft erfüllt ist: Für einen $\mathbb{K}$–Vektorraum Z und eine k–lineare, symmetrische Abbildung $\varphi : V^k \to Z$ gibt es eine eindeutige lineare Abbildung $\tilde{\varphi} : S^kV \to Z$, sodass $\tilde{\varphi}(v_1 \vee \dots \vee v_k) = \varphi(v_1, \dots, v_k)$ gilt.

(2) Ist W noch ein $\mathbb{K}$–Vektorraum und $f : V \to W$ eine lineare Abbildung, dann erhalten wir eine induzierte lineare Abbildung $S^kf : S^kV \to S^kW$, die durch

$$(S^kf)(v_1 \vee \dots \vee v_k) = f(v_1) \vee \dots \vee f(v_k)$$

charakterisiert ist. Es gilt $S^k(g \circ f) = S^kg \circ S^kf$ und $S^k \operatorname{id}_V = \operatorname{id}_{S^kV}$.

(3) Ist V endlichdimensional und $\{v_1, \dots, v_n\}$ eine Basis für V, dann bilden die Elemente $v_{i_1} \vee \dots \vee v_{i_k}$ mit $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$ eine Basis für S^kV. Insbesondere ist $\dim(S^kV) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$.

BEWEIS. (1) Man betrachtet die Tensorpotenz $\otimes^kV$ und den Teilraum $\mathcal{A} \subset \otimes^kV$, der von allen Elementen der Form

$$\xi \otimes v_1 \otimes v_2 \otimes \eta - \xi \otimes v_2 \otimes v_1 \otimes \eta$$

mit $v_1, v_2 \in V$, ξ ein Tensorprodukt von i Elementen von V mit $1 \leq i \leq k-2$ und η ein Tensorprodukt von $k-2-i$ Elementen von V. Dann definiert man $S^kV := \otimes^kV/\mathcal{A}$ und $v_1 \vee \dots \vee v_k := \pi(v_1 \otimes \dots \otimes v_k)$, wobei $\pi : \otimes^kV \to \otimes^kV/\mathcal{A}$ die kanonische Surjektion bezeichnet.

Die Abbildung $(v_1, \dots, v_k) \mapsto v_1 \vee \dots \vee v_k$ ist offensichtlich k–linear. Außerdem gilt nach Konstruktion, dass sich am Wert von $v_1 \vee \dots \vee v_k$ nichts ändert, wenn man zwei benachbarte Vektoren vertauscht, weil dann die Differenz der Tensorprodukte in $\mathcal{A} = \operatorname{Ker}(\pi)$ liegt. Nachdem man aber jede Vertauschung durch iterierte Vertauschung benachbarter Elemente erzeugen kann, ist die Abbildung symmetrisch.

Sei nun $\varphi : V^k \to Z$ eine symmetrische, k–lineare Abbildung. Dann gibt es nach der universellen Eigenschaft der Tensorpotenz eine eindeutige lineare Abbildung $\hat{\varphi} : \otimes^kV \to Z$, sodass $\hat{\varphi}(v_1 \otimes \dots \otimes v_k) = \varphi(v_1, \dots, v_k)$ gilt. Da φ symmetrisch ist, folgt sofort, dass jedes der erzeugenden Elemente von $\mathcal{A}$ von oben im Kern von $\hat{\varphi}$ liegt. Damit gilt aber $\mathcal{A} \subset \operatorname{Ker}(\hat{\varphi})$ und man findet eine eindeutige lineare Abbildung $\tilde{\varphi} : S^kV \to Z$, sodass $\tilde{\varphi} \circ \pi = \hat{\varphi}$ gilt. Das liefert genau die gewünschte Eigenschaft für $\tilde{\varphi}$.

(2) Die Abbildung $V^k \to S^kW$, die durch $(v_1, \dots, v_k) \mapsto f(v_1) \vee \dots \vee f(v_k)$ definiert ist, ist offensichtlich k–linear und symmetrisch, induziert also nach Teil (1) eine eindeutige lineare Abbildung $S^kf : S^kV \to S^kW$. Die weiteren Behauptungen folgen sofort aus dieser Charakterisierung.

(3) Nach Teil (4) von Lemma 11.4 bilden die Elemente der Form $v_{i_1} \otimes \dots \otimes v_{i_k}$ für beliebige $i_1, \dots, i_k \in \{1, \dots, n\}$ ein Erzeugendensystem für $\otimes^kV$. Damit bilden ihre Bilder unter π ein Erzeugendensystem für S^kV. Aus der Symmetrie des $\vee$–Produkts folgt, dass man die Faktoren in der angegebenen Reihenfolge anordnen kann. Damit bilden die angegebenen Elemente ein Erzeugendensystem.

Für ein Tupel $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$ definieren wir nun eine lineare Abbildung $f : \otimes^k V \to \mathbb{K}$, indem wir das Element $v_{i_1} \otimes \dots \otimes v_{i_k}$ und alle Elemente, die daraus durch eine Permutation der Vektoren hervorgehen, auf Eins abbilden und alle übrigen Basiselemente auf 0. Dann folgt sofort, dass $\mathcal{A} \subset \text{Ker}(f)$ gilt, also erhalten wir eine lineare Abbildung $\lambda_{i_1 \dots i_k} := \underline{f} : S^k V \to \mathbb{K}$. Dann gilt nach Konstruktion offensichtlich $\lambda_{i_1 \dots i_k}(v_{i_1} \vee \dots \vee v_{i_k}) = 1$, aber für jedes andere Tupel mit $1 \leq j_1 \leq j_2 \leq \dots \leq j_k \leq n$ muss $\lambda_{i_1 \dots i_k}(v_{j_1} \vee \dots \vee v_{j_k}) = 0$ sein. Daraus folgt sofort die lineare Unabhängigkeit der angegebenen Elemente.

Um die Formel für die Dimension zu beweisen, verwenden man einen einfachen Trick. Wenn $1 \leq i_1 \leq \dots \leq i_k \leq n$ gilt, dann gilt $1 \leq i_1 < i_2 + 1 < \dots < i_k + (k-1) \leq n + k - 1$. Damit erhalten wir aus dem Tupel $(i_1, \dots, i_k)$ die k–elementige Teilmenge $\{i_1, i_2+1, \dots, i_k+(k-1)\}$ von $\{1, \dots, n+k-1\}$. Haben wir umgekehrt eine k–elementige Teilmenge von $\{1, \dots, n+k-1\}$ gegeben, dann können sie der Reihe nach ordnen und erhalten $1 \leq \ell_1 < \ell_2 < \dots < \ell_k \leq n+k-1$. Dann folgt aber für jedes r, dass $\ell_r \leq n+r-1$ gilt und somit ist $1 \leq \ell_1 \leq \ell_2 - 1 \leq \ell_3 - 2 \leq \dots \leq \ell_k - k + 1 \leq n$. Damit sehen wir, dass die angegebenen Tupel in bijektiver Korrespondenz mit den k–elementigen Teilmengen einer Menge mit $n+k-1$ Elementen steht, woraus sich die Anzahl der Tupel ergibt. □

11.9. Äußere Potenzen. In Beispiel 8.5 und bei den Determinanten in Kapitel 6 haben wir schon gesehen, dass es neben dem Konzept von symmetrischen multilinearen Abbildungen auch das "umgekehrte" Konzept gibt. Man nennt eine k–lineare Abbildung $\varphi : V^k \to Z$ *alternierend*, wenn $\varphi(v_1, \dots, v_k) = 0$ gilt, sofern $v_i = v_j$ für ein $i \neq j$ gilt. Analog wie in 6.4 und 6.6 verifiziert man, dass daraus die Gleichung $\varphi(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \text{sgn}(\sigma)\varphi(v_1, \dots, v_k)$ für jede Permutation $\sigma \in \mathfrak{S}_k$ mit Signum $\text{sgn}(\sigma)$ folgt. Wie in 11.8 gibt es auch für diese Abbildungen ein universelles Objekt, das überraschenderweise wahrscheinlich sogar wichtiger ist, als die symmetrische Potenz.

Proposition 11.9. *Sei V ein Vektorraum über $\mathbb{K}$ und sei $k \geq 2$. Dann gilt:*

(1) Es gibt einen Vektorraum $\Lambda^k V$ und eine k–lineare, alternierende Abbildung $V^k \to \Lambda^k V$, die man als $(v_1, \dots, v_k) \mapsto v_1 \wedge \dots \wedge v_k$ schreibt, sodass folgende universelle Eigenschaft erfüllt ist: Für einen $\mathbb{K}$–Vektorraum Z und eine k–lineare, alternierende Abbildung $\varphi : V^k \to Z$ gibt es eine eindeutige lineare Abbildung $\tilde{\varphi} : \Lambda^k V \to Z$, sodass $\tilde{\varphi}(v_1 \wedge \dots \wedge v_k) = \varphi(v_1, \dots, v_k)$ gilt.

(2) Ist W noch ein $\mathbb{K}$–Vektorraum und $f : V \to W$ eine lineare Abbildung, dann erhalten wir eine induzierte lineare Abbildung $\Lambda^k f : \Lambda^k V \to \Lambda^k W$, die durch

$$\Lambda^k f(v_1 \wedge \dots \wedge v_k) = f(v_1) \wedge \dots \wedge f(v_k)$$

charakterisiert ist. Es gilt $\Lambda^k(g \circ f) = \Lambda^k g \circ \Lambda^k f$ und $\Lambda^k \text{id}_V = \text{id}_{\Lambda^k V}$.

(3) Ist V endlichdimensional und $\{v_1, \dots, v_n\}$ eine Basis für V, dann bilden die Elemente $v_{i_1} \wedge \dots \wedge v_{i_k}$ mit $1 \leq i_1 < i_2 < \dots < i_k \leq n$ eine Basis für $\Lambda^k V$. Insbesondere ist $\dim(\Lambda^k V) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$ für $k \leq n$ und $\Lambda^k V = \{0\}$ für $k \geq n$.

Beweis. Man betrachtet den Teilraum $\mathcal{A} \subset \otimes^k(V)$, der von allen Tensorprodukten von Elementen von V aufgespannt wird, in denen ein Vektor $v \in V$ mindestens zwei mal vorkommt. Dann definiert man $\Lambda^k V := \otimes^k V / \mathcal{A}$ und $v_1 \wedge \dots \wedge v_k := \pi(v_1 \otimes \dots \otimes v_k)$, wobei $\pi : \otimes^k V \to \otimes^k V / \mathcal{A}$ die kanonische Surjektion bezeichnet.

Die Abbildung $(v_1, \dots, v_k) \mapsto v_1 \wedge \dots \wedge v_k$ ist nach Konstruktion k–linear und alternierend. Die universelle Eigenschaft und damit Teil (1), sowie Teil (2) folgen nun völlig analog wie im symmetrischen Fall.

(3) Wie im symmetrischen Fall folgt sofort, dass die Elemente $v_{i_1} \wedge \dots \wedge v_{i_k}$ mit $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ein Erzeugendensystem für $\Lambda^k V$ bilden. Für ein Tupel

$1 \leq i_1 < \cdots < i_k \leq n$ definieren wir eine lineare Abbildung $f : \otimes^k V \to \mathbb{K}$ indem wir für jede Permutation $\sigma \in \mathfrak{S}_k$ das Element $v_{i_{\sigma(1)}} \otimes \cdots \otimes v_{i_{\sigma(k)}}$ auf $\operatorname{sgn}(\sigma)$ und alle anderen Basiselemente auf Null abbilden. Analog wie im symmetrischen Fall liefert das ein lineares Funktional auf $\Lambda^k V$, das $v_{i_1} \wedge \cdots \wedge v_{i_k}$ auf Eins und alle anderen Elemente in unserem Erzeugendensystem auf Null abbildet. Daraus folgt wieder, dass die angegebenen Elemente eine Basis bilden. Die Behauptung über die Dimension folgt sofort, weil die Basiselemente durch die k–elementigen Teilmengen der Menge $\{1, \dots, n\}$ indiziert werden. □

Im Gegensatz zu den symmetrischen Potenzen, wachsen die Dimensionen der äußern Potenzen eines endlichdimensionalen Raumes nicht mit dem Grad. Ist $\dim(V) = n$, dann ist $\Lambda^k V = \{0\}$ für $k > n$ und $\dim(\Lambda^n V) = 1$, also $\Lambda^n V \cong \mathbb{K}$. Die letzte Aussage entspricht genau der Aussage aus Satz 6.7, dass es bis auf Vielfache genau eine n–lineare alternierende Funktion $V^n \to \mathbb{K}$ gibt. Solche Funktionen entsprechen ja genau linearen Abbildungen $\Lambda^n V \to \mathbb{K}$. Analog zum Beweis von Satz 6.7 kann man auch sofort sehen, dass für $k > n$ eine k–lineare alternierende Abbildung $V^k \to \mathbb{K}$ automatisch Null sein muss. Man setzt einfach beliebige Vektoren ein und entwickelt sie in eine Linearkombination von Basiselementen. Wegen der k–Linearität kann man das Ergebnis als eine Summe von Termen schreiben, in denen immer nur Basisvektoren in die Abbildung eingesetzt werden und dabei treten immer mindestens zwei gleiche Vektoren auf.

Auch das Resultat aus Satz 6.8 können wir nun schön interpretieren. Sei V ein Vektorraum der Dimension n und $F : V^n \to \mathbb{K}$ eine n–lineare, alternierende Funktion $\neq 0$. Dann ist $\tilde{F} : \Lambda^n V \to \mathbb{K}$ ein linearer Isomorphismus. Ist nun $f : V \to V$ eine lineare Abbildung, dann erhalten wir die induzierte lineare Abbildung $\Lambda^n f : \Lambda^n V \to \Lambda^n V$. Wegen $\dim(V) = 1$ ist diese Abbildung einfach durch Multiplikation mit einem Element $a \in \mathbb{K}$ gegeben. Wir können diese Zahl bestimmen, indem wir Elemente $v_1 \dots, v_n \in V$ wählen, sodass $v_1 \wedge \cdots \wedge v_n \neq 0$ gilt. (Das ist genau dann der Fall, wenn die Elemente eine Basis bilden.) Dann ist $\Lambda^n f(v_1 \wedge \cdots \wedge v_n) = f(v_1) \wedge \cdots \wedge f(v_n) = a(v_1 \wedge \cdots \wedge v_n)$, und somit $F(f(v_1), \dots, f(v_n)) = aF(v_1, \dots, v_n)$. Aus Satz 6.8 folgt dann $a = \det(f)$, also erhalten wir so eine basisunabhängige Beschreibung der Determinante einer linearen Abbildung.

Interessanterweise kann man eine ähnliche Beschreibung für alle Koeffizienten des charakteristischen Polynoms, die sogenannten *charakteristischen Koeffizienten*, geben.

Satz 11.9. *Sei V ein $\mathbb{K}$–Vektorraum der Dimension $n > 0$ und sei $f : V \to V$ eine lineare Abbildung mit charakteristischem Polynom $p_f := \sum_{k=0}^n a_k x^k$. Dann gilt $a_k = (-1)^k \operatorname{tr}(\Lambda^{n-k} f)$ für alle $k < n$.*

Beweisskizze. Indem man zu einer Matrixdarstellung A von f übergeht und dann A als Matrix über dem algebraischen Abschluss von $\mathbb{K}$ betrachtet, darf man ohne Beschränkung der Allgemeinheit annehmen, dass p_f in ein Produkt von Polynomen ersten Grades zerfällt. Also finden wir n (nicht notwendigerweise verschiedene) Eigenwerte $\lambda_1, \dots, \lambda_n$ von f und $p_f = (\lambda_1 - x) \cdots (\lambda_n - x)$. Dann zeigt man leicht induktiv, dass $a_k = (-1)^k \sum_{1 \leq i_1 < i_2 < \cdots < i_{n-k} \leq n} (\lambda_{i_1} \cdots \lambda_{i_{n-k}})$ gilt.

Nach Bemerkung 10.1 finden wir eine geordnete Basis $\mathcal{B} := \{v_1, \dots, v_n\}$ für V, sodass $[f]_{\mathcal{B}}$ eine obere Dreiecksmatrix ist. Dann stehen auf der Hauptdiagonale die Eigenwerte λ_i, also können diese so ordnen, dass $f(v_i)$ immer als $\lambda_i v_i$ plus eine Linearkombination der v_j mit $j < i$ geschrieben werden kann.

Nach Proposition 11.9 erhalten wir eine induzierte Basis für jeden der Räume $\Lambda^k V$. Um mit diesen Basen handlich arbeiten zu können, betrachten wir Multiindizes $\mathbf{i} := (i_1, \dots, i_k)$ mit $1 \leq i_1 < \cdots < i_k \leq n$, und schreiben die Elemente der induzierten Basis als $v_{\mathbf{i}} := v_{i_1} \wedge \cdots \wedge v_{i_k}$. Wir ordnen diese Basiselemente dadurch an, dass wir $\mathbf{i} < \mathbf{j}$ genau dann setzen, wenn für das erste ℓ, sodass $i_\ell \neq j_\ell$ gilt, $i_\ell < j_\ell$ ist. Das definiert eine Totalordnung und wir benutzen diese, um die Reihenfolge der Elemente in der Basis $\{v_{\mathbf{i}}\}$ für $\Lambda^k V$ festzulegen.

Betrachten wir nun $\Lambda^k f(v_{\mathbf{i}}) = f(v_{i_1}) \wedge \cdots \wedge f(v_{i_k})$. Wir wissen, dass man, für jedes i, das Bild $f(v_i)$ als Summe von $\lambda_i v_i$ und einer Linearkombination von v_ℓ mit $\ell < i$ schreiben kann. Setzt man das ein, dann zerfällt $\Lambda^k f(v_{\mathbf{i}})$ in eine Linearkombination der Basisvektoren $v_{\mathbf{j}}$. Setzt man für $f(v_{i_1})$ ein, dann erhält man $\lambda_{i_1} v_{i_1}$ plus eine Linearkombination von v_ℓ mit $\ell < i_i$. Nun gilt aber für jeden Multiindex $\mathbf{j}$, der einen Eintrag $\ell < i_1$ enthält, automatisch $\mathbf{j} < \mathbf{i}$. Damit sehen wir, dass man $\Lambda^k f(v_{\mathbf{i}})$ als $\lambda_{i_1} v_{i_1} \wedge f(v_{i_2}) \wedge \cdots \wedge f(v_{i_k})$ plus eine Linearkombination von $v_{\mathbf{j}}$ mit $\mathbf{j} < \mathbf{i}$ schreiben kann. Analog erfüllt jeder Multiindex $\mathbf{j}$, der i_1 und einen Index $\ell < i_2$ enthält, automatisch $\mathbf{j} < \mathbf{i}$. Damit ist aber $\Lambda^k f(v_{\mathbf{i}})$ gleich $\lambda_{i_1}\lambda_{i_2} v_{i_1} \wedge v_{i_2} \wedge f(v_{i_3}) \wedge \cdots \wedge f(v_{i_k})$ plus eine Linearkombination von $v_{\mathbf{j}}$ mit $\mathbf{j} < \mathbf{i}$. Setzt man das fort, dann sieht man, dass $\Lambda^k f(v_{\mathbf{i}})$ eine Summe von $\lambda_{i_1} \dots \lambda_{i_k} v_{\mathbf{i}}$ und einer Linearkombination von $v_{\mathbf{j}}$ mit $\mathbf{j} < \mathbf{i}$ ist.

Das sagt aber gerade, dass die Matrixdarstellung von $\Lambda^k f$ bezüglich der Basis $\{v_{\mathbf{i}}\}$ eine obere Dreiecksmatrix ist, wobei die Hauptdiagonaleintragungen gerade $\prod_{\ell=1}^k \lambda_{i_\ell}$ sind. Somit ist $\operatorname{tr}(\Lambda^k f)$ gerade die Summe aller dieser Ausdrücke und die Behauptung folgt. □

Tensoralgebra, äußere Algebra und Cliffordalgebren

11.10. Algebren und die Tensoralgebra. Eine *Algebra* über einem Körper $\mathbb{K}$ ist ein $\mathbb{K}$–Vektorraum A zusammen mit einer bilinearen Abbildung $A \times A \to A$, die man als *Multiplikation* bezeichnet. Wenn diese Multiplikation assoziativ ist, spricht man von einer *assoziativen Algebra*, ist sie zusätzlich kommutativ, von einer *kommutativen Algebra*. Die Bilinearität stellt sicher, dass die Multiplikation auf beiden Seiten distributiv bezüglich der Addition ist, weil ja $(a+b)c = ac+bc$ und $a(b+c) = ab+ac$ gelten müssen. Zusätzlich wird noch eine Verträglichkeit mit der Skalarmultiplikation verlangt, in dem Sinne, dass $(ra)b = a(rb) = r(ab)$ für $r \in \mathbb{K}$ und $a, b \in A$ gelten muss. Ein *Homomorphismus* von $\mathbb{K}$–Algebren ist eine lineare Abbildung, die zusätzlich im offensichtlichen Sinn mit den Multiplikationen verträglich ist.

Wir haben schon einige Beispiele von Algebren kennen gelernt. Für eine $\mathbb{K}$–Vektorraum V ist der Raum $L(V, V)$ mit der Komposition $\circ$ als Multiplikation eine assoziative Algebra, die aber für $\dim(V) > 1$ nicht kommutativ ist. Analog mach die Matrizenmultiplikation den Raum $M_n(\mathbb{K})$ der $n \times n$–Matrizen mit Eintragungen aus $\mathbb{K}$ zu einer assoziativen aber nicht kommutative Algebra. Ist V endlichdimensional, $\dim(V) = n$ und $\mathcal{B}$ eine Basis für V, dann definiert $f \mapsto [f]_{\mathcal{B}}$ einen Homomorphismus (sogar einen Isomorphismus) von $\mathbb{K}$–Algebren.

Der Polynomring $\mathbb{K}[x]$ mit Koeffizienten in $\mathbb{K}$ ist eine unendlichdimensionale kommutative $\mathbb{K}$–Algebra. Für eine Vektorraum V und eine lineare Abbildung $f : V \to V$ definiert die Evaluationsabbildung $p \mapsto p(f)$ einen Algebrahomomorphismus $\mathbb{K}[x] \to L(V, V)$. Die in Satz 10.4 bewiesenen Eigenschaften für diese Abbildung gelten ganz analog allgemein: Das Bild eines Algebrahomomorphismus ist eine Teilalgebra (im offensichtlichen Sinn) und der Kern eines Algebrahomomorphismus ist immer ein Teilraum und ein Ideal.

Mit Hilfe von Tensorprodukten kann man nun zu jedem Vektorraum V eine "universelle" Algebra konstruieren, die sogenannte *Tensoralgebra* $\otimes V$. Dazu betrachten wir die Tensorpotenzen $\otimes^k V$ für $k \in \mathbb{N}$, wobei wir $\otimes^0 V$ als den Grundkörper $\mathbb{K}$ und $\otimes^1 V := V$

definieren. Dann definiert man $\otimes V := \oplus_{k\in\mathbb{N}}(\otimes^k V)$. Ein Element von $\otimes V$ ist also nach Definition eine endliche Summe von Elementen der Räume $\otimes^k V$, wobei es zwischen der Räumen für verschiedene k keine Relationen gibt. (Man kann auch direkte Summen von unendlich vielen Räumen allgemein definieren und zeigen, dass sie eine universelle Eigenschaft besitzen, das würde hier aber zu weit führen.)

Aus der Assoziativität des Tensorprodukts aus Proposition 11.6 folgt induktiv sofort, dass es einen natürlichen Isomorphismus $(\otimes^k V)\otimes(\otimes^\ell V)\to\otimes^{k+\ell}V$ gibt. (Für $k=0$ benutzt man einfach den Isomorphismus $\mathbb{K}\otimes(\otimes^\ell V)\to\otimes^\ell V$ aus Proposition 11.6.) Die entsprechenden bilineare Abbildung $(\otimes^k V)\times(\otimes^\ell V)\to\otimes^{k+\ell}V$ ist durch

$$(v_1\otimes\dots\otimes v_k, w_1\otimes\dots\otimes w_\ell)\mapsto v_1\otimes\dots\otimes v_k\otimes w_1\otimes\dots\otimes w_\ell$$

(bzw. durch $(r, w_1\otimes\dots\otimes w_\ell)=r(w_1\otimes\dots\otimes w_\ell)$) charakterisiert. Wir schreiben dieses Produkt als $(\varphi,\psi)\mapsto\varphi\otimes\psi$ für $\varphi\in\otimes^k V$ und $\psi\in\otimes^\ell V$.

Diese Abbildung können wir nun benutzen, um eine Multiplikation $\otimes V\times\otimes V\to\otimes V$ zu definieren. Nehmen wir Elemente $\varphi_1+\dots+\varphi_r$ und $\psi_1+\dots+\psi_s$ von $\otimes V$, wobei $\varphi_i\in\otimes^{k_i}V$ und $\psi_j\in\otimes^{\ell_j}$ liegt. Dann können wir für jedes i und j das Element $\varphi_i\otimes\psi_j\in\otimes^{k_i+\ell_j}V$ definieren und dann einfach diese rs Elemente aufsummieren. Meist schreibt man diese Elemente wieder als $\varphi,\psi\in\otimes V$ und ihr Produkt als $\varphi\otimes\psi$.

Proposition 11.10. *Sei V ein Vektorraum über $\mathbb{K}$. Dann gilt:*

(1) Die oben definierte Multiplikation macht $\otimes V$ zu einer assoziativen (aber im Allgemeinen nicht kommutativen) Algebra. Das Element $1\in\mathbb{K}=\otimes^0 V$ ist multiplikativ neutral.

(2) Die Tensoralgebra $\otimes V$ hat folgenden universelle Eigenschaft. Ist A eine assoziative Algebra mit Einselement über $\mathbb{K}$ und $f:V\to A$ eine lineare Abbildung, dann gibt es einen eindeutigen Homomorphismus $\tilde f:\otimes V\to A$ von Algebren mit Einselement, der auf $\otimes^1 V=V$ mit f übereinstimmt.

Beweisskizze. (1) Unsere Definition impliziert sofort, dass die Multiplikation bilinear ist, also erhalten wir eine Algebra. Auch dass $1\in\mathbb{K}$ ein neutrales Element für die Multiplikation ist, folgt sofort aus der Definition. Wegen der Bilinearität muss man die Assoziativität nur für Elemente $\varphi\in\otimes^k V$, $\psi\in\otimes^\ell V$ und $\omega\in\otimes^m V$ überprüfen. Nochmals wegen der Bilinearität darf man sogar annehmen, dass alle drei Elemente Tensorprodukte von Elementen von V sind. Für diese ist die Assoziativität aber offensichtlich. Nachdem wir schon wissen, dass im Allgemeinen für linear unabhängige Vektoren $v_1,v_2\in V$ die Elemente $v_1\otimes v_2$ und $v_2\otimes v_1$ in $\otimes^2 V$ verschieden sind, ist die Multiplikation für $\dim(V)>1$ nicht kommutativ.

(2) Haben wir $f:V\to A$ gegeben, dann definieren wir $f_k:\otimes^k V\to A$ wie folgt. Für $k=0$ setzen wir $f_0(r):=r\cdot 1_A$, wobei 1_A das Einselement der Algebra A ist und $\cdot$ die Skalarmultiplikation auf A bezeichnet. Natürlich setzen wir $f_1:=f:V=\otimes^1 V\to A$. Für $k=2$ betrachten wir die Abbildung $V^2\to A$, die durch $(v_1,v_2)\mapsto f(v_1)f(v_2)$ (Multiplikation in A) gegeben ist. Wegen der Bilinearität der Multiplikation in A ist diese Abbildung bilinear, induziert also eine lineare Abbildung $f_2:\otimes^2 V\to A$, die $f_2(v_1\otimes v_2)=f(v_1)f(v_2)$ erfüllt. Da die Algebra A assoziativ ist, erhält man ganz analog für $k>2$, eine lineare Abbildung $f_k:\otimes^k V\to A$, die $f_k(v_1\otimes\dots\otimes v_k)=f(v_1)\cdots f(v_k)$ erfüllt.

Nun definiert man $\tilde f:\otimes V\to A$, indem man

$$\tilde f(\varphi_1+\dots+\varphi_r):=f_{i_1}(\varphi_1)+\dots+f_{i_r}(\varphi_r)$$

für Elemente $\varphi_j\in\otimes^{i_j}V$ definiert. Nach Konstruktion ist das eine lineare Abbildung und man verifiziert leicht direkt, dass $\tilde f$ ein Algebrahomomorphismus ist.

Die Eindeutigkeit von $\tilde{f}$ folgt sofort, weil man $v_1 \otimes \cdots \otimes v_k$ als Produkt in der Tensoralgebra interpretieren kann. Deshalb muss $\tilde{f}$ dieses Element auf $\tilde{f}(v_1) \cdots \tilde{f}(v_k) \in A$ abbilden. Verlangt man, dass $\tilde{f}$ auf $\otimes^1 V = V$ mit f übereinstimmt, dann legt das natürlich $\tilde{f}$ auf $\otimes^k V$ fest und die Eindeutigkeit folgt. □

11.11. Ideale und Quotienten von Algebren. Die Tensoralgebra ist vor allem deswegen ein sehr flexibles Werkzeug, weil man leicht Quotienten bilden kann. Den richtigen Rahmen dafür bilden wieder Ideale, die wir (im Fall eines Polynomringes) schon in Definition 10.5 kennen gelernt haben. Für Algebren modifiziert man den Begriff so, dass man ein *Ideal* in einer $\mathbb{K}$–Algebra A als einen Teilraum $I \subset A$ definiert, sodass für jedes $a \in A$ und $b \in I$, sowohl ab als auch ba in I liegt.

Proposition 11.11. *Sei A eine Algebra über eine Körper $\mathbb{K}$.*

(1) Ist $I \subset A$ ein Ideal, dann induziert die Multiplikation auf A eine Multiplikation auf dem Quotientenraum A/I. Dabei übertragen sich Assoziativität, Kommutativität und Existenz eines Einselments auf den Quotienten.

(2) Ist $\{I_\alpha\}$ eine beliebige Familie von Idealen in A, dann ist der Durchschnitt $\cap_\alpha I_\alpha$ ebenfalls ein Ideal in A.

(3) Ist $B \subset A$ eine beliebige Teilmenge, dann gibt es ein eindeutig bestimmtes Ideal $I_B \subset A$, sodass $B \subset I_B$ gilt und sodass jedes Ideal $I \subset A$, das B enthält auch I_B enthält.

(4) Das Ideal I_B aus (3) stimmt mit dem von $B \cup \{a_1 b a_2 : a_1, a_2 \in A, b \in B\}$ erzeugten Teilraum von A überein.

Beweis. (1) Man definiert die Multiplikation auf A/I durch $(a_1 + I) \cdot (a_2 + I) := a_1 a_2 + I$. Um zu sehen, dass das wohldefiniert ist, genügt es zu beobachten, dass für $b_1, b_2 \in I$ die Gleichung $(a_1+b_1)(a_2+b_2) = a_1a_2+b_1a_2+a_1b_2+b_1b_2$ wegen der Bilinearität der Multiplikation gilt. Da aber I ein Ideal ist, liegen die letzten drei Summanden (und damit auch ihre Summe) in I, und die Wohldefiniertheit folgt. Aus der Definition von Addition und Skalarmultiplikation auf A/I folgt sofort, dass sich die Bilinearität sowie Assoziativität, Kommutativität und Existenz eines neutralen Elements von der ursprünglichen Multiplikation auf die Multiplikation auf A/I übertragen.

(2) folgt direkt aus der Definition eines Ideals und aus Korollar 2.4, das zeigt, dass $\cap_\alpha I_\alpha$ ein Teilraum von A ist. Dann erhält man aber (3) sofort, indem man I_B als den Durchschnitt über all jene Ideale von A definiert, die B enthalten.

(4) Sei V der von $\tilde{B} := B \cup \{a_1 b a_2 : a_1, a_2 \in A, b \in B\}$ erzeugte Teilraum. Dann liegen natürlich für $a \in A$ und $\tilde{b} \in \tilde{B}$ klarerweise auch $a\tilde{b}$ und $\tilde{b}a$ in $\tilde{B}$, also in V. Zusammen mit der Bilinearität der Multiplikation impliziert das sofort, dass $V \subset A$ ein Ideal ist. Natürlich gilt $B \subset V$, also folgt $I_B \subset V$ aus (3).

Umgekehrt folgt aus $B \subset I_B$ und der Tatsache, dass I_B ein Ideal ist, natürlich $\tilde{B} \subset I_B$. Da I_B aber insbesondere ein Teilraum von A ist, ist auch der von $\tilde{B}$ erzeugte Teilraum V in I_B enthalten. □

Analog zum Fall von Teilräumen nennt man das Ideal I_B aus Teil (3) das *von B erzeugte Ideal* in A.

Beispiel 11.11. Wir können die Flexibilität dieses Konzepts schon in einem einfachen Beispiel erkennen. Der Übergang von $\mathbb{R}$ zu $\mathbb{C}$ ist dadurch motiviert, dass in $\mathbb{R}$ das Polynom $x^2 + 1$ keine Nullstelle besitzt. Betrachten wir nun die kommutative Algebra $\mathbb{R}[x]$ und, dann ist das von $p := x^2 + 1$ erzeugte Ideal I gerade $p\mathbb{R}[x]$, also ist der Quotient $R := \mathbb{R}[x]/I$ ein kommutativer Ring mit Einselement. Analog wie im Beweis von Proposition 10.6 folgt sofort, dass $1 + I$ und $x + I$ eine Basis für R bilden, und das 1 ein multiplikativ neutrales Element ist. Nun ist aber $(x+I)^2 = x^2+I = -1+I = -(1+I)$, also gilt $(x+I)^2 = -1$ in R. Daraus schließt man sofort, dass R ein Körper und isomorph zu $\mathbb{C}$ ist.

Tatsächlich kann man leicht zeigen, dass für jeden Körper $\mathbb{K}$ und für jedes *irreduzible* Polynome $p \in \mathbb{K}[x]$ der Quotient $L := \mathbb{K}[x]/p\mathbb{K}[x]$ ein Körper ist. Dazu nimmt man ein Element $0 \neq u \in L$ und betrachte $uL = \{uv : v \in L\}$. Man sieht sofort, dass das ein Ideal in L ist. Betrachtet man die kanonische Surjektion $\pi : \mathbb{K}[x] \to L$, dann folgt sofort, dass $I := \{r \in \mathbb{K}[x] : \pi(r) \in uL\}$ ein Ideal in $\mathbb{K}[x]$ ist. Nach dem Hauptidealsatz (Satz 10.5) ist $I = q\mathbb{K}[x]$ für ein monisches Polynom $q \in \mathbb{K}[x]$. Nach Definition ist $p \in I$, also q ein Teiler von p, aber I eine echte Obermenge von $p\mathbb{K}[x] = \mathrm{Ker}(\pi)$, also $q \neq p$. Wegen der Irreduzibilität von p folgt $q = 1$, also $I = \mathbb{K}[x]$. Damit ist aber $1 \in uL$ also gibt es ein Element $v \in L$, sodass $uv = 1$ gilt, also hat jedes Element $\neq 0$ in L ein multiplikativ Inverses. Man erhält also einen systematischen Weg zur Konstruktion von Körpern aus irreduziblen Polynomen.

11.12. Die äußere Algebra. Wir können die Ideen über das Bilden von Quotienten in einem einfachen Fall auf die Tensoralgebra $\otimes V$ eines Vektorraumes V anwenden. Im Beweis von Proposition 11.9 haben wir für $k \geq 2$ die äußeren Potenz $\Lambda^k V$ als Quotientenraum von $\otimes^k V$ nach einem Teilraum $\mathcal{A}$ definiert, den wir nun mit $\mathcal{A}_k \subset \otimes^k V$ bezeichnen. Nach Definition ist $\mathcal{A}_k$ der Teilraum, der von all jenen Tensorprodukten von Elementen von V aufgespannt wird, in denen ein Element $v \in V$ mindestens zwei Mal vorkommt. Setzten wir nun noch $\mathcal{A}_0 = \{0\} \subset \mathbb{K}$ und $\mathcal{A}_1 = \{0\} \subset V$, dann können wir offensichtlich $I := \oplus_{k\in\mathbb{N}}\mathcal{A}_k$ als Teilraum von $\oplus_{k\in\mathbb{N}} \otimes^k V = \otimes V$ auffassen.

Nun liegt aber nach Definition ein Produkt eines Elements von $\mathcal{A}_k$ mit $v_1 \otimes \cdots \otimes v_\ell \in \otimes^\ell V$ immer in $\mathcal{A}_{k+\ell}$ (egal in welcher Reihenfolge man die beiden Elemente multipliziert). Daraus folgt sofort, dass $I = \oplus_{k\in\mathbb{N}}\mathcal{A}_k$ ein Ideal in der Algebra $\otimes V$ ist. Somit ist $\Lambda V := \otimes V/I$ eine assoziative Algebra mit Einselement die man als die *äußere Algebra* von V bezeichnet. Aus der Konstruktion folgt sofort, dass $\Lambda V = \oplus_{k\in\mathbb{N}}\Lambda^k V$ gilt, wobei $\Lambda^0 V := \mathbb{K}$ und $\Lambda^1 V = V$ gesetzt wird. Die Multiplikation in ΛV bildet $\Lambda^k V \times \Lambda^\ell V$ nach $\Lambda^{k+\ell} V$ ab und wird (analog zum Tensorprodukt) mit $(\varphi, \psi) \mapsto \varphi \wedge \psi$ bezeichnet.

Proposition 11.12. *Sei V ein Vektorraum über $\mathbb{K}$ mit äußerer Algebra ΛV. Dann gilt:*

(1) Die Multiplikation auf ΛV ist assoziativ und graduiert kommutativ, d.h. für $\varphi \in \Lambda^k V$ und $\psi \in \Lambda^\ell V$ gilt $\psi \wedge \varphi = (-1)^{k\ell}\varphi \wedge \psi$, und $1 \in \mathbb{K} = \Lambda^0 V$ ist ein multiplikativ neutrales Element.

(2) Ist V endlichdimensional und $\dim(V) = n$, *dann ist auch ΛV endlichdimensional mit* $\dim(\Lambda V) = 2^n$.

Beweis. (1) Assoziativität und die Tatsache, dass 1 ein neutrales Element ist, folgen sofort aus den entsprechenden Eigenschaften der Tensoralgebra und aus Proposition 11.11, also bleibt nur die graduierte Kommutativität zu zeigen. Da $\otimes^k V$ von Tensorprodukten von Elementen von V aufgespannt wird, spannen die Elemente der Form $v_1 \wedge \cdots \wedge v_k$ mit $v_i \in V$ den Raum $\Lambda^k V$ auf. Damit genügt es aber nach der Bilinearität der Multiplikation, die graduierte Kommutativität für solche Elemente zu zeigen. Nach Definition der Multiplikation ist aber

$$(v_1 \wedge \cdots \wedge v_k) \wedge (w_1 \wedge \cdots \wedge w_\ell) = v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_\ell$$

und wir wissen schon, dass eine Permutation von Elementen in so einem Produkt von Elementen von V gerade Multiplikation mit dem Signum der Permutation bewirkt. Nun können wir aber $w_1 \wedge \cdots \wedge w_\ell \wedge v_1 \wedge \cdots \wedge v_k$ folgendermaßen erzeugen. Wir tauschen erst

w_1 ganz nach rechts, was k Vertauschungen, also eine Multiplikation mit $(-1)^k$ bewirkt. Dann tauschen wir w_2 hinter w_1, was wiederum eine Multiplikation mit $(-1)^k$ bewirkt und so weiter. Insgesamt erhalten wir einen Faktor $(-1)^{k\ell}$ und das Resultat folgt.

(2) Wir wissen bereits aus Proposition 11.9, dass für endlichdimensionales V auch $\Lambda^k V$ endlichdimensional ist und $\Lambda^k V = \{0\}$ für $k > n = \dim(V)$ gilt. Wir haben wir auch schon gezeigt, dass wir eine Basis für $\Lambda^k V$ erhalten, die durch die k–elementigen Teilmengen von $\{1, \dots, n\}$ indiziert ist. Für $\Lambda^1 V = V$ erhalten wir eine Basis mit genau n Elementen (also der Anzahl der einelementigen Teilmengen von $\{1, \dots, n\}$) und für $\Lambda^0 V = \mathbb{K}$ eine Basis mit einem Element. Insgesamt sehen wir, dass wir eine Basis für ΛV erhalten, die durch die Menge aller Teilmengen von $\{1, \dots, n\}$ indiziert ist, also 2^n Elemente hat. □

11.13. Anwendung: Differentialformen. Die Schiefsymmetrie von Abbildungen und die graduierte Kommutativität der Multiplikation in der äußeren Algebra wirkt vielleicht ungewohnt. Trotzdem haben äußere Algebren wichtige Anwendung in der Analysis und in der Differentialgeometrie. Die wesentliche Quelle dieser Anwendungen ist der Zusammenhang der Determinante für $\mathbb{K} = \mathbb{R}$ mit Volumsbegriffen. Wir haben diesen Zusammenhang in Dimension drei in Abschnitt 8.12 kurz besprochen. Dort haben wir gesehen, dass man für 3 Vektoren $u, v, w \in \mathbb{R}^3$, der Absolutbetrag $|\det(u, v, w)|$ genau das Volumen des von u, v und w aufgespannten Parallelepipeds ist. (Für diese Interpretation war das innere Produkt und das Kreuzprodukt nötig. Man kann das Volumen aber auch einfach über den Betrag der Determinante definieren und die Rechnung in 8.12 als Bestätigung für die Sinnhaftigkeit dieses Begriffs interpretieren.) Das Vorzeichen der Determinante hängt mit dem Begriff der *Orientierung* zusammen, weshalb die Determinante oft als *orientiertes Volumen* bezeichnet wird.

Die eigentliche Rechtfertigung der Interpretation der Determinante als Volumen und die Verbindung zur Analysis stellt der Transformationssatz für Mehrfachintegrale dar. Seien U und V offene Teilmengen in $\mathbb{R}^n$ und $\Phi : U \to V$ ein Diffeomorphismus (also eine stetig differenzierbare, bijektive Funktion, sodass auch $\Phi^{-1} : V \to U$ stetig differenzierbar ist). Für geeignete Funktionen $f : V \to \mathbb{R}$ stellt der Transformationssatz dann den Zusammenhang zwischen $\int_V f$ und $\int_U (f \circ \Phi)$ her, wobei der zentrale Faktor der Absolutbetrag der Determinante der Ableitung von Φ ist.

In der Analysis und der Differentialgeometrie möchte man Integrale nicht nur über offene Teilmengen von $\mathbb{R}^n$ bilden, sondern auch über gewisse schönen Teilmengen ("Teilmannigfaltigkeiten") von $\mathbb{R}^n$ (Kurvenintegrale, Oberflächenintegrale, etc.) oder über noch allgemeinere Objekte (Mannigfaltigkeiten). In klassischen Formulierungen werden dabei Funktionen integriert, wobei aber noch Objekte wie "Volumselemente" und "Oberflächenelemente" berücksichtigt werden müssen. Einen alternativen, allgemeinen Zugang zu diesen Fragen bietet der Begriff der *Differentialform*. Hier interpretiert man den Transformationssatz für Mehrfachintegrale so, dass man nicht über Funktionen integriert sondern über Objekte, deren Werte in jedem Punkt alternierende, multilineare Abbildungen sind.

Für $0 \leq k \leq n$ definiert man eine *Differentialform vom Grad k* (oder einfach eine *k–form*) ω auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ als eine Funktion, die jedem Punkt $x \in u$ eine k–lineare, alternierende Abbildung $\omega(x) : (\mathbb{R}^n)^k \to \mathbb{R}$ zuordnet. Diese Objekte kann man dann in natürlicher Weise über k–dimensionale Teilmannigfaltigkeiten von U integrieren. Um mit diesen Objekten operieren zu können, benutzt man üblicherweise die folgende Beschreibung.

PROPOSITION 11.13. *Sei V ein endlichdimensionaler Vektorraum über $\mathbb{R}$.*

(1) Es gibt einen natürlichen Isomorphismus von $\Lambda^k V^*$ auf den Raum L^k aller k–linearen, alternierenden Abbildungen $V^k \to \mathbb{R}$ und damit einen natürlichen Isomorphismus $\Lambda^k V^* \cong (\Lambda^k V)^*$.

(2) Über den Isomorphismus aus (1) induziert das Produkt auf der äußeren Algebra ΛV^* eine Operation auf den alternierenden multilinearen Abbildung auf V, die explizit wie folgt gegeben ist. Für $\varphi : V^k \to \mathbb{R}$ und $\psi : V^\ell \to \mathbb{R}$ ist $\varphi \wedge \psi : V^{k+\ell} \to \mathbb{R}$ gegeben durch

$$(\varphi \wedge \psi)(v_1, \dots, v_{k+\ell}) = \tfrac{1}{k!\ell!} \textstyle\sum_{\sigma \in \mathfrak{S}_{k+\ell}} \operatorname{sgn}(\sigma)\varphi(v_{\sigma(1)}, \dots, v_{\sigma(k)})\psi(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}).$$

Beweis. (1) Der Raum L^k aller k–linearen alternierende Abbildungen $V^k \to \mathbb{R}$ ist ein Vektorraum unter den punktweisen Operationen. Nach der universellen Eigenschaft von $\Lambda^k V$ induziert jede Abbildung $\varphi \in L^k$ eine eindeutige lineare Abbildung $\tilde{\varphi} : \Lambda^k V \to \mathbb{R}$, also ein Element in $(\Lambda^k V)^*$. Aus der Eindeutigkeit folgt leicht, dass $\widetilde{\varphi + \lambda\psi} = \tilde{\varphi} + \lambda\tilde{\psi}$ gilt, wir also eine lineare Abbildung $L^k \to (\Lambda^k V)^*$ definiert haben. Umgekehrt ist für eine lineare Abbildung $f : \Lambda^k V \to \mathbb{R}$ natürlich die Abbildung $(v_1, \dots, v_k) \mapsto f(v_1 \wedge \dots \wedge v_k)$ in L^k und man verifiziert sofort, dass das eine Inverse definiert, also gilt $L^k \cong (\Lambda^k V)^*$.

Für lineare Funktionale $\varphi_1, \dots, \varphi_k \in V^*$ betrachten wir die Abbildung $V^k \to \mathbb{R}$, die $v_1, \dots, v_k \in V$ auf die Determinante der Matrix $A = (a_{ij})$ abbildet, die durch $a_{ij} = \varphi_i(v_j)$ gegeben ist. Offensichtlich ist diese Abbildung k–linear und falls zwei der v's übereinstimmen hat A zwei gleiche Spalten, also erhalten wir eine Element von L^k. Betrachten wir nun diese Zuordnung als Abbildung $(V^*)^k \to L^k$, dann folgt aus der Linearität der Determinante in jeder Zeile leicht, dass die entsprechende Abbildung k–linear ist. Ist $\varphi_i = \varphi_j$ für $i \neq j$, dann hat die Matrix A immer zwei gleiche Zeilen, also Determinante Null. Nach der universellen Eigenschaft aus Proposition 11.9 erhalten wir eine lineare Abbildung $\Lambda^k V^* \to L^k$. Betrachten wir als vereinfachte Notation einfach $\varphi_1 \wedge \dots \wedge \varphi_k$ als Element von L^k, dann ist diese Abbildung charakterisiert durch

$$(\varphi_1 \wedge \dots \wedge \varphi_k)(v_1, \dots, v_k) = \textstyle\sum_{\sigma \in \mathfrak{S}_k} \operatorname{sgn}(\sigma)\varphi_1(v_{\sigma(1)}) \cdots \varphi_k(v_{\sigma(k)}),$$

wobei wir die Leibniz Formel für die Determinante aus Satz 6.7 benutzt haben. Aus Proposition 11.9 wissen wir auch, dass die Räume $\Lambda^k V^*$ und $(\Lambda^k V)^*$ (und damit L^k) die gleiche Dimension haben, also genügt es zu zeigen, dass unsere Abbildung injektiv ist, um den Beweis von (1) zu vervollständigen. Dafür wählen wir eine Basis $\{v_1, \dots, v_n\}$ für V und betrachten die duale Basis $\{\varphi_1, \dots, \varphi_n\}$ für V^*. Dann bilden nach Proposition 11.9 die Elemente $\varphi_{i_1} \wedge \dots \wedge \varphi_{i_k}$ für $1 \leq i_1 < i_2 < \dots < i_k \leq n$ eine Basis für $\Lambda^k V^*$. Setzt man aber k verschiedene Basiselmente v_ℓ in die entsprechenden Abbildungen ein, dann liefert genau eine der Abbildung 1 und alle anderen Null. Daraus folgt aber sofort, dass die Menge $\{\varphi_{i_1} \wedge \dots \wedge \varphi_{i_k} : 1 \leq i_1 < i_2 < \dots < i_k \leq n\}$ linear unabhängig in L^k ist, und damit die Injektivität.

(2) Das Produkt auf $\Lambda^k V^*$ ist natürlich charakterisiert durch

$$(\varphi_1 \wedge \dots \wedge \varphi_k) \wedge (\psi_1 \wedge \dots \wedge \psi_\ell) = \varphi_1 \wedge \dots \wedge \psi_\ell.$$

Wegen der Bilinearität genügt es auch, die behauptete Formel in diesem Fall, also für $\varphi = \varphi_1 \wedge \dots \wedge \varphi_k$ und $\psi = \psi_1 \wedge \dots \wedge \psi_\ell$ zu beweisen. Aus dem Beweis von (1) wissen wir, das $\varphi_1 \wedge \dots \wedge \psi_\ell$ die Vektoren $v_1, \dots, v_{k+\ell}$ auf

$$\textstyle\sum_{\sigma \in \mathfrak{S}_{k+\ell}} \operatorname{sgn}(\sigma)\varphi_1(v_{\sigma(1)}) \cdots \varphi_k(v_{\sigma(k)})\psi_1(v_{\sigma(k+1)}) \cdots \psi_\ell(v_{\sigma(k+\ell)})$$

abbildet. Um aber für $\varphi = \varphi_1 \wedge \dots \wedge \varphi_k$, den Wert $\varphi(v_{\sigma(1)}, \dots, v_{\sigma(k)})$ zu berechnen, müssen wir über alle Permutationen der Einträge summieren. Die $k!$ entsprechenden Terme treten natürlich auch in der Summe für andere Permutationen σ auf. Analoges gilt für $\psi(v_{\sigma(k+1)},\dots,v_{\sigma(k+\ell)})$ wobei man hier $\ell!$ Summanden erhält. Insgesamt folgt, dass man die obige Summe als

$$\tfrac{1}{k!\ell!}\sum_{\sigma\in\mathfrak{S}_{k+\ell}}\operatorname{sgn}(\sigma)\varphi(v_{\sigma(1)},\dots,v_{\sigma(k)})\psi(v_{\sigma(k+1)},\dots,v_{\sigma(k+\ell)})$$

schreiben kann und damit die Behauptung. □

Damit ist eine Differentialformen auf U einfach eine Funktion auf U mit Werten im endlichdimensionalen Vektorraum $\Lambda^k\mathbb{R}^{n*}$, also kann man problemlos von stetigen oder differenzierbaren Differentialformen sprechen. Außerdem kann man natürlich das punktweise Produkt von zwei Differentialformen bilden. Das macht die Differentialformen zu einer assoziativen und graduiert kommutativen Algebra, die eine der Grundstrukturen der Differentialgeometrie darstellt.

11.14. Die symmetrische Algebra. Das kommutative Gegenstück zur äußeren Algebra ist die symmetrische Algebra eines Vektorraumes, die wir nur relativ kurz besprechen werden. Hier sind die Konstruktionen eher einfacher als im Fall der äußeren Algebra, das Resultat ist aber immer unendlichdimensional.

In 11.8 haben wir für einen $\mathbb{K}$–Vektorraum V die symmetrische Potenz S^kV als Quotient von $\otimes^kV$ nach einem Teilraum $\mathcal{A}_k$ definiert. Wiederum kann man die $\mathcal{A}_k$ zu einem Teilraum $J:=\oplus_{k\in\mathbb{N}}\mathcal{A}_k\subset\otimes V$ zusammenfassen, der nach Definition ein Ideal in der Tensoralgebra bildet. Somit ist der Quotient $SV:=\otimes V/J$ eine Algebra, die *symmetrische Algebra* von V. Definiert man wie zuvor $S^0V:=\mathbb{K}$ und $S^1V:=V$, dann ist $SV=\oplus_{k\in\mathbb{N}}S^kV$. Analog zum Fall der äußeren Algebra bezeichnen wir die Multiplikation in SV mit $(\varphi,\psi)\mapsto\varphi\vee\psi$.

Proposition 11.14. *Sei V ein Vektorraum über $\mathbb{K}$ mit symmetrischer Algebra SV.*

(1) Die Algebra SV ist assoziativ und kommutativ und $1\in\mathbb{K}=S^0V$ ist ein multiplikativ neutrales Element.

(2) Die symmetrische Algebra über V besitzt folgende universelle Eigenschaft. Ist A eine beliebige assoziative und kommutative Algebra über $\mathbb{K}$ und $f:V\to A$ eine lineare Abbildung, dann gibt es einen eindeutigen Homomorphismus $\tilde f:SV\to A$ von $\mathbb{K}$–Algebren, der auf dem Teilraum $S^1V=V$ mit f übereinstimmt.

Beweis. (1) Außer der Kommutativität folgen alle Eigenschaften direkt aus denen der Tensoralgebra. Um die Kommutativität zu beweisen genügt es zu zeigen, dass für $\varphi\in\otimes^kV$ und $\psi\in\otimes^\ell V$ immer $\varphi\otimes\psi-\psi\otimes\varphi\in\mathcal{A}_{k+\ell}$ gilt. Das beweist man mit Induktion nach $k+\ell$, wobei der Induktionsanfang $k+\ell=2$ nach Definition von $\mathcal{A}_2$ offensichtlich ist. Für dreifache Produkte rechnet man einfach

$$v_1\otimes v_2\otimes w-w\otimes v_1\otimes v_2=v_1\otimes v_2\otimes w-v_1\otimes w\otimes v_2+v_1\otimes w\otimes v_2-w\otimes v_1\otimes v_2$$

und sowohl die erste als auch die zweite Differenz liegt in nach Definition in $\mathcal{A}_3$. Das allgemeine Resultat beweist man analog mit Induktion.

(2) Nach der universellen Eigenschaft der Tensoralgebra liefert $f:V\to A$ einen Homomorphismus $\otimes V\to A$ den wir mit $\hat f$ bezeichnen. Nach Definition kann man jedes Element von J als Summe von Elementen der Form $\xi\otimes v_1\otimes v_2\otimes\eta-\xi\otimes v_2\otimes v_1\otimes\eta$. Da $\hat f$ ein Algebrahomomorphismus ist, gilt

$$\hat f(\xi\otimes v_1\otimes v_2\otimes\eta)=\hat f(\xi)f(v_1)f(v_2)\hat f(\eta),$$

wobei auf der rechten Seite in A multipliziert wird. Da aber die Multiplikation in A kommutativ ist, ändert sich das Ergebnis nicht, wenn v_1 und v_2 vertauscht werden. Damit gilt $J\subset\operatorname{Ker}(\hat f)$. Damit faktorisiert $\hat f$ zu einer Abbildung $\tilde f:\otimes V/J=SV\to A$, die nach Konstruktion eine Algebrahomomorphismus ist und (wie $\hat{f}$) auf $V = \otimes^1 V = S^1 V$ mit f übereinstimmt. Die Eindeutigkeit folgt genau wie in Proposition 11.10. $\square$

Bemerkung 11.14. (1) Es mag überraschend erscheinen, dass wir zwar für die Tensoralgebra und für die symmetrische Algebra eine universelle Eigenschaft bewiesen haben, nicht aber für die äußere Algebra. Die äußere Algebra besitzt eine universelle Eigenschaft, um diese zu formulieren benötigt man allerdings den Begriff von graduierten Algebren, der hier zu weit führen würde.

(2) Wie im Fall der äußeren Algebra kann man für einen $\mathbb{K}$–Vektorraum V mit Dualraum V^* die Symmetrische Algebra SV^* mit symmetrischen Multilinearformen in Verbindung bringen. Dazu ordnet man $\varphi_1, \dots, \varphi_k \in V^*$ die Abbildung $V^k \to \mathbb{K}$ zu, die $(v_1, \dots, v_k)$ auf

$$\sum_{\sigma\in\mathfrak{S}_k} \varphi_1(v_{\sigma(1)}) \cdots \varphi_k(v_{\sigma(k)})$$

abbildet. Diese ist k–linear und symmetrisch und liefert daher nach der universellen Eigenschaft der symmetrischen Potenz eine lineare Abbildung $S^k V \to \mathbb{K}$, also ein Element von $(S^k V)^*$.

Das liefert eine Abbildung $(V^*)^k \to (S^k V)^*$ von der man leicht nachrechnet, dass sie k–linear und symmetrisch ist. Nochmals nach der universellen Eigenschaft erhält man eine lineare Abbildung $S^k V^* \to (S^k V)^*$. Für endlichdimensionales V und $\mathbb{K} = \mathbb{R}$ liefert das wieder einen Isomorphismus, den man benutzen kann, um das Produkt von SV^* auf den Raum der symmetrischen Multilinearformen zu übertragen.

(3) Betrachten wir die Konstruktion auf (2) für den $\mathbb{R}$–Vektorraum $\mathbb{R}^n$ und bezeichnen wir mit $x_1, \dots, x_n \in \mathbb{R}^{n*}$ die Koordinatenabbildungen (also die Elemente der dualen Basis zur Standardbasis). Schreibt man x_i^j für ein $\vee$–Produkt von j Kopien von x_i, dann wissen wir aus Proposition 11.8, dass die Elemente der Form $x_1^{j_1} \vee \dots \vee x_n^{j_n}$ mit $0 \leq j_i$ und $j_1 + \dots + j_n = k$ eine Basis für $S^k V$. Schreibt man dieses Element einfach als $x_1^{j_1} \dots x_n^{j_n}$ besteht SV^* gerade aus endlichen Summen der Form $\sum a_{j_1 \cdots j_n} x_1^{j_1} \dots x_n^{j_n}$. Somit sieht man, dass man in diesem Fall einfach eine Polynomalgebra in n Variablen erhält. Man kann also SV^* als Algebra der Polynome auf V betrachten.

11.15. Clifford–Algebren. Zum Abschluss des Kapitels besprechen wir noch eine Klasse von endlichdimensionalen Algebren, die eine wichtige Rolle in einigen Teilgebieten der Mathematik und der theoretischen Physik spielen. Man benutzt diese Algebren um tiefer liegende Eigenschaften der orthogonalen Gruppen zu studieren. Außerdem liefern sie die Grundlage für Spinoren, die sowohl in der Mathematik als auch in der Physik wichtig sind.

Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, sei V ein endlichdimensionaler Vektorraum über $\mathbb{K}$ und $b : V \times V \to \mathbb{K}$ eine symmetrische Bilinearform (auch für $\mathbb{C}$, wir betrachten hier keine Sesquilinearformen). Wir wollen eine Algebra $\mathrm{Cl}(V, b)$ konstruieren, die in natürlicher Weise V enthält und die eine universelle Eigenschaft für gewisse lineare Abbildungen von V in $\mathbb{K}$–Algebren mit Einselement besitzt. Und zwar interessieren wir uns für lineare Abbildungen $f : V \to A$, sodass $f(v)f(v) = b(v, v)1$ gilt, wobei auf der linken Seite die Multiplikation in A und auf der rechten Seite das Einselement von A benutzt wird. Diese Relation sieht an dieser Stelle wahrscheinlich ziemlich aus der Luft gegriffen aus. Am Anfang der Theorie standen Versuche, die komplexe Einheit i (die ja $i^2 = -1$ erfüllt) zu verallgemeinern, ähnlich wie wir es in 9.15 für die Hamilton'schen Quaternionen besprochen haben. Die Verbindungen zur Physik entstehen aus den Versuchen, den Laplace Operator als Quadrat eines Operators erster Ordnung zu schreiben. Das gelang dem Physiker P.A.M. Dirac in den 1930er Jahren, und die entsprechenden Operatoren heißen heute Dirac–Operatoren.

Die oben angesprochene Beziehung zu komplexen Einheiten können wir auch benutzen um an Beispielen zu sehen, dass diese Relationen nicht so seltsam sind, wie sie vielleicht auf den ersten Blick aussehen. Betrachten wir $V = \mathbb{R}$ mit dem Negativen des Standard–inneren Produkts, also $b(t,t) = -t^2$. Dann hat die Abbildung $f : V \to \mathbb{C}$, die durch $f(t) = ti$ gegeben ist, gerade die gewünschte Eigenschaft. Etwas komplizierter können wir $V = \mathbb{R}^2$ mit dem negativen des Standard–inneren Produkts betrachten und $f : \mathbb{R}^2 \to \mathbb{H}$ durch $f\binom{s}{t} := si + tj$ definieren. Nun ist $(si + tj)(si + tj) = s^2i^2 + stij + tsji + t^2j^2$ und wie wir aus 9.15 wissen, gelten in $\mathbb{H}$ die Relationen $i^2 = j^2 = -1$ und $ij = -ji = k$. Damit erhalten wir aber $(si + tj)(si + tj) = -(s^2 + t^2)1$ und damit genau die gewünschte Relation.

Proposition 11.15. *Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$, V ein endlichdimensionaler $\mathbb{K}$–Vektorraum und b eine symmetrische Bilinearform auf V. Dann gibt es eine assoziative $\mathbb{K}$–Algebra $\mathrm{Cl}(V,b)$ mit Einselement und eine lineare Abbildung $i : V \to \mathrm{Cl}(V,b)$, sodass $i(v)i(v) = b(v,v)1$ für alle $v \in V$ gilt und die folgende universelle Eigenschaft erfüllt ist. Ist A eine assoziative Algebra mit Einselement und $f : V \to A$ eine lineare Abbildung, sodass $f(v)f(v) = b(v,v)1$ gilt, dann gibt es einen eindeutigen Homomorphismus $\tilde{f} : \mathrm{Cl}(V,b) \to A$ von Algebren, sodass $\tilde{f} \circ i = f$ gilt. Durch diese Eigenschaft ist $\mathrm{Cl}(V,b)$ bis auf natürliche Isomorphie eindeutig bestimmt.*

Beweis. Der Beweis ist eigentlich ganz simpel. Man betrachtet die Tensoralgebra $\otimes V$ und darin das Ideal I_B, das von der Menge $B := \{v \otimes v - b(v,v)1 : v \in V\}$ erzeugt wird (siehe Proposition 11.11). Dann definiert man $\mathrm{Cl}(V,b) := \otimes V/I_B$, bezeichnet mit $\pi : \otimes V \to \mathrm{Cl}(V,b)$ die kanonische Surjektion und definiert $i : V \to \mathrm{Cl}(V,b)$ als die Einschränkung von π auf $\otimes^1 V = V$. Dann ist $\mathrm{Cl}(V,b)$ nach Konstruktion eine assoziative $\mathbb{K}$–Algebra mit Einselement und $i : V \to \mathrm{Cl}(V,b)$ ist eine lineare Abbildung. Für $v \in V = \otimes^1 V$ gilt nach Definition der Multiplikation $i(v)i(v) = \pi(v \otimes v)$ und $1 = \pi(1)$. Da $v \otimes v - b(v,v)1$ in I_B liegt, erhalten wir $0 = \pi(v \otimes v) - b(v,v)\pi(1)$, also $i(v)i(v) = b(v,v)1$.

Sei nun A eine assoziative Algebra mit Einselement und $f : V \to A$ eine lineare Abbildung, sodass $f(v)f(v) = b(v,v)1$ für alle $v \in V$ gilt. Dann liefert die universelle Eigenschaft der Tensoralgebra einen Homomorphismus $\hat{f} : \otimes V \to A$ von Algebren mit Einselement, der auf $V = \otimes^1 V$ mit f übereinstimmt. Insbesondere gilt

$$\hat{f}(v \otimes v - b(v,v)1) = f(v)f(v) - b(v,v)1 = 0.$$

Damit ist aber $\mathrm{Ker}(\hat{f})$ ein Ideal in $\otimes V$, dass B enthält. Nach Proposition 11.11 folgt $I_B \subset \mathrm{Ker}(\hat{f})$, also gibt es eine lineare Abbildung $\tilde{f} : \otimes V/I_B = \mathrm{Cl}(V,b) \to A$, sodass $\tilde{f} \circ \pi = \hat{f}$. Daraus folgt einerseits sofort, dass $\tilde{f}$ ein Algebrahomomorphismus ist. Andererseits gilt $\tilde{f} \circ i = \hat{f}|_{\otimes^1 V} = f$. Damit ist die Existenz von $\tilde{f}$ bewiesen.

Zur Eindeutigkeit von $\tilde{f}$: Wir wissen bereits, dass die Elemente der Form 1 und $v_1 \otimes \cdots \otimes v_k$ mit $1 \leq k$ und $v_i \in V$ für alle $i = 1, \dots, k$ den Vektorraum aufspannen. Da π surjektiv ist, spannt die 1 gemeinsame mit den Elementen der Form $\pi(v_1 \otimes \cdots \otimes v_k)$ den Vektorraum $\mathrm{Cl}(V,b)$ auf. Nun ist aber nach Definition der Multiplikation

$$\pi(v_1 \otimes \cdots \otimes v_k) = \pi(v_1) \cdots \pi(v_k) = i(v_1) \cdots i(v_k),$$

also wird der Vektorraum $\mathrm{Cl}(V,b)$ von allen Produkten von Elementen der Form $i(v)$ mit $v \in V$ aufgespannt. Somit ist ein Homomorphismus von Algebren mit Einselement von $\mathrm{Cl}(V,b)$ in eine beliebige $\mathbb{K}$–Algebra durch seine Werte auf den Elementen der Form $i(v)$ eindeutig bestimmt.

Die Tatsache, dass $\mathrm{Cl}(V,b)$ durch die universelle Eigenschaft bis auf Isomorphie eindeutig bestimmt ist, beweist man völlig analog wie unseren früheren Resultate dieser Art. □

An dieser Stelle sieht alles wunderbar einfach aus und man könnte den Eindruck bekommen, dass man mit solchen universellen Konstruktionen ganz leicht alle gewünschten Eigenschaften erzwingen kann. Bis zu einem gewissen Grad stimmt das auch, es gibt allerdings ein gravierendes Problem. Wir können nämlich für's Erste gar nicht sagen, wie die Algebra $\mathrm{Cl}(V,b)$ tatsächlich aussieht. Dadurch, dass wir einfach das von einer gewissen Teilmenge $B \subset \otimes V$ erzeugte Ideal I_B betrachtet haben, haben wir die Kontrolle darüber verloren, wie groß I_B tatsächlich ist. Es könnte ja durchaus passieren, dass $I_B = \otimes V$ und damit $\mathrm{Cl}(V,b) = \{0\}$ gilt, das haben wir in unserem Beweis nirgends ausgeschlossen. (Das würde natürlich auch bedeuten, dass eine lineare Abbildung $f : V \to A$, die $f(v)f(v) = b(v,v)1$ erfüllt, automatisch 0 sein müsste, aber auch dagegen spricht im Moment nichts.) Mit schlecht gewählten Relationen passiert es leicht, dass das entsprechende Ideal viel größer ist, als man auf den ersten Blick vermuten würde.

Für die Clifford Algebren ist das natürlich nicht der Fall, aber der Beweis dafür ist relativ aufwändig. Ein erster Hinweis ergibt sich, indem man Spezialfälle betrachtet. Im Fall $b = 0$ wird das Ideal I_B von der Menge $B = \{v \otimes v : v \in V\}$ erzeugt. Aus der Definition der Multiplikation auf der Tensoralgebra folgt damit sofort, dass $I_B \subset \oplus_{k\geq 2} \otimes^k V$ gilt und damit insbesondere die Abbildung $i : V \to \mathrm{Cl}(V,0)$ injektiv ist. Hier können wir aber I_B explizit beschrieben. Betrachten wir nämlich das Ideal I, das in der Konstruktion der äußeren Algebra ΛV in 11.12 verwendet wurde. Nach Definition wird I von allen Tensorprodukten von Elementen von V aufgespannt, in denen mindestens ein Faktor zwei Mal vorkommt. Insbesondere ist natürlich $I \cap \otimes^2 V$ gerade die lineare Hülle von $B \subset \otimes^2 V$. Damit folgt $B \subset I$, also $I_B \subset I$. Andererseits ist I_B ein Teilraum, muss also die lineare Hülle von B enthalten. Damit folgt $I \cap \otimes^2 V \subset I_B$.

Wir behaupten nun, dass $I \subset I_B$ und damit $I = I_B$ und $\mathrm{Cl}(V,0) \cong \Lambda V$ gilt. Dazu zeigt man durch Induktion nach k, dass $I \cap \otimes^k V \subset I_B$ gilt, wobei wir den Induktionsanfang $k = 2$ bereits erledigt haben. Nehmen wir also $k > 2$ an, dann genügt es zu zeigen, dass ein Tensorprodukt von k Elementen von V, in dem mindestens zwei Faktoren gleich sind, schon in I_B liegt. Betrachten wir also $v_1 \otimes \cdots \otimes v_k$. Ist $v_i = v_j$ für $i, j > 1$ und $i \neq j$, dann liegt $v_2 \otimes \cdots \otimes v_k$ in $I \cap \otimes^{k-1} V$, also nach Induktionsvoraussetzung in I_B. Weil I_B ein Ideal ist, liegt auch das Tensorprodukt von v_1 mit diesem Element in I_B. Analog behandelt man den Fall $i, j < k$, also müssen wir nur noch zeigen, dass $v \otimes w_1 \otimes \cdots \otimes w_{k-2} \otimes v \in I_B$ gilt. Nun liegen aber nach Definition $v \otimes v$, $w_1 \otimes w_1$ und $(v + w_1) \otimes (v + w_1)$ in I_B, also auch $v \otimes w_1 + w_1 \otimes v$. Bildet man das Tensorprodukt mit $w_2 \otimes \cdots \otimes w_{k-2} \otimes v$ dann liegt das in I_B und ist die Summe des gesuchten Elements mit einem Element, das nach Induktionsvoraussetzung in I_B liegt, also folgt die Behauptung. Man kann also Clifford Algebren als eine Verallgemeinerung der äußeren Algebra betrachten.

Der zweite wichtige Spezialfall ist, dass die symmetrische Bilinearform nicht degeneriert ist. Analog wie im Fall $b = 0$ bemerken wir, dass für $v, w \in V$ das Element $(v + w) \otimes (v + w) + b(v + w, v + w)1$ in I_B liegt. Expandiert man in beiden Summanden, dann erhält man die Summe der Elemente zu v und w, die in I_B liegen mit $v \otimes w + w \otimes v + 2b(v,w)1$, also liegt auch dieses Element in I_B. Damit folgt aber sofort, dass in $\mathrm{Cl}(V,b)$ immer $i(v)i(w) + i(w)i(v) = 2b(v,w)1$. Damit sieht man sofort, dass $i : V \to \mathrm{Cl}(V,b)$ injektiv ist, denn aus $i(v) = 0$ folgt $0 = 2b(v,w)$ für alle $w \in V$, also $v = 0$. Aus Abschnitt 9.11 wissen wir, dass wir eine Basis $\{v_1, \dots, v_n\}$ für V finden können, sodass $b(v_j, v_j) =: \varepsilon_j = \pm 1$ für alle $j = 1, \dots, n$ und $b(v_j, v_k) = 0$ für $j \neq k$ gilt. In der Clifford Algebra gilt somit $i(v_j)i(v_j) = \varepsilon_j 1$ und $i(v_j)i(v_k) = -i(v_k)i(v_j)$ für $j \neq k$.

Die Tatsache, dass iterierte Tensorprodukte von Basiselementen die Tensoralgebra $\otimes V$ aufspannen zeigt nun, dass die iterierten Produkte der Form $i(v_{j_1}) \dots i(v_{j_\ell})$ die Cliffordalgebra $\mathrm{Cl}(V, b)$ aufspannen. Kommen in so einem Produkt zwei gleiche Faktoren vor, dann kann man sie nebeneinander tauschen, was das Element entweder unverändert lässt oder mit -1 multipliziert und dann weglassen, was wiederum bestenfalls das Vorzeichen ändert. Also genügt es, Produkte dieser Form mit lauter verschiedenen Faktoren zu betrachten. In diesen kann man aber dann die Faktoren umordnen, was wieder nur das Vorzeichen ändern kann. Insgesamt sieht man, dass das Element 1 gemeinsam mit den Produkten $i(v_{j_1}) \dots i(v_{j_\ell})$ mit $1 \leq j_1 < j_2 < \dots < j_\ell \leq n$ die Cliffordalgebra aufspannen. Insbesondere ist $\dim(\mathrm{Cl}(V, b)) \leq 2^n$.

Als nächstes zeigt man, dass für nicht degeneriertes b die oben aufgelisteten Elemente immer linear unabhängig sind, also $\dim(\mathrm{Cl}(V, b)) = 2^n$ gilt. Dann kann man das Resultat von oben, dass $\mathrm{Cl}(V, 0) \cong \Lambda V$ ist, benutzen um das auf den Fall allgemeiner Bilinearformen erweitern. Es gilt also immer $\dim(\mathrm{Cl}(V, b)) = 2^n$ und man kann eine Basis für die Cliffordalgebra analog wie oben konstruieren.

Wichtig für Anwendungen sind hauptsächlich die Cliffordalgebren für nicht degeneriertes b. Das liegt einerseits an einem Zusammenhang mit orthogonalen Gruppen. Da jede Cliffordalgebra assoziativ ist, bilden die invertierbaren Elemente in $\mathrm{Cl}(V, b)$ immer eine Gruppe. Indem man die Struktur der Clifford Algebra genauer analysiert, kann man (ähnlich wie in den Konstruktionen mit Quaternionen in 9.16 und 9.17) darin Untergruppen finden, die man als lineare Abbildungen $f : V \to V$ interpretieren kann, die $b(f(v), f(w)) = b(v, w)$ für alle $v, w \in V$ erfüllen. Das führt zur Theorie der sogenannten Spin–Gruppen.

Andererseits kann man zu gegebenem (V, b) die Komplexifizierung $V_{\mathbb{C}} := V \otimes_{\mathbb{R}} \mathbb{C}$ betrachten, und sieht leicht, dass b darauf eine komplexe Bilinearform $b_{\mathbb{C}}$ induziert. Man sieht dann leicht, dass man $\mathrm{Cl}(V, b)$ als reelle Teilalgebra der komplexen Clifford Algebra $\mathrm{Cl}(V_{\mathbb{C}}, b_{\mathbb{C}})$ betrachten kann. Wie wir in 9.11 besprochen haben, ist aber die Klassifikation der komplexen symmetrischen Bilinearformen wesentlich einfacher als im reellen Fall. Damit kann man beweisen, dass $\mathrm{Cl}(V_{\mathbb{C}}, b_{\mathbb{C}})$ als Algebra entweder isomorph zu einer Matrixalgebra $M_N\mathbb{C}$ für passendes N oder zu einer direkten Summe von zwei solchen Matrixalgebren ist. Das liefert dann Realisierungen der oben erwähnten Spin Gruppen durch lineare Abbildungen auf $\mathbb{C}^N$, die sogenannten Spin Darstellungen, die in einigen Teilgebieten der Mathematik und der theoretischen Physik wichtig sind.

Index