LINEARE ALGEBRA I, 2012/2013

RAINER SCHULZE-PILLOT

Inhaltsverzeichnis

0. Einleitung

Die Vorlesung "Lineare Algebra" hat Probleme als Ausgangspunkt, die Sie aus dem Mathematikunterricht der Oberstufe gut kennen.

Wir schauen uns als Beispiel Abituraufgaben an:

Abitur Saarland 2003, Aufgabe 2:

1. Gegeben ist eine gerade Pyramide (siehe Zeichnung) (hier nicht wiedergegeben) mit quadratischer Grundfläche $ABCD$ und Spitze S. Die Seitenlänge des in der x_1x_2-Ebene liegenden Quadrates $ABCD$ beträgt 80 m; die Pyramide hat eine Höhe von 60 m.

1.1 Stellen Sie eine Normalengleichung der Ebene e auf, in der die Seitenfläche ABS liegt.

1.2 Berechnen Sie den Winkel, den die Ebene $e : 3x_2 + 2x_3 - 120 = 0$ (Teil 1.1) mit der Pyramidenkante $\overline{DS}$ bildet.

1.3 Im angegebenen Koordinatensystem der Pyramide ist ein Richtungsvektor der Sonnenstrahlen $\vec{u} = \begin{pmatrix} 2 \\ 4 \\ -3 \end{pmatrix}$. Der Schattenpunkt S' der Pyramidenspitze S liegt in der x_1x_2-Ebene. Berechnen Sie die Koordinaten von S'.

1.4 Wie weit ist der Punkt $S'(40|80|0)$ von den Eckpunkten A und B der Pyramide entfernt?

1.5 Begründen Sie:
Jeder Punkt der Pyramidenhöhe $\overline{OS}$ hat von den vier Seitenflächen der Pyramide den gleichen Abstand.
Bestimmen Sie den Punkt von $\overline{OS}$, der sowohl von den vier Seitenflächen als auch von der Grundfläche der Pyramide den gleichen Abstand hat.

2. Zeigen Sie mit den Mitteln der Vektorrechnung:
In einem Trapez, in dem die eine Grundseite doppelt so lang ist wie die andere, teilen sich die Diagonalen im Verhältnis $2 : 1$.

   Hinweis: Die zueinander parallelen Seiten eines Trapezes heißen Grundseiten.

Zunächst 1.1: Zur Berechnung der Normalengleichung $\vec{x}\vec{n} + d = 0$ (oder ausmultipliziert: $x_1n_1 + x_2n_2 + x_3n_3 + d = 0$) haben wir verschiedene Möglichkeiten:

a) Die (hier fehlende) Zeichnung gibt ein Koordinatensystem, in dem wir haben: $A = (-40|40, 0), B = (40|40|0), S = (0|0|60)$. Wir setzen die drei Punkte $A,\ B,\ S$ ein, die in der Ebene liegen und erhalten die drei Gleichungen

$$\begin{aligned} -40n_1 + 40n_2 + d &= 0 \\ 40n_1 + 40n_2 + d &= 0 \\ 60n_3 + d &= 0. \end{aligned}$$

Rechnung ergibt rasch, dass alle Lösungen proportional zu der Lösung $n_1 = 0$, $n_2 = 3$, $n_3 = 2$, $d = -120$ sind.

Man setzt also $\vec{n} = \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$, $d = -120$ und hat die Normalengleichung.

b) Der Normalenvektor $\vec{n}$ muss senkrecht auf $\vec{AB}$ und $\vec{SB}$ stehen, also proportional zum Vektorprodukt (Kreuzprodukt) dieser beiden Vektoren sein.

$$\vec{AB} \times \vec{SB} = \begin{pmatrix} 80 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 40 \\ 40 \\ -60 \end{pmatrix} = \begin{pmatrix} 0 \\ 4800 \\ 3200 \end{pmatrix}$$

ist proportional zu $\begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$.

Wir setzen $\vec{n} = \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$ und berechnen d aus $\vec{x}\vec{n} = -d$ für jedes $\vec{x}$ in der Ebene.

Setzen wir etwa für $\vec{x}$ den Ortsvektor von A ein, so erhalten wir wieder $d = -120$.

1.2 lassen wir aus, 1.3 geht offenbar so:

Sei $\vec{s}$ der Ortsvektor von S, $\vec{s'}$ der Ortsvektor des Schattenpunkts S'. Man erhält S', indem man von S so weit in Richtung des Vektors $\vec{u}$ geht, bis man auf die x_1x_2-Ebene trifft.

Also: $\vec{s'} = \vec{s} + t \begin{pmatrix} 2 \\ 4 \\ -3 \end{pmatrix}$, wobei t so gewählt wird, dass die letzte Koordinate von S' gleich 0 ist.

$$\begin{pmatrix} s'_1 \\ s'_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 60 \end{pmatrix} + t \begin{pmatrix} 2 \\ 4 \\ -3 \end{pmatrix}$$

$$\Rightarrow t = 20, \quad s'_1 = 40, \quad s'_2 = 80.$$

1.4 und 1.5 lassen wir wieder aus und schauen uns 2. an:

Wir können die Ecken des Trapezes als die Punkte $(0|0)$, $(2|0)$, $(a|b)$, $(a+1|b)$ mit beliebigen von 0 verschiedenen a, b annehmen. Die beiden Diagonalen haben dann die Parameterdarstellungen

$$t \begin{pmatrix} a+1 \\ b \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} a \\ b \end{pmatrix} + s \begin{pmatrix} 2-a \\ -b \end{pmatrix}.$$

Im Schnittpunkt sind die Koordinaten gleich, wir erhalten die Gleichungen

$$\begin{aligned} t(a+1) &= a + s(2-a) \\ tb &= b(1-s). \end{aligned}$$

Es folgt: $t = 1 - s$, also

$$\begin{aligned} &(1-s)(a+1) = a + s(2-a) \\ &\Rightarrow s = \tfrac{1}{3}, \quad t = \tfrac{2}{3}, \end{aligned}$$

die Diagonalen teilen sich also gegenseitig im Verhältnis 2:1 wie behauptet.

Ganz ähnlich ist die

Aufgabe 2 von 2004:

1. Gegeben sind die Punkte $A(4|2|5)$, $B(6|0|6)$ und die Gerade g : $\vec{x} = \begin{pmatrix} 6 \\ 6 \\ 9 \end{pmatrix} + \lambda \cdot \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$.

1.1 Berechnen Sie eine Koordinatengleichung der Ebene e, die den Punkt A und die Gerade g enthält und weisen Sie nach, dass auch der Punkt B in dieser Ebene liegt.

1.2 Auf der Geraden g gibt es einen Punkt C so, dass die Strecken $\overline{AB}$ und $\overline{BC}$ senkrecht aufeinander stehen. Berechnen Sie die Koordinaten des Punktes C.
(Zur Kontrolle: $C(7|2|8)$)

1.3 Ergänzen Sie das rechtwinklige Dreieck ΔABC durch Berechnung des Punktes D zum Rechteck $ABCD$ und zeigen Sie dann, dass dieses Rechteck sogar ein Quadrat ist.

1.4 Das Quadrat $ABCD$ ist die Grundfläche einer geraden quadratischen Pyramide, deren Spitze S in der $x-z$-Ebene liegt. Berechnen Sie die Koordinaten der Pyramidenspitze S und das Volumen der Pyramide $ABCDS$.
(Zur Kontrolle: $S(1,5|0|10,5)$)

1.5 Es gibt eine Kugel, die durch alle Eckpunkte der Pyramide $ABCDS$ geht. Berechnen Sie die Koordinaten des Mittelpunktes M dieser Kugel.

2. Ein Würfel mit der Kantenlänge a ist gemäß folgender Abbildung in einem kartesischen Koordinatensystem positioniert.

2.1 Berechnen Sie das Maß des Winkekls zwischen zwei Raumdiagonalen des Würfels.

2.2 Zeigen Sie: Der Abstand der Würfelecke P_2 von der Raumdiagonalen $\overline{P_5P_3}$ beträgt $\frac{1}{3}a\sqrt{6}$.

1.1 sucht eine Koordinatengleichung $ax + by + cz + d = 0$.

In der Ebene liegen die Punkte $A(4|2|5)$, $(6|6|9)$ $(\lambda = 0)$ und $\begin{pmatrix} 5 \\ 10 \\ 10 \end{pmatrix}$

($\lambda = 1$). Wir setzen ein und erhalten die Gleichungen

$$\begin{aligned} 4a + 2b + 5c + d &= 0 \\ 6a + 4b + 9c + d &= 0 \\ 5a + 10b + 10c + d &= 0, \end{aligned}$$

geschicktes Auflösen liefert die Lösung

$$a = 1, \quad b = 1, \quad c = -2, \quad d = 0,$$

zu der alle anderen Lösungen proportional sind.

Alternativ hätten wir wieder das Vektorprodukt zweier Vektoren auszurechnen, die Punkte in der Ebene verbinden. Einsetzen zeigt, dass B in der Ebene liegt.

1.2:

$$\begin{aligned} \vec{AB} &= \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \\ \vec{BC} &= \begin{pmatrix} 6 - \lambda - 6 \\ 6 + 4\lambda \\ 9 + \lambda - 6 \end{pmatrix} = \begin{pmatrix} -\lambda \\ 6 + 4\lambda \\ 3 + \lambda \end{pmatrix}, \end{aligned}$$

wenn wir C mit Hilfe der gegebenen Parameterdarstellung von g schreiben.

$\vec{AB}$ und $\vec{BC}$ stehen genau dann senkrecht, wenn das Skalarprodukt dieser beiden Vektoren 0 ist, das liefert die Beziehung

$$\begin{aligned} &-2\lambda - 12 - 8\lambda + 3 + \lambda = 0, \text{ also} \\ &\lambda = -1, \quad C = (7|2|8). \end{aligned}$$

Die weiteren Rechnungen ersparen wir uns und überlegen statt dessen, wohin wir uns von diesem Ausgangspunkt bewegen wollen - sicher wird es nicht darum gehen, ein weiteres halbes Jahr mit dem Rechnen von Aufgaben dieses wohlbekannten Typs zuzubringen, das wäre ja auch langweilig.

Was war unseren Aufgaben gemeinsam?

Es ging um geometrische Probleme im Raum bzw. in der Ebene, die mit algebraischen Methoden gelöst werden: Der Anschauungsraum wird über die Einführung von Koordinaten bezüglich eines kartesischen Koordinatensystems mit dem $\mathbb{R}^3$ identifiziert, man rechnet mit den Koordinaten der Punkte bzw. der Vektoren, die als Differenz zweier Punkte auftreten (bzw. als die Translation, die einen Punkt in den anderen verschiebt).

Diese Rechnungen übersetzen das gegebene geometrische Problem in die Aufgabe, ein System von linearen Gleichungen in einer, zwei, drei oder vier Variablen zu lösen, was durch (mehr oder minder) geschicktes Eliminieren von Variablen geschieht.

Eine (in der Schule meist benutzte) Vereinfachung lieferte die Möglichkeit, das Vektorprodukt zu benutzen. Statt bei der Aufgabe 1.1 durch Einsetzen von drei Punkten $A(x_1|x_2|x_3)$, $B(x_1'|x_2'|x_3')$, $C(x_1''|x_2''|x_3'')$ in die zu findende Ebenengleichung $ax + by + cz + d = 0$ die drei Gleichungen

$$\begin{array}{lrcl} (A) & ax_1 + bx_2 + cx_3 + d & = & 0 \\ (B) & ax_1' + bx_2' + cx_3' + d & = & 0 \\ (C) & ax_1'' + bx_2'' + cx_3'' + d & = & 0 \end{array}$$

in den Unbekannten a, b, c, d zu erhalten, bilden wir die Differenzen $\vec{AB}$, $\vec{AC}$ und suchen einen Vektor $\vec{n}$, der auf diesen senkrecht steht: Mit

$$\vec{AB} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1' - x_1 \\ x_2' - x_2 \\ x_3' - x_3 \end{pmatrix}, \quad \vec{AC} = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} x_1'' - x_1 \\ x_2'' - x_2 \\ x_3'' - x_3 \end{pmatrix}$$

erhalten wir die zwei Gleichungen

$$\begin{array}{rcl} n_1y_1 + n_2y_2 + n_3y_3 & = & 0 \\ n_1y_1' + n_2y_2' + n_3y_3' & = & 0, \end{array}$$

oder äquivalent durch Einsetzen (und mit $a = n_1$, $b = n_2$, $c = n_3$):

$$\begin{array}{lrcl} (B') & a(x_1' - x_1) + b(x_2' - x_2) + c(x_3' - x_3) & = & 0 \\ (C') & a(x_1'' - x_1) + b(x_2'' - x_2) + c(x_3'' - x_3) & = & 0, \end{array}$$

die wir auch als $(B') = (B) - (A)$ bzw. $(C') = (C) - (A)$ aus dem ersten Gleichungssystem erhalten.

Bilden des Vektorprodukts liefert uns dann die Lösung

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \times \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix},$$

also

$$\begin{array}{rcl} a & = & y_2y_3' - y_3y_2' \\ b & = & y_3y_1' - y_1y_3' \\ c & = & y_1y_2 - y_2y_1'. \end{array}$$

Der Satz, dass das Vektorprodukt senkrecht auf beiden Faktoren steht, erspart uns also hier das Lösen des Gleichungssystems durch Elimination von Variablen: Wir haben eine Formel, die uns die Lösung für die zwei Gleichungen (B'), (C') liefert.

Damit haben wir auch schon die ersten Programmpunkte für diese Vorlesung:

- Beschreibe ein Lösungsverfahren für (beliebig große) lineare Gleichungssysteme (in beliebig vielen Variablen).
- Suche eine Formel für Lösungen.

Allerdings werden wir bei der Lösung dieser Aufgaben anders vorgehen, als Sie es aus der Schule gewöhnt sind:

- Im Vordergrund steht nicht das Üben und Beherrschen von Lösungstechniken, sondern das Studium abstrakter Begriffe, die zunächst beim Lösen linearer Gleichungssysteme und beim Behandeln analytisch-geometrischer Probleme entstehen und anschließend in allen mathematischen Disziplinen, rein oder angewandt, Grundlage der Überlegungen sind. Dies beginnt mit dem schon aus der Schule bekannten Begriff des Vektorraums, es folgen Symmetrien, Gruppen, Abbildungen und vieles mehr.
- Besonderen Wert legen wir auf saubere Begründungen der Lehrsätze durch logisch einwandfreie Beweise. Ziel des Studiums ist ja, dass Sie nicht nur erlernte Methoden anwenden können sondern in der Lage sind, selbst neue Methoden zu finden oder Analoga zu bekannten Methoden in neuen Situationen einzuführen. Das ist nur möglich, wenn man neue Verfahren und Ideen in nachprüfbarer Weise begründen kann.
- Speziell möchte ich hier die Lehramtsstudierenden ansprechen. Es wird in der letzten Zeit viel davon gesprochen, dass es notwendig sei, das Lehramtsstudium nicht mit letzten Endes überflüssigem Fachwissen zu überfrachten, worauf es ankomme, sei eine verlässliche Beherrschung des Schulstoffs sowie pädagogische Fähigkeiten. Natürlich ist Überfrachten schädlich (sonst finge das Wort nicht mit "Über" an), natürlich sind Ihre pädagogischen Fähigkeiten (und vor allem Ihr Interesse am Umgang mit Schülern) wichtig für Ihren Beruf. Wenn Sie aber lebendigen Unterricht geben wollen, mit dem Sie Schülerinnen und Schüler für die Mathematik interessieren und vielleicht sogar begeistern wollen, dann müssen Sie erheblich mehr können als das, was Sie Tag für Tag beibringen. Auch von Lehrerinnen und Lehrern für Geschichte oder Englisch (z.B.) erwartet man ja, dass sie nicht nur den Schulstoff beherrschen, sie sollen etwa für Geschichte wissen, wie Geschichtswissenschaft arbeitet, wie das entsteht, was in Geschichtsbüchern steht und wie die verschiedenen Unterrichtsgegenstände zusammenhängen, sie sollten etwa für Englisch die Kultur und Geschichte des Landes kennen und die Literatur kennen und lieben.

  Genauso sollten Lehrerinnen und Lehrer für Mathematik die exakte Begründung der Differentialrechnung kennen, um selbst beurteilen zu können, welche Ausschnitte man im Unterricht präsentiert, sie sollten abstrakte algebraische Strukturen ebenso kennen wie die Techniken der angewandten Mathematik, um den Lehrstoff richtig einordnen zu können und bei zweifelsohne anstehenden Lehrplanreformen treibende Kraft und nicht geplagtes Opfer zu sein.

- Die Beherrschung von Lösungsverfahren steht zwar nicht im Vordergrund, ist aber auch nicht unwichtig; man muss erlernte Lösungsverfahren schon alleine deshalb ein paarmal durchführen, um sicher zu sein, dass man sie richtig verstanden hat. Virtuosität anzustreben, lohnt sich allerdings in der Regel nicht, dafür rechnet der Computer zu gut. Wir werden daher stets auch anschauen, wie man konkrete Rechenverfahren mit Hilfe eines Computeralgebrasystems, etwa MAPLE, rasch und kraftsparend durchführt. Schon ein paar einfache Beispiele für lineare Gleichungssysteme dürften davon überzeugen, dass man hier nicht mit dem Rechner konkurrieren sondern lieber lernen sollte, ihn sinnvoll einzusetzen.

Zum Abschluss dieses einleitenden Teils möchte ich ganz kurz skizzieren, womit ich mich wissenschaftlich beschäftige:

Mein Spezialgebiet ist Zahlentheorie. Die Zahlentheorie beschäftigt sich mit grundlegenden Eigenschaften der ganzen Zahlen und mit dem Studium ganzzahliger Lösungen von Gleichungen.

Klassische Sätze der Zahlentheorie sind etwa:

- Eine Primzahl p lässt sich ganau dann als $p = x^2+y^2$ mit $x, y \in \mathbb{Z}$ schreiben, wenn $p-1$ durch 4 teilbar ist.
- Eine (positive) ganze Zahl n lässt sich genau dann als $n = x^2 + y^2 + z^2$ mit $x, y, z \in \mathbb{Z}$ schreiben, wenn n nicht von der Form
$$4^j(8k+7) \text{ mit } j, k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$$
ist.
- Bezeichnet $\pi(X)$ die Anzahl der Primzahlen $p \leq X$, so strebt der Quotient
$$\frac{\pi(X) \cdot \log\ X}{X} \text{ für } X \longrightarrow \infty \text{ gegen } 1.$$

Ein nicht ganz so klassischer Satz der Zahlentheorie wurde vor 17 Jahren von Andrew Wiles bewiesen:

> Die Gleichung $x^n + y^n = z^n$ hat für natürliches $n \geq 3$ keine Lösung x, y, z mit $x, y, z \in \mathbb{Z}$, $x \cdot y \cdot z \neq 0$.

Eine spannende (aber nicht ganz korrekte) Schilderung der Entdeckung dieses Beweises wird in dem Buch „Fermats letzter Satz“ von Simon Singh gegeben.

Eine offene Frage der Zahlentheorie ist:

> Gilt die Vermutung von Goldbach: Jede gerade Zahl kann man als Summe von zwei Primzahlen schreiben?

In dem Roman „Onkel Petros und die Goldbach'sche Vermutung“ von A. Doxiadis geht es um einen Mathematiker, der versucht, diese Vermutung zu beweisen.

Man weiß über diese Vermutung außer numerischer Evidenz für sie nur:

> Jede hinreichend große Zahl lässt sich als Summe $p_1 + p_2 p_3$ oder als $p_1 + p_2$ mit Primzahlen p_1, p_2, p_3 schreiben (das wurde von dem chinesischen Mathematiker Chen bewiesen).

1. Grundlagen.

1.1. **Mengenlehre, Logik.** Eine axiomatische Einführung in die Mengenlehre würde (wenigstens) ein ganzes Semester beanspruchen, man schaue sich etwa das Lehrbuch von Deiser an.

Wir begnügen uns daher mit einem "naiven" Standpunkt, ergänzt durch ein paar Vorsichtsmaßregeln und einen Einblick in die Axiomatik:

Begründung durch Cantor (1845–1918):

> "Eine Menge ist jede Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens"

Was sind "Denken", "Anschauung", "Zusammenfassung"? Wann sind Objekte "wohlunterschieden", wann nicht?

Die Problematik wird klar durch "Antinomien". Am bekanntesten (und grundlegendsten):

Wie üblich schreiben wir: $x \in M$ (x ist Element von M), die Elemente von M sind die in M zusammengefassten Objekte.

Mengen sind Objekte unseres Denkens, wir können Sie zu Mengen zusammenfassen.

Wir bilden daher die Menge M_1, deren Objekte sämtliche Mengen sind. Es gilt: $M \in M_1$ für alle Mengen M. Insbesondere: $M_1 \in M_1$ (verwirrend, aber offensichtlich denkbar). Es gibt sicher Mengen, die nicht Elemente von sich selbst sind. Bilde daher

$$M_2 := \{M \text{ Menge} \mid M \notin M\}.$$

Frage: Gilt $M_2 \in M_2$?
Unterscheide:
$M_2 \notin M \Rightarrow M_2 \in M_2$ Wdsp.
$M_2 \in M \Rightarrow M_2 \notin M_2$ Wdsp.
Die Bildung von M_2, obwohl legal, führt auf Widersprüche. Daher: Axiomatischer Zugang.
Die Mengenlehre handelt von zwei Arten von Objekten: "Mengen", "Dinge".
Für eine Menge M und ein Ding x ist die Aussage "x ist Element von M" ($x \in M$) wahr oder falsch (schreibe dann: $x \notin M$). Es gelten die Zermelo (1871–1953)-Fraenkel (1891-1965)-Axiome, für die wieder auf das Buch von Deiser verwiesen sei.

Die für uns wichtigsten sind:

- **Extensionalität:** Die Mengen A und B sind genau dann gleich, wenn sie die selben Elemente haben (also: $x \in A \Leftrightarrow x \in B$).

- **Aussonderungsaxiom:** Zu jeder Menge A und jeder Aussage P über Elemente von A gibt es eine Menge B, so dass gilt: B besteht genau aus den Elementen von A, für die Aussage P wahr ist.

Wichtig also:

- Wir betrachten nur die Elementbeziehung. Sieht man Mengen als Beutel an, in denen die Objekte gesammelt werden, so gilt: Beutel können nicht aus verschiedenen Stoffen oder verschieden farbig sein: Nur der Inhalt zählt. Ein Element gehört zur Menge oder nicht (nicht doppelt oder vielleicht).
- Wir können durch beliebige Aussagen Teilmengen bilden: Schreibweise:
$$N = \{x \in M \mid P(x) \text{ ist wahr } \}$$
(etwa $\mathbb{R}_{>0} = \{x \in \mathbb{R} \mid x > 0\}$).

Aber: Die Bildung beliebiger Zusammenfassungen ist nicht garantiert.

Wir fassen die wichtigsten Schreibweisen zusammen:

Schreibweisen der Mengenlehre:

$x \in M$	x ist Element von M
$x \notin M$	x ist nicht Element von M
$M_1 \subseteq M_2$	M_1 ist enthalten in M_2, M_1 ist Teilmenge von M_2: Für alle $x \in M_1$ gilt $x \in M_2$. Man schreibt auch: $M_1 \subset M_2$.
$M_1 \subsetneqq M_2$	$M_1 \subseteq M_2$ mit $M_1 \neq M_2$
$M_2 \supseteq M_1$	M_2 ist Obermenge von M_1: gleichwertig zu $M_1 \subseteq M_2$.
$\emptyset$	Die leere Menge. Sie hat kein Element und ist Teilmenge jeder Menge. $x \notin \emptyset$ ist stets wahr, $x \in \emptyset$ stets falsch.
$M_1 \cup M_2$	Vereinigung von M_1 und M_2: x ist genau dann Element von $M_1 \cup M_2$, wenn x Element von M_1 oder Element von M_2 ist (oder von beiden Mengen, s.u.).
$M_1 \cap M_2$	Durchschnitt von M_1 und M_2: x ist genau dann Element von $M_1 \cap M_2$, wenn x Element von M_1 und Element von M_2 ist. Ist $M_1 \cap M_2 = \emptyset$, so sagt man, M_1 und M_2 seien *disjunkt* oder *elementfremd*.
$M_1 \setminus M_2$	Differenz von M_1 und M_2: x ist genau dann Element von $M_1 \setminus M_2$, wenn $x \in M_1$ und $x \notin M_2$ gilt. Ist M_2 Teilmenge von M_1, so heißt $M_1 \setminus M_2$ auch das Komplement von M_1 in M_2.
$M_1 \times M_2$	Das kartesische Produkt Es besteht aus allen geordneten Paaren (m_1, m_2) mit $m_1 \in M_1, m_2 \in M_2$. Geordnete Paare heißt: Zwei Paare $(m_1, m_2), (m_1', m_2')$ sind genau dann gleich, wenn $m_1 = m_1'$ und $m_2 = m_2'$ gilt. (Man kann das Paarsymbol auch definieren, indem man setzt: $(a, b) = \{\{a\}, \{a, b\}\}$.) Entsprechend: $M_1 \times \cdots \times M_n$: geordnete n-Tupel $(m_1, \ldots, m_n)$.
$\mathfrak{P}(M)$	Die Potenzmenge von M. Ihre Elemente sind alle Teilmengen von M.

Wir halten ein paar grundlegende Regeln fest:

- $A \subseteq A$ ist stets wahr.
- $A = B$ ist äquivalent zu: $A \subseteq B$ und $B \subseteq A$.
- Mengen können durch Auflisten ihrer Elemente beschrieben werden:

  $$A = \{1, 2\},\ \emptyset = \{\ \},\ \mathfrak{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$

  oder mit Hilfe des Aussonderungsaxioms durch Angabe einer Obermenge und einer definierenden Eigenschaft:

  $$\begin{aligned} \{1, 2\} &= \{x \in \mathbb{N} \mid x \leq 2\} = \{x \in \mathbb{N} \mid x^2 \leq 4\} \\ &= \{1, 2, 2\} = \{1\} \cup \{2\} = \{1, 2, 3\} \setminus \{3, 4\}. \end{aligned}$$

  Liefern zwei Schreibweisen die selben Elemente, so sind die Mengen gleich (Extensionalitätsaxiom), etwa

  $$\emptyset = \{x \in \mathbb{N} \mid x < 0\}.$$

Weitere wichtige Regeln:

Satz 1.1. *Sind A, B, C Mengen, so gilt:*

a) *Assoziativgesetze:*

$$\begin{aligned} (A \cup B) \cup C &= A \cup (B \cup C) \\ (A \cap B) \cap C &= A \cap (B \cap C) \end{aligned}$$

b) *Kommutativgesetze:*

$$\begin{aligned} A \cap B &= B \cap A \\ A \cup B &= B \cup A \end{aligned}$$

c) *Distributivgesetze:*

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned}$$

Beweise: Siehe Übung

Satz 1.2. *A, B, C seien Mengen. Dann gilt:*

a) $A \subseteq A \cup B$, $B \subseteq A \cup B$
b) $A \cap B \subseteq A$, $A \cap B \subseteq B$
c) $A \cup \emptyset = A$
d) $A \cap C = A \Leftrightarrow A \subseteq C \Leftrightarrow A \cup C = C$

Beweise: Siehe Übung.

Definition 1.3. *Seien $A \subseteq M$ Mengen. Das Komplement von A in M ist $\mathcal{C}_M A := M \setminus A = \{x \in M \mid x \notin A\}$.*

Satz 1.4. *Es gilt für $A, B \subseteq M$:*

a) $\mathcal{C}_M(\mathcal{C}_M A) = A$
b) $\mathcal{C}_M(M) = \emptyset$, $\mathcal{C}_M(\emptyset) = M$

c) $\mathcal{C}_M(A \cup B) = \mathcal{C}_M A \cap \mathcal{C}_M B$,
$\mathcal{C}_M(A \cap B) = \mathcal{C}_M A \cup \mathcal{C}_M B$
(de Morgansche Gesetze)

Beweise: Siehe Übung

Logik: Auch hier könnte man eine ganze Vorlesung halten, das Thema gehört eher in die Philosophie.

Wir benötigen Folgendes:

(A) Alle Aussagen, die wir behandeln, sind wahr (hat Wahrheitswert W) oder falsch (hat Wahrheitswert F). "nicht ganz falsch", "sehr wahr" und weitere Graustufen oder Farbwerte kommen nicht vor.

(B) Zu jeder Aussage A gibt es die Verneinung (Negation) $\neg A$ ("nicht A"). Diese hat den zu A entgegengesetzten Wahrheitswert:
Ist A wahr, so ist $\neg A$ falsch.
Ist A falsch, so ist $\neg A$ wahr.
Es gilt also: $\neg(\neg A) = A$ hat gleichen Wahrheitswert wie A.

(C) Zu A und B haben wir die Verknüpfungen:
A und B ($A \wedge B$), A oder B ($A \vee B$).

Wahrheitstafeln:

A und B:

$A \setminus B$	W	F
W	W	F
F	F	F

A oder B:

$A \setminus B$	W	F
W	W	W
F	W	F

Insbesondere: "oder" ist stets "nicht-ausschließend" (im Gegensatz zur Alltagssprache). Ausschließendes "oder" wäre für uns die Aussage

$$(A \vee B) \wedge \neg(A \wedge B).$$

((A oder B) und (nicht (A und B)).
Die folgenden Aussagen haben jeweils den gleichen Wahrheitswert (sind äquivalent):

$$\neg(A \vee B) \text{ äquivalent zu } (\neg A) \wedge (\neg B)$$
$$\neg(A \wedge B) \text{ äquivalent zu } (\neg A) \vee (\neg B)$$

("Negation vertauscht die Faktoren", analog zu den de Morganschen Gesetzen der Mengenlehre.)

(D) Zu A und B betrachten wir die Verknüpfungen $A \Rightarrow B$ (aus A folgt B, A impliziert B, wenn A, dann B).

$A \Rightarrow B$:

$A \setminus B$	W	F
W	W	F
F	W	W

A heißt auch Voraussetzung, B Konklusion (Folgerung). B heißt notwendige Bedingung für A, A heißt hinreichende Bedingung für B.

Die Konvention weicht von der Umgangssprache teilweise ab, sie impliziert insbesondere keine Kausalität. Gilt $A \Rightarrow B$ und $B \Rightarrow A$, so haben A und B den gleichen Wahrheitswert (können aber beide falsch sein). Wir schreiben dann: $A \Leftrightarrow B$ (A ist äquivalent zu B).

Die Aussage $\neg B \Rightarrow \neg A$ hat die gleiche Wahrheitstafel, ist also äquivalent zu $A \Rightarrow B$.

Wir sagen $\neg B \Rightarrow \neg A$ entsteht aus $A \Rightarrow B$ durch Kontraposition. Also etwa: $x \in \mathbb{Q} \Rightarrow x^2 \neq 2$ ist äquivalent zu $x^2 = 2 \Rightarrow x \notin \mathbb{Q}$.
Das Prinzip des Widerspruchsbeweises für die Aussage A ist:
Zeige für eine (geeignete) Aussage B, dass $\neg A \Rightarrow B$ gilt, zeige, dass B falsch ist. Dann kann $\neg A$ nicht wahr sein, also ist A wahr. (Beispiel: A: Es gibt unendlich viele Primzahlen). Besonders erstaunlich, wenn A eine Existenzaussage ist (sehen wir später an Beispielen).

(E) Quantoren: Sei M eine Menge, für jedes $x \in M$ sei $P(x)$ eine Aussage (über x).
   a) $\forall x \in M : P(x)$ heißt:
      Für alle $x \in M$ ist Aussage $P(x)$ wahr.
      $\forall$: Der All-Quantor.
   b) $\exists x \in M : P(x)$ heißt :
      Es gibt (wenigstens) ein $x \in M$, für das $P(x)$ wahr ist.
      $\exists$: Existenzquantor. Verneinung $\nexists$.

Negation vertauscht Quantoren:

$$\neg(\exists x \in M : \quad P(x)) \Leftrightarrow \forall x \in M : \neg P(x),$$
$$\neg(\forall x \in M) : \quad P(x)) \Leftrightarrow \exists x \in M : \neg P(x).$$

Beispiel: $\exists x \in \mathbb{Q} : x^2 = 2$ hat Negation:
$\forall x \in \mathbb{Q} : x^2 \neq 2$.
$\forall x \in \mathbb{Q} : x^2 > 0$ hat Negation
$\exists x \in \mathbb{Q} : x^2 \leq 0$ (nämlich $x = 0$).

Beachte: Falscher Umgang mit Quantoren und deren Negation ist eine häufige Fehlerquelle.

1.2. **Relationen, Abbildungen.**

Definition 1.5. *Sei M eine Menge. Eine* Relation*auf M ist eine Teilmenge $R \subseteq M \times M$. Für $(x_1, x_2) \in R$ schreibt man*

$$x_1 \underset{R}{\sim} x_2 \quad (\text{ oder nur } x_1 \sim x_2)$$

Die Relation heißt symmetrisch*, wenn*

$$x_1 \underset{R}{\sim} x_2 \Rightarrow x_2 \underset{R}{\sim} x_1$$

gilt. Sie heißt reflexiv*, wenn $x \underset{R}{\sim} x \ \forall x \in M$ gilt. Sie heißt* transitiv*, wenn*

$$x \underset{R}{\sim} y \text{ und } y \underset{R}{\sim} z \Rightarrow x \underset{R}{\sim} z$$

gilt.

Beispiel:

- $\Delta = \Delta_M = \{(x, x) \mid x \in M\}$ liefert die Gleichheit:

  $$x_1 \underset{\Delta}{\sim} x_2 \Leftrightarrow x_1 = x_2$$

  Diese Relation ist reflexiv, symmetrisch, transitiv.
- $T = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \leq x_2\}$ ist die Relation

  $$x \underset{T}{\sim} y \Leftrightarrow x \leq y.$$

  Sie ist reflexiv und transitiv, aber nicht symmetrisch.
- $x \underset{R}{\sim} y \Leftrightarrow x$ ist älter und größer als y und ist eine Relation auf der Menge der Anwesenden.
- $x \underset{R}{\sim} y$: x ist Kusin/Kusine ersten Grades von y.

Definition 1.6. *Die Relation $\sim$ auf M heißt* Äquivalenzrelation*, wenn sie reflexiv, symmetrisch und transitiv ist.*

Satz 1.7. *Sei M eine Menge mit der Äquivalenzrelation $\sim$. Für $x \in M$ sei*

$$[x]_\sim := [x] := \{y \in M \mid x \sim y\}$$

die Äquivalenzklasse*von x unter $\sim$. Dann gilt:*

a) *Für $x, y \in M$ gilt $[x] = [y]$ oder $[x] \cap [y] = \emptyset$. (Die Äquivalenzklassen sind paarweise disjunkt).*
b) $M = \bigcup_{x \in M} [x] = \bigcup_{[x] \subseteq M} [x]$

Beweis. Wegen der Reflexivität gilt b).
Zu a): Ist $[x] \neq [y]$, so gibt es $z \in M$ mit $x \sim z$, $y \not\sim z$.
Ist dann $[x] \cap [y] \neq \emptyset$, so sei w im Durchschnitt: $x \sim w$, $y \sim w$.
Also: $z \sim x$ und $x \sim w \Rightarrow z \sim w \Rightarrow w \sim z$
$y \sim w$, $w \sim z \Rightarrow y \sim z$ Widerspruch. □

Definition 1.8. *Sei M eine Menge mit der Äquivalenzrelation $\sim$.*

a) *Ist A eine Äquivalenzklasse von $\sim$ in M, so heißt jedes $x \in A$ ein Repräsentant der Klasse A.*

b) *Eine Teilmenge $X \subseteq M$ von M heißt ein (vollständiges) Repräsentantensystem von M bezüglich $\sim$ wenn gilt:*
Für jede Äquivalenzklasse A von $\sim$ in M gibt es genau ein $x \in X$ mit $x \in A$.

Beispiel:

a) (Konstruktion von $\mathbb{Z}$): Auf $\mathbb{N}_0 \times \mathbb{N}_0$ ($\mathbb{N}_0 = \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$) definieren wir eine Äquivalenzrelation $\sim$ durch

$$(a, b) \sim (a', b') \Leftrightarrow a + b' = a' + b.$$

Nachrechnen: Das ist eine Äquivalenzrelation.
Bezeichne die Äquivalenzklassen von (a, b) mit $[(a, b)]$. Es gilt (nachprüfen):
Ist $(a, b) \sim (a', b')$, $(c, d) \sim (c', d')$ (also $[(a, b)] = [(a', b')]$), so ist

$$\begin{aligned} (a + c, b + d) &\sim (a' + c', b' + d') \\ (ac + bd, ad + bc) &\sim (a'c' + b'd', a'd' + b'c') \end{aligned}$$

Wir können daher definieren:

$$\begin{aligned} [(a, b)] + [(c, d)] &:= [(a + c, b + d)], \\ [(a, b)] \cdot ([c, d]) &:= [ac + bd, ad + bc]. \end{aligned}$$

Man hat dann:

$$\begin{aligned} [(a, b)] + [(b, a)] = [(a + b), (a + b)] &= [(0, 0)] \\ [(a, b)] + [(0, 0)] &= [(a, b)] \end{aligned}$$

Wir schreiben: $[(b, a)] = -[(a, b)]$ und $[(0, 0)] = 0$.

Die Menge der Äquivalenzklassen bezeichnen wir mit $\mathbb{Z}$ und identifizieren für $n \in \mathbb{N}_0$ die Klasse von $(n, 0)$ mit n; dann haben wir mit den Klassen $[(n, 0)]$ eine Kopie von $\mathbb{N}_0$ in der Äquivalenzklassenmenge $\mathbb{Z}$, in der genau so gerechnet wird wie in $\mathbb{N}$:

Es gilt $[(n, 0)] + [(m, 0)] = [(n + m, 0)]$ und $[(n, 0)] \cdot [(m, 0)] = [(n \cdot m, 0)]$.

Ferner haben wir $[(n, m)] = [(n, 0)] - [(m, 0)]$ und insbesondere $[(0, n)] = -[(n, 0)]$. Nimmt man noch hinzu, dass man für $a, b \in \mathbb{N}_0, a \geq b$ hat:

$[(a, b)] = [(a - b, 0)], \quad , [(b, a)] = [(0, a - b)] = -[(a, b)]$,

so sieht man, dass man eine mengentheoretische Konstruktion der Menge $\mathbb{Z}$ der ganzen Zahlen gegeben hat.

b) (Restklassen modulo m): Auf $\mathbb{Z}$ wird für $m \in \mathbb{N}$ eine Relation $\equiv \bmod m$ definiert durch: $a \equiv b \bmod m \Leftrightarrow b - a$ ist durch m teilbar $\Leftrightarrow a$ und b lassen bei Division mit Rest durch m den gleichen Rest. Die Relation ist offenbar eine Äquivalenzrelation. Bezeichne die Klasse von a mit $[a]_m$. Dann hat man etwa für

$m = 3$ die Klassen

$$[0]_3 = [3]_3 = [-3]_3 = ...$$
$$[1]_3 = [-2]_3 = [4]_3 = ...$$
$$[2]_3 = [-1]_3 = [5]_3 = ...$$

die Menge der Äquivalenzklassen wird mit $\mathbb{Z}/m\mathbb{Z}$ bezeichnet. Es gilt (Übung): $[a]_m = [a']_m$ und $[b]_m = [b']_m \Rightarrow [a+b]_m = [a'+b']_m$, $[a \cdot b]_m = [a' \cdot b']_m$.
Man kann daher auf der Menge $\mathbb{Z}/m\mathbb{Z}$ eine Addition und eine Multiplikation definieren durch

$$[a]_m \underset{m}{+} [b]_m := [a+b]_m$$
$$[a]_m \underset{m}{\cdot} [b]_m := [a \cdot b]_m.$$

Man rechnet leicht nach, dass für diese Rechenoperationen die vom Zahlenrechnen gewohnten Assoziativ-, Kommutativ- und Distributivgesetze gelten und die Klasse $[0]_m$ neutrales Element bezüglich der Addition, die Klasse $[1]_m$ neutrales Element bezüglich der Multiplikation ist.

Steht der Modul m fest, so schreibt man auch $\bar{a}$ für die Äquivalenzklasse $[a]_m$ von a.

Speziell für $m = 2$ erhalten wir $\mathbb{F}_2 := \mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\}$ mit $\bar{0} + \bar{0} = \bar{0}, \bar{0} + \bar{1} = \bar{1} + \bar{0} = \bar{1}, \bar{1} + \bar{1} = \bar{0}, \bar{0} \cdot \bar{0} = \bar{0}, \bar{0} \cdot \bar{1} = \bar{1} \cdot \bar{0} = \bar{0}, \bar{1} \cdot \bar{1} = \bar{1}$. Wir kommen auf dieses Beispiel im nächsten Abschnitt nach der allgemeinen Definition von Körpern und Ringen zurück und sehen dort unter anderem, dass $\mathbb{F}_2$ ein Körper ist.

Das Studium von Abbildungen zwischen verschiedenen Mengen (in der Schule meistens "Zuordnungen" genannt) spielt in allen Zweigen der Mathematik eine zentrale Rolle. Wir stellen die Definitionen zusammen und verschaffen uns eine Übersicht über die wichtigsten Eigenschaften von Abbildungen und deren Zusammenhänge.

Definition 1.9. *Seien X, Y Mengen. Eine* Abbildung *(oder* Funktion*) $f : X \longrightarrow Y$ ordnet jedem $x \in X$ genau ein $y \in Y$ zu. Man schreibt: $x \mapsto y = f(x)$.*

$f : X \longrightarrow Y$ heißt:

- *injektiv, wenn gilt: Ist $f(x_1) = f(x_2)$, so ist $x_1 = x_2$*
- *surjektiv, wenn gilt: Für jedes $y \in Y$ gibt es (wenigstens) ein $x \in X$ mit $f(x) = y$*
- *bijektiv, wenn f injektiv und surjektiv ist (äquivalent: Für jedes $y \in Y$ gibt es genau ein $x \in X$ mit $y = f(x)$).*

Ist $f : X \longrightarrow Y$ eine Abbildung, so heißt $g : Y \longrightarrow X$ Umkehrabbildung von f, wenn gilt

- *Für jedes $x \in X$ ist $g(f(x)) = x$*

- *Für jedes $y \in Y$ ist $f(g(y)) = y$.*

Man schreibt dann: $g = f^{-1}$
Ist $f : X \longrightarrow Y$ eine Abbildung, so ist

$$f(X) := \mathrm{Im}(f) := \{y \in Y \mid \text{es gibt } x \in X \text{ mit } y = f(x)\} \tag{1.1}$$

das Bild von f. Analog schreibt man für jede Teilmenge $M \subseteq X$:

$$f(M) := \{y \in Y \mid \text{es gibt } x \in M \text{ mit } y = f(x)\} \tag{1.2}$$

und nennt diese Menge das Bild von M unter f.
Ist $N \subseteq Y$ eine Teilmenge von Y, so schreibt man

$$f^{-1}(N) := \{x \in X \mid f(x) \in N\} \tag{1.3}$$

und nennt diese Menge das Urbild von N unter f (diese Bezeichnung ist etwas irritierend, denn eine Umkehrabbildung f^{-1} muss nicht existieren. Sie ist aber dennoch üblich).
Sind $f : X \longrightarrow Y$ und $g : Y \longrightarrow Z$ Abbildungen, so ist die Abbildung $g \circ f : X \longrightarrow Z$ definiert durch $(g \circ f)(x) = g(f(x))$ für alle $x \in X$ (Komposition von Abbildungen, Hintereinanderausführung).
Ist $f : X \longrightarrow Y$ Abbildung, $M \subseteq X$ eine Teilmenge, so wird die Abbildung $f|_M : M \longrightarrow Y$ definiert durch $f|_M(x) = f(x)$ für $x \in M$ (Einschränkung oder Restriktion von f auf M.

Bemerkung.
- Zwei Abbildungen $f, g : X \longrightarrow Y$ sind gleich, wenn $f(x) = g(x)$ für alle $x \in X$ gilt. Es ist also unerheblich, ob sie eventuell durch verschiedene Vorschriften gegeben sind, die am Ende die gleiche Wirkung haben.
- Man kann auch die folgende mengentheoretische Definition einer Abbildung geben: Eine Abbildung $f : X \longrightarrow Y$ wird gegeben durch eine Teilmenge $\Gamma_f \subseteq X \times Y$ mit:

  Für jedes $x \in X$ gibt es genau ein $y \in Y$, so dass $(x, y) \in \Gamma_f$ gilt.

  Für $(x, y) \in \Gamma_f$ schreibt man $y = f(x)$, $f : x \longmapsto y$. Die Menge $\Gamma_f \subseteq X \times Y$ heißt auch der (mengentheoretische) Graph von f.
- Nicht jede Abbildung besitzt eine Umkehrabbildung (siehe nächstes Lemma).
- Man sieht: f ist genau dann surjektiv, wenn $f(X) = Y$ gilt.
- Ist f umkehrbar mit Umkehrabbildung $g = f^{-1}$, so ist das Urbild einer Teilmenge $N \subseteq Y$ von Y unter f das Bild von N unter $g = f^{-1}$, die beiden denkbaren Bedeutungen von $f^{-1}(N)$ stimmen also überein. Ist f nicht umkehrbar, so hat das Symbol f^{-1} für sich genommen keine Bedeutung, so dass „Urbild von N unter f“ die einzige mögliche Bedeutung von $f^{-1}(N)$ ist.

  Das Symbol f^{-1} hat im Übrigen gar nichts mit dem Kehrwert („eins durch f“) zu tun, der ja auch in beliebigen Mengen kein sinnvoller Begriff ist.

- Bei der Komposition $g \circ f$ zweier Abbildungen wird zuerst f und dann g angewendet, also quasi von Rechts nach Links gelesen.

Beispiel:

- Mit $X = Y = \{1, 2, 3\}$ wird durch $f(1) = 3, f(2) = 1, f(1) = 1$ keine Abbildung gegeben: $f(1)$ ist nicht eindeutig definiert und $f(3)$ ist überhaupt nicht definiert. Dagegen wird durch $f(1) = 2, f(2) = 1, f(3) = 1$ eine Abbildung gegeben, die allerdings weder injektiv noch surjektiv ist.
- Mit $X = Y = \mathbb{R}$ wird durch $f(x) = \sqrt{x}$ keine Abbildung gegeben, weil $f(x)$ für $x < 0$ nicht definiert ist (jedenfalls nicht als reelle Zahl). Ersetzt man hier X durch $X' = \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} \mid x \geq 0\}$ (und legt fest, dass die positive Wurzel genommen werden soll), so wird durch die gleiche Vorschrift eine Abbildung $f : X' \longrightarrow \mathbb{R}$ gegeben, die injektiv, aber nicht surjektiv ist. Ersetzt man auch noch Y durch $Y' = \mathbb{R}_{\geq 0}$, so erhält man eine bijektive Abbildung $X' \longrightarrow Y'$.
- Sei $X = Y = \mathbb{R}$ sowie $f : X \longrightarrow Y$ durch $f(x) = x + 1$ und $g : X \longrightarrow Y$ durch $g(x) = x^2$ gegeben.

  Dann ist $(g \circ f)(x) = (x + 1)^2 = x^2 + 2x + 1$ für alle $x \in X$, dagegen ist $(f \circ g)(x) = x^2 + 1$ für alle $x \in X$.
- Sei X eine Menge. Die identische Abbildung $Id_X : X \longrightarrow X$ von X ist durch $Id_X(x) = x$ für alle $x \in X$ definiert. Mit dieser Notation wird die Umkehrabbildung g einer Abbildung $f : X \longrightarrow Y$ durch
$$g \circ f = Id_X, \quad f \circ g = Id_Y$$
  charakterisiert.

Lemma 1.10. (i) *Sind $f : X \longrightarrow Y$, $g : Y \longrightarrow Z$, $h : Z \longrightarrow W$ Abbildungen, so ist $h \circ (g \circ f) = (h \circ g) \circ f$ (die Komposition von Abbildungen ist assoziativ, d. h., man darf Klammern verschieben).*

(ii) *$f : X \longrightarrow Y$ besitzt genau dann eine Umkehrabbildung, wenn f bijektiv ist; die Umkehrabbildung ist in diesem Fall eindeutig bestimmt und ebenfalls bijektiv.*

(iii) *Sind $f : X \longrightarrow Y$, $g : Y \longrightarrow Z$ bijektive Abbildungen, so ist auch $g \circ f$ bijektiv, und es gilt $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.*

Beweis. (i) Für jedes $x \in X$ ist

$$\begin{aligned} (h \circ (g \circ f))(x) &= h((g \circ f)(x)) \\ &= h(g(f(x)) \\ &= (h \circ g)(f(x)) \\ &= ((h \circ g) \circ f)(x). \end{aligned} \tag{1.4}$$

Da die beiden Abbildungen $(h \circ (g \circ f)$ und $(h \circ g) \circ f$ in allen $x \in X$ den gleichen Wert annehmen, sind sie gleich.

(ii) Ist f bijektiv, so definiere man $g : Y \longrightarrow X$ wie folgt: Für $y \in Y$ gibt es nach Definition genau ein $x \in X$, für das $f(x) = y$ gilt. Dann setze man $g(y) := x$. Dadurch wird jedem $y \in Y$ ein eindeutig bestimmtes $x \in X$ zugeordnet, man hat also eine Abbildung $g : Y \longrightarrow X$ definiert. Dass diese die behauptete Eigenschaft hat, ist jetzt klar.

Hat umgekehrt f eine Umkehrabbildung g, so müssen wir zeigen, dass f injektiv und surjektiv ist. Sind zunächst $x_1, x_2 \in X$ mit $f(x_1) = f(x_2)$, so ist $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$, f ist also injektiv. Ist $y \in Y$ beliebig, so ist $y = f(g(y))$ Bild des Elements $g(y)$ von X unter f, also ist f auch surjektiv.

Die Umkehrabbildung g ist eindeutig bestimmt, denn für $y \in Y$ gibt es genau ein $x \in X$ mit $f(x) = y$, wegen $(g \circ f)(x) = x$ ist dann zwangsläufig $g(y) = x$. Dass die Umkehrabbildung ebenfalls bijektiv ist, rechne man als Übung nach.

(iii) Sind $x_1, x_2 \in X$ mit $(g \circ f)(x_1) = (g \circ f)(x_2)$, so ist $g(f(x_1) = g(f(x_2)$, also gilt $f(x_1) = f(x_2)$ wegen der vorausgesetzten Injektivität von g. Da auch f injektiv ist, folgt $x_1 = x_2$, und man sieht, dass $g \circ f$ injektiv ist.

Ist $z \in Z$, so gibt es ein $y \in Y$ mit $g(y) = z$, da g surjektiv ist. Zu diesem $y \in Y$ gibt es ein $x \in X$ mit $f(x) = y$, da auch f nach Voraussetzung surjektiv ist. Nimmt man beide Gleichungen zusammen, so erhält man $(g \circ f)(x) = z$. Das beliebig angenommene $z \in Z$ ist also Bild des Elements $x \in X$ unter $g \circ f$, und damit folgt die Surjektivität von $g \circ f$. Insgesamt sieht man, dass $g \circ f$ bijektiv ist.

Dass $f^{-1} \circ g^{-1}$ die Umkehrabbildung von $g \circ f$ ist, rechnet man mit Hilfe des Assoziativgesetzes für die Komposition von Abbildungen nach.

□

1.3. **Natürliche Zahlen und Induktion.** Wir müssen uns vergewissern, was $\mathbb{N}$ ist, bzw. welche Eigenschaften von $\mathbb{N}$ wir als grundlegend ansehen und für die weiteren logischen Schlüsse verwenden wollen.

Eine axiomatische Charakterisierung von $\mathbb{N}$ wird durch das Axiomensystem von Peano (1858–1932) (Dedekind 1831–1916) gegeben.

Definition 1.11. *Die Menge der natürlichen Zahlen ist eine Menge* $\mathbb{N}$ *mit folgenden Eigenschaften:*

P0: *Es gibt eine Abbildung* $s : \mathbb{N} \longrightarrow \mathbb{N}$*; für* $n \in \mathbb{N}$ *heißt* $s(n)$ *der Nachfolger (successor) von* n.

P1: *Es gibt eine Element* $1 \in \mathbb{N}$.

P2: s *ist injektiv: Ist* $n \neq m$*, so ist* $s(n) \neq s(m)$.

P3: *Induktionsaxiom: Enthält eine Menge* M *von natürlichen Zahlen die Zahl* 1 *und mit jeder natürlichen Zahl* $n \in M$ *auch deren Nachfolger* $n' = s(n)$*, so ist* $M = \mathbb{N}$.

Wir fügen zu $\mathbb{N}$ *noch ein Symbol* 0 *mit* $s(0) = 1$ *hinzu und erhalten* $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Bemerkung. a) $\mathbb{N}$ *kann man mit Hilfe des Unendlichkeitsaxioms der Mengenlehre rein mengentheoretisch konstruieren:*
$1 = \{\emptyset\}$, $s(1) = 1 \cup \{1\} = \{\emptyset, \{\emptyset\}\}$, $s(n) = n \cup \{n\}$

b) *Alle weiteren aus der Schule bekannten Eigenschaften von* $\mathbb{N}$ *kann man aus diesen Axiomen herleiten. Z.B. definiert man die Addition durch:*

$$\begin{aligned} m+1 &:= s(m) \\ m+s(n) &= s(m+n) \quad \textit{induktiv (rekursiv).} \end{aligned}$$

Man kann dann mit $2 := s(1)$ *beweisen, dass (mit* $3 = s(2)$, $4 = s(3)$*)* $2+2=4$ *gilt.*
(Also: $s(1) + s(1) = s(s(s(1)))$ *).*
Ebenfalls beweisen kann man, dass für die Addition natürlicher Zahlen das Assoziativ- und das Kommutativgesetz gelten.
Multiplikation:

$$\begin{aligned} m \cdot 1 &= m \\ m \cdot s(n) &= m \cdot n + m. \end{aligned}$$

c) *Das Induktionsaxiom ist Grundlage eines wichtigen Beweisverfahrens, des Beweises durch vollständige Induktion.*

Satz 1.12. *Für jedes* $n \in \mathbb{N}$ *sei eine Aussage* $P(n)$ *gegeben. Es gelte:*

a) $P(1)$ *ist wahr. Induktionsanfang (Induktionsannahme).*
b) *Ist* $n \in \mathbb{N}$ *und* $P(n)$ *wahr, so ist auch* $P(n+1)$ *wahr. Induktionsschritt.*
Dann ist $P(n)$ *für alle* $n \in \mathbb{N}$ *wahr.*

Beweis. Sei $M := \{n \in \mathbb{N} \mid P(n) \text{ ist wahr }\}$. M hat die im Induktionsaxiom genannten Eigenschaften, also ist $M = \mathbb{N}$, d.h., $P(n)$ ist wahr für alle $n \in \mathbb{N}$. □

Variante: $P(j)$ sei wahr für alle $j \leq n$.

Beispiel:

a) Für alle $n \in \mathbb{N}$ ist

$$\sum_{j=1}^{n} j = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Beweis. $P(1)$ sagt $1 = 1$
Sei $P(n)$ wahr, also

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}.$$

Dann ist

$$\begin{aligned}\sum_{j=1}^{n+1} j &= n+1+\sum_{j=1}^{n} j = n+1+\frac{n(n+1)}{2}\\ &= \frac{2n+2+n^2+n}{2} = \frac{n^2+3n+2}{2}\\ &= \frac{(n+1)(n+2)}{2},\end{aligned}$$

also gilt $P(n+1)$.

b) Alle Hörer dieser Vorlesung haben die gleiche Augenfarbe.

Beweis. Ich formuliere die Aussage um. $P(n)$ ist die Aussage: Je n Hörer haben die gleiche Augenfarbe. In dieser Form zeige ich dann durch vollständige Induktion, dass $P(n)$ für alle $n \in \mathbb{N}$ wahr ist.

Induktionsanfang: $P(1)$ ist offensichtlich wahr.
Induktionsannahme (oder -Voraussetzung): Sei $n \in \mathbb{N}$, $P(j)$ sei wahr für alle $j \leq n$.
Ich teile eine Menge $\{H_1, \ldots, H_{n+1}\}$ von $n+1$ Hörern auf als $\{H_1, H_2, \ldots, H_n\} \cup \{H_2, \ldots, H_n, H_{n+1}\}$. Jede dieser beiden Mengen hat n Elemente, also haben die Hörer in ihr nach Induktionsannahme die gleiche Augenfarbe. H_n gehört zu beiden Mengen, also haben sowohl die Hörer in der ersten Menge als auch die in der zweiten die gleiche Augenfarbe wie H_n, also haben alle Hörer diese Augenfarbe.

Definition 1.13. *In $\mathbb{N}_0$ werden Addition und Multiplikation definiert durch:*

$$\begin{aligned}0+n &= n \qquad \forall n\\ 0\cdot n &= 0\\ s(m)\cdot n &= m\cdot n+n\end{aligned}$$

Satz 1.14. *Für alle $p, q, r \in \mathbb{N}_0$ gilt:*

$$\begin{aligned}(p+q)+r &= p+(q+r)\\ p+q &= q+p\\ (p+q)r &= pr+qr\\ r(p+q) &= rp+rq\end{aligned}$$

Beweis. Übung. □

2. Körper und lineare Gleichungssysteme

Definition 2.1. *Sei K eine Menge mit zwei* Verknüpfungen *genannten Abbildungen $+ : K \times K \to K$ (geschrieben als $(a,b) \mapsto a+b$) und $\cdot : K \times K \to K$ (geschrieben als $(a,b) \mapsto a \cdot b =: ab$) sowie zwei ausgezeichneten Elementen $0 = 0_K$ und $1 = 1_K \neq 0_K$. Die Menge K mit diesen Verknüpfungen und den beiden ausgezeichneten Elementen heißt ein* Körper*, wenn gilt:*

- A1 *(*Assoziativgesetz *der Addition): Für alle $a, b, c \in K$ gilt $(a+b)+c = a+(b+c)$.*
- A2 *(*Kommutativgesetz *der Addition): Für alle $a, b \in K$ gilt $a+b = b+a$.*
- A3 *(*Neutrales Element *der Addition): Für alle $a \in K$ gilt $a+0 = 0+a = a$.*
- A4 *(*Inverses Element *der Addition): Für alle $a \in K$ gibt es genau ein mit $-a$ bezeichnetes Element von K, für das $a+(-a) = (-a)+a = 0$ gilt.*
- M1 *(Assoziativgesetz der Multiplikation): Für alle $a, b, c \in K$ gilt $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.*
- M2 *(Kommutativgesetz der Multiplikation): Für alle $a, b \in K$ gilt $a \cdot b = b \cdot a$.*
- M3 *(Neutrales Element der Multiplikation): Für alle $a \in K$ gilt $a \cdot 1 = 1 \cdot a = a$.*
- M4 *(Inverses Element der Multiplikation): Für alle $a \in K, a \neq 0$ gibt es genau ein mit a^{-1} bezeichnetes Element von K, für das $a \cdot a^{-1} = a^{-1} \cdot a = 1$ gilt.*
- D Distributivgesetz*:*

$$a \cdot (b+c) = a \cdot b + a \cdot c \qquad \textit{für alle } a, b, c \in K$$

Beispiele: $\mathbb{Q}$ und $\mathbb{R}$ sind Körper, ebenso $\mathbb{F}_2 := \mathbb{Z}/2\mathbb{Z}$. Die beiden erstgenannten Körper mit ihren Rechenregeln sind aus der Schule vertraut. Der Körper $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ ist etwas gewöhnungsbedürftig, in ihm gilt $1+1 = 0$. Er spielt bei Anwendungen in der Informatik eine wichtige Rolle, ist aber auch vom Standpunkt der reinen Algebra ein interessanter Gegenstand.
Allgemeiner gilt: $\mathbb{Z}/n\mathbb{Z}$ mit den Verknüpfungen $+$ und $\cdot$ ist genau dann ein Körper, wenn $n = p$ eine Primzahl ist, er wird dann mit $\mathbb{F}_p$ bezeichnet (Übung). In diesem Fall ist

$$p \cdot 1 := \underbrace{1 + \cdots + 1}_{p-\text{ mal}} = 0,$$

und p ist die kleinste aller natürlichen Zahlen r, für die $r \cdot 1 = 0$ gilt. Man sagt, der Körper $\mathbb{Z}/p\mathbb{Z}$ habe *Charakteristik* p.

Bemerkung.

a) In einem Körper ist stets $1 \neq 0$, also haben Körper wenigstens 2 Elemente.
b) Lässt man das Axiom M4 fort, das die Existenz von multiplikativen Inversen fordert, so erhält man die Definition eines kommutativen Ringes mit Einselement $1 \neq 0$. Zum Beispiel hat die Menge $\mathbb{Z}$ der ganzen Zahlen diese Eigenschaften.

   Lässt man obendrein das Kommutativgesetz M2 für die Multiplikation fort und fordert dafür in D zusätzlich das zweite Distributivgesetz $(a+b)c = ac + bc$ für alle $a, b, c \in K$, so hat man einen Ring mit Einselement $1 \neq 0$ definiert. Beispiele dafür sehen wir später.
c) Verlangt man zwar M4, aber für die Multiplikation nicht die Gültigkeit des Kommutativgesetzes M2 und in dafür in D zusätzlich das zweite Distributivgesetz $(a+b) \cdot c = a \cdot c + b \cdot c$ für alle $a, b, c \in K$, so erhält man die Definition eines *Schiefkörpers*. Beispiele dafür sehen wir später.
d) Wir werden bei der nachfolgenden Behandlung linearer Gleichungssysteme sehen, dass wir für Koeffizienten und Variable des Gleichungssystems nur die in Definition 2.1 festgelegten Eigenschaften benötigen. Diese Eigenschaften heißen die *Körperaxiome*. Die abstrakte axiomatische Vorgehensweise hat den Vorteil, dass alle Sätze, die wir für lineare Gleichungssysteme herleiten, automatisch für Gleichungssysteme über einem beliebigen Körper gelten, wir also z. B. die in vieler Hinsicht sehr verschiedenen Körper $\mathbb{R}$ und $\mathbb{F}_2$ nicht getrennt behandeln müssen.

Lemma 2.2. *In jedem Körper K gilt für alle $a, b, c, d \in K$:*

a) *(2. Distributivgesetz):* $(a+b) \cdot c = a \cdot c + b \cdot c$.
b) $-(a+b) = (-a) + (-b)$.
c) $0 \cdot a = a \cdot 0 = 0$.
d) $(-1) \cdot a = -a$.
e) $(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$.
f) *Ist $ab = 0$ so ist $a = 0$ oder $b = 0$.*
g) *Ist $a \neq 0 \neq b$, so ist $(a \cdot b)^{-1} = a^{-1} \cdot b^{-1}$.*
h) *Ist $b \neq 0 \neq d$ und schreibt man $\frac{a}{b} := a \cdot b^{-1}$, so ist*
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}, \quad \frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + b \cdot c}{b \cdot d}.$$

Beweis. Übung. □

Definition 2.3. *Das System von Gleichungen*

$$\begin{array}{lcl} a_{11}x_1 & + \ldots + & a_{1n}x_n = b_1 \\ & \vdots & \\ a_{p1}x_1 & + \ldots + & a_{pn}x_n = b_p \end{array} \tag{2.1}$$

mit Koeffizienten $a_{11}, \ldots, a_{pn}$ und $b_1, \ldots b_p$ in dem Körper K heißt lineares Gleichungssystem *(über dem Körper K) in den n Unbekannten*

$x_1, \ldots, x_n$. *Es heißt* homogen, *wenn* $b_1 = \ldots = b_p = 0$ *gilt,* inhomogen *sonst.*
Ein Vektor

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n,$$

dessen Komponenten $x_1, \ldots, x_n$ *die Gleichungen lösen, heißt ein* Lösungsvektor *(oder einfach eine Lösung) für das Gleichungssystem.*
Die pn *Koeffizienten des Gleichungssystems werden in der* $(p \times n)$-Matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{p1} & \ldots & a_{pn} \end{pmatrix}$$

zusammengefaßt; diese hat die p *Zeilen*

$${}^t\mathbf{z}_1 = (a_{11}, \ldots, a_{1n}), \ldots, \quad {}^t\mathbf{z}_p = (a_{p1}, \ldots, a_{pn})$$

und die n *Spalten*

$$\mathbf{s}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{p1} \end{pmatrix}, \ldots, \quad \mathbf{s}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{pn} \end{pmatrix}$$

Die Menge der $(p \times n)$*-Matrizen mit Einträgen aus* K *heißt* $M(p \times n, K)$ *(oder* $\mathrm{Mat}_{p,n}(K), K^{p,n}$*).*
Für das System (2.1) schreiben wir auch abkürzend

$$A\mathbf{x} = \mathbf{b}.$$

Das System, das abgekürzt als $A\mathbf{x} = \mathbf{0}$ *geschrieben wird, heißt das zu* $A\mathbf{x} = \mathbf{b}$ *gehörende homogene Gleichungssystem.*

Aus der Schulmathematik ist die geometrische Interpretation eines solchen linearen Gleichungssystems in den Fällen $n = 2$ und $n = 3$ bekannt:
Ist $n = 2$, so besteht die Lösungsmenge der i-ten Gleichung in (2.1) aus den Ortsvektoren der Punkte einer Geraden g_i (falls nicht alle a_{ij} Null sind, die Gleichung also weder trivial noch widersprüchlich ist). Die Lösungsmenge des ganzen Gleichungssystems besteht dann aus den Ortsvektoren derjenigen Punkte, die auf allen Geraden $g_i (1 \leq i \leq p)$ liegen, sie besteht aus den Punkten auf einer Geraden (falls alle g_i gleich sind), einem Punkt oder ist leer (falls die Geraden keinen gemeinsamen Schnittpunkt haben).
Ähnlich ist die Situation für $n = 3$: Die Lösungsmenge der i-ten Gleichung in (2.1) besteht jetzt im nichttrivialen Fall aus den Ortsvektoren der Punkte einer Ebene E_i. Die Lösungsmenge des ganzen Gleichungssystems besteht dann aus den Ortsvektoren derjenigen Punkte, die auf allen Ebenen $E_i (1 \leq i \leq p)$ liegen (in denen sich also die Ebenen

schneiden), sie besteht aus den Punkten auf einer Ebene (falls alle E_i gleich sind), einer Geraden, einem Punkt oder ist leer (falls die Ebenen keinen gemeinsamen Schnittpunkt haben).
In beiden Fällen gehört der Ursprung $\mathbf{0}$ genau dann zur Lösungsmenge, wenn bei allen Gleichungen die rechte Seite 0 ist, das Gleichungssystem also homogen ist.
Wir werden in diesem Abschnitt sehen, dass sich lineare Gleichungssysteme in mehr als 3 Variablen, bei denen es keine direkte geometrische Interpretation gibt, im Prinzip ähnlich verhalten.
Um die Lösungen eines linearen Gleichungssystems explizit zu bestimmen verwendet man in der Regel ein „Gauß - Elimination“ (Carl Friedrich Gauß, 1777-1855) genanntes algorithmisches Verfahren, das aber im Prinzip schon lange vor Gauß in China bekannt war (unter dem Namen „fang cheng“ kommt es in den „Neun Kapiteln über die Kunst der Mathematik“ aus der Zeit der Han-Dynastie, vermutlich im ersten Jahrhundert vor Christus vor); es verallgemeinert und formalisiert das von Gleichungssystemen in zwei oder drei Variablen vertraute Verfahren, durch geschicktes Addieren von Gleichungen und Multiplizieren von Gleichungen mit Zahlen $\neq 0$ das Gleichungssystem auf eine Gleichung in einer Unbekannten zu reduzieren, die man dann leicht lösen kann.
Bevor wir damit beginnen, führen wir für einen Körper K noch Rechenregeln für das Rechnen in K^n ein:

Definition 2.4. *Sei K ein Körper.*
Für $c \in K, \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n, \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in K^n$ *sei*

$$\mathbf{x} + \mathbf{y} := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, c\mathbf{x} := \begin{pmatrix} cx_1 \\ \vdots \\ cx_n \end{pmatrix}.$$

Lemma 2.5. *Für die Addition von Vektoren des K^n gelten das Assoziativ- und das Kommutativgesetz, der Nullvektor* $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ *ist neutrales Element der Addition, und zum Vektor* $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ *ist der Vektor* $-\mathbf{x} = \begin{pmatrix} -x_1 \\ \vdots \\ -x_n \end{pmatrix}$ *additiv invers, also* $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ *für jedes* $\mathbf{x} \in K^n$.
Ferner gilt $c(\mathbf{x}+\mathbf{y}) = c\mathbf{x}+c\mathbf{y}$ *und* $b(c\mathbf{x}) = (bc)\mathbf{x}$ *für alle* $b, c \in K, \mathbf{x}, \mathbf{y} \in K^n$.

Beweis. Klar. □

Zunächst schreiben wir eine Form eines Gleichungssystems auf, in der es (wie wir gleich sehen werden) besonders leicht zu lösen ist:

Definition 2.6. *Sei $A \in M(p \times n, K)$. Man sagt, A (bzw. das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$) habe* Zeilenstufenform, *wenn gilt:*

i) *Es gibt $0 \leq r \leq p$, so daß gilt:*
Ist $i > r$, so ist $a_{ij} = 0$ für alle j,
ist $1 \leq i \leq r$, so gibt es ein j mit $a_{ij} \neq 0$.

ii) *Für $1 \leq i \leq r$ sei $s(i) := \min\{j \mid a_{ij} \neq 0\}$ die Nummer der ersten Spalte von links, die in der i-ten Zeile ein Element $\neq 0$ enthält. Mit dieser Bezeichnung gilt:*

$$s(1) < s(2) < \ldots < s(r).$$

Die Elemente $a_{i,s(i)}$ heißen die Pivotelemente *der Matrix in Zeilenstufenform. Bildlich:*

$$\begin{pmatrix}
0 & \ldots & 0 & |\underline{a_{1,s(1)}} & \ldots & \ldots & \ldots & \ldots & \ldots & a_{1,s(r)+1} & \ldots & a_{1,n} \\
0 & \ldots & 0 & 0 & \ldots & 0 & |\underline{a_{2,s(2)}} & \ldots & \ldots & a_{2,s(r)+1} & \ldots & a_{2,n} \\
\vdots & & & \ddots & & & & & & & & \\
0 & \ldots & \ldots & & & \ldots & \ldots & 0 & |\underline{a_{r,s(r)}} & a_{r,s(r)+1} & \ldots & a_{r,n} \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & & & & & & & & & & & \vdots \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0
\end{pmatrix}$$

Wir sagen, A habe reduzierte *Zeilenstufenform, wenn überdies gilt:*

iii) $a_{i,s(i)} = 1$ *für* $1 \leq i \leq r$

iv) $a_{k,s(i)} = 0$ *für* $1 \leq k < i \leq r$

Die Matrix hat dann die Form:

$$\begin{pmatrix}
0 & \ldots & 0 & |\underline{1} & \ldots & \ldots & 0 & \ldots & 0 & a_{1,s(r)+1} & \ldots & a_{1,n} \\
0 & \ldots & 0 & 0 & \ldots & 0 & |\underline{1} & \ldots & 0 & a_{2,s(r)+1} & \ldots & a_{2,n} \\
\vdots & & & \ddots & & & & & & & & \\
0 & \ldots & \ldots & & & \ldots & \ldots & 0 & |\underline{1} & a_{r,s(r)+1} & \ldots & a_{r,n} \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\
\vdots & & & & & & & & & & & \vdots \\
0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0
\end{pmatrix}$$

Ist ein Gleichungssystem bzw. seine Matrix in Zeilenstufenform, so lassen sich seine Lösungen leicht bestimmen.

Satz 2.7. a) *Ist das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ mit $A \in M(p \times n, K)$ in Zeilenstufenform mit $r = r(A)$, so hat es keine Lösungen, wenn nicht $b_{r+1} = \ldots = b_p = 0$ gilt.*

b) *Ist das lineare Gleichungssystem* $A\mathbf{x} = \mathbf{b}$ *mit* $A \in M(p \times n, K)$ *in reduzierter Zeilenstufenform lösbar (also* $b_{r+1} = \cdots = b_p = 0$*) und zusätzlich* $s(i) = i$ *für* $1 \leq i \leq r$*, so sind mit*

$$\mathbf{l}_1 = \begin{pmatrix} -a_{1,r+1} \\ \vdots \\ -a_{r,r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad \mathbf{l}_{n-r} = \begin{pmatrix} -a_{1,n} \\ \vdots \\ -a_{r,n} \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

die Lösungen von $A\mathbf{x} = \mathbf{b}$ *genau die sämtlichen Vektoren*

$$\mathbf{x} = \begin{pmatrix} b_1 \\ \vdots \\ b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} + t_1\mathbf{l}_1 + \cdots + t_{n-r}\mathbf{l}_{n-r}$$

mit $t_1, \ldots, t_{n-r} \in K$*, und jede Lösung lässt sich in eindeutiger Weise so schreiben.*
Die Vektoren $\mathbf{l}_1, \ldots, \mathbf{l}_{n-r}$ *heißen ein System von Fundamentallösungen des homogenen Gleichungssystems* $A\mathbf{x} = \mathbf{0}$*. Ist hier* $r = n$*, so ist* $\{\mathbf{l}_1, \ldots, \mathbf{l}_{n-r}\} = \emptyset$ *und es gibt nur die eine Lösung*

$$\mathbf{x} = \begin{pmatrix} b_1 \\ \vdots \\ b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

($r > n$ *ist nicht möglich)*

Beweis. a) ist klar, denn für $i > r$ ist die i-te Gleichung

$$0 = b_i,$$

und das ist offensichtlich unlösbar für $b_i \neq 0$.
b) ist nicht viel schwerer: Geben wir beliebige Werte $t_1, \ldots, t_{n-r}$ der Variablen $x_{r+1}, \ldots, x_n$ vor, so wird die i-te Gleichung zu

$$x_i + \sum_{j=r+1}^{n} a_{ij} t_{j-r} = b_i,$$

eine Lösung des Gleichungssystems mit diesen Werten der Variablen $x_{r+1}, \ldots, x_n$ ist also gleich

$$\begin{pmatrix} b_1 \\ \vdots \\ b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} + t_1\mathbf{l}_1 + \cdots + t_{n-r}\mathbf{l}_{n-r},$$

und umgekehrt ist jeder Vektor dieses Typs eine Lösung des Gleichungssystems. □

Als nächsten Schritt stellen wir die Umformungen zusammen, die wir an einem Gleichungssystem vornehmen wollen:

Definition 2.8. *Sei $A = (a_{ij}) \in M(p \times n, K)$ eine Matrix mit Zeilen ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_p \in K^n$. Eine* elementare Zeilenumformung *von A ist gegeben durch:*

i) *Addition der mit $\lambda \in K$ multiplizierten j-ten Zeile zur i-ten Zeile (also ${}^t\mathbf{z}_i \longmapsto {}^t\mathbf{z}'_i = {}^t\mathbf{z}_i + \lambda {}^t\mathbf{z}_j$) für $i \neq j$.*
ii) *Multiplikation der i-ten Zeile mit $\lambda \in K^\times := K \setminus \{0\}$.*
iii) *Vertauschen von i-ter Zeile und j-ter Zeile.*

Eine elementare Zeilenumformung des linearen Gleichungssystems $A\mathbf{x} = \mathbf{b}$ ist eine elementare Zeilenumformung der erweiterten Matrix

$$(A|\mathbf{b}) := \begin{pmatrix} a_{11} & \ldots & a_{1n} & b_1 \\ \vdots & & & \\ a_{p1} & \ldots & a_{pn} & b_p \end{pmatrix}$$

zu der erweiterten Matrix $(A'|\mathbf{b}')$, gefolgt vom Übergang zum linearen Gleichungssystem $A'\mathbf{x} = \mathbf{b}'$.

Die beschriebenen Umformungen sind zum Lösen des Gleichungssystems brauchbar, weil sie die Lösungsmenge nicht verändern:

Lemma 2.9. *Geht das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ durch elementare Zeilenumformungen in das Gleichungssystem $A'\mathbf{x} = \mathbf{b}'$ über, so ist $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$ genau dann eine Lösung von $A\mathbf{x} = \mathbf{b}$, wenn es eine Lösung von $A'\mathbf{x} = \mathbf{b}'$ ist.*
Geht A' aus A durch Vertauschen der j-ten Spalte der Matrix mit der k-ten Spalte hervor, so entspricht das einer Vertauschung der Variablen x_j mit der Variablen x_k im Gleichungssystem, also einer Vertauschung der j-ten mit der k-ten Komponente in den Lösungsvektoren.

Beweis. Klar. □

Nützlich werden die Umformungen dadurch, dass sie es erlauben, ein beliebiges Gleichungssystem (in algorithmischer, also programmierbarer Weise) in die beschriebene einfache Gestalt (Zeilenstufenform) überzuführen:

Satz 2.10. *Jede Matrix $A \in M(p \times n, K)$ kann durch wiederholte elementare Zeilenumformungen in (reduzierte) Zeilenstufenform gebracht werden.*
Lässt man noch Spaltenvertauschungen zu, so lässt sich sogar $s(i) = i$ für $1 \leq i \leq r$ erreichen.

Beweis. Wir beweisen das durch vollständige Induktion nach der Anzahl p der Gleichungen (bzw. Zeilen der Matrix).

Induktionsanfang: Ist $p = 1$, so hat man nur eine Gleichung. Ist j_0 minimal mit $a_{1j_0} \neq 0$, so setze man $s(1) = j_0$ und multipliziere die (einzige) Gleichung mit $a_{1j_0}^{-1}$, das überführt sie in reduzierte Zeilenstufenform. Sind alle $a_{ij} = 0$, so ist die Gleichung bereits in reduzierter Zeilenstufenform.

Induktionsannahme: Sei $p > 1$ und die Behauptung bewiesen für Gleichungssysteme mit weniger als p Gleichungen.

Induktionsschritt: Sind alle $a_{ij} = 0$, so ist das Gleichungssystem bereits in reduzierter Zeilenstufenform. Andernfalls sei j_0 das Minimum aller j, für die ein Element $\neq 0$ in der j-ten Spalte der Matrix steht, sei i_0 so gewählt, dass $a_{i_0j_0} \neq 0$ gilt.
Wir subtrahieren jetzt für alle $i \neq i_0$ die mit $a_{ij_0} \cdot a_{i_0j_0}^{-1}$ multiplizierte i_0-te Gleichung (bzw. Zeile der Matrix) von der i-ten Gleichung (bzw. Zeile der Matrix). Danach stehen in allen Zeilen der Matrix außer der i_0-ten nur Nullen in der j_0-ten Spalte.
Anschließend multiplizieren wir die i_0-te Gleichung (bzw. Zeile der Matrix) mit $a_{i_0j_0}^{-1}$ und vertauschen dann die neue i_0-te Gleichung (bzw. Zeile der Matrix) (die jetzt mit $0 \cdots 0 \quad 1$ beginnt) mit der ersten Gleichung (bzw. Zeile der Matrix).
Jetzt haben wir eine Matrix A_1 erreicht, bei der links von der j_0-ten Spalte alle Einträge 0 sind und in der j_0-ten Spalte in der ersten Zeile 1, in allen anderen Zeilen 0 steht.
Sei A' die $(p-1) \times n$-Matrix, die man aus A_1 durch Streichen der ersten Zeile erhält. Nach Induktionsannahme kann man diese Matrix durch elementare Zeilenumformungen in reduzierte Zeilenstufenform bringen. Führt man diese Umformungen mit A_1 durch (unter Beachtung der Nummerierung der Zeilen: Die i-te Zeile von A' entspricht der $i+1$-ten Zeile von A_1), so ändert sich nichts an der ersten Zeile und an den ersten j_0 Spalten von A_1, und die resultierende $p \times n$ - Matrix $A_2 = (a_{ij}^{(2)})$ erfüllt alle Bedingungen der reduzierte Zeilenstufenform, außer eventuell der Bedingung $a_{1,s(i)}^{(2)} = 0$ für $2 \leq i \leq r$. Diese erreichen wir

dadurch, dass wir für $2 \leq i \leq r$ (in aufsteigender Reihenfolge) noch die mit $a^{(2)}_{1,s(i)}$ multiplizierte i-te Zeile von A_2 von der ersten Zeile abziehen. Da $a^{(2)}_{i,s(i)} = 1$ dabei jeweils das erste von 0 verschiedene Element in der i-ten Zeile ist, erreichen wir am Ende $a^{(2)}_{1,s(i)} = 0$ für $2 \leq i \leq r$ und damit die gewünschte reduzierte Zeilenstufenform.
Dass man durch abschließendes Ordnen der Spalten erreichen kann, dass $s(i) = i$ für $1 \leq i \leq r$ gilt, ist klar. □

Mit dem Lemma und dem Satz sind wir jetzt in der Lage, ein beliebiges lineares Gleichungssystem zu lösen bzw. als unlösbar nachzuweisen sowie gleichzeitig die Struktur der Lösungsmenge zu bestimmen. Wir fassen die Aussagen in einem Satz und drei Korollaren zusammen, deren Beweise sich aus den bisher bewiesenen Aussagen unmittelbar ergeben (man beachte, dass die Behauptungen über die Struktur der Lösungsmenge sich bei Umnummerieren der Variablen nicht ändern, wir also in der Zeilenstufenform die durch solches Umnummerieren erreichbare spezielle Gestalt mit $s(i) = i$ für alle i annehmen dürfen):

Satz 2.11. *Sei $A \in M(p \times n, K)$, $\mathbf{b} \in K^p$. Dann gilt: Entweder hat das Gleichungssystem $A\mathbf{x} = \mathbf{b}$ keine Lösungen, oder es gibt $r \in \mathbb{N}$ $(0 \leq r \leq p)$ und Vektoren $\mathbf{x}_0, \mathbf{l}_1, \ldots, \mathbf{l}_{n-r} \in K^n$, so daß gilt:*
Jede Lösung $\mathbf{x} \in K^n$ von $A\mathbf{x} = \mathbf{b}$ läßt sich auf genau eine Weise als $\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{l}_1 + \cdots + t_{n-r}\mathbf{l}_{n-r}$ mit $t_1, \ldots, t_{n-r} \in K$ schreiben, und alle solchen Vektoren $\mathbf{x} \in K^n$ sind Lösungen von $A\mathbf{x} = \mathbf{b}$.
Jedes System $\mathbf{l}_1, \ldots, \mathbf{l}_{n-r}$ von Vektoren aus K^n mit dieser Eigenschaft heißt System von Fundamentallösungen des zugehörigen homogenen Gleichungssystems $A\mathbf{x} = \mathbf{0}$. Ist hier $r = n$, so ist $\{\mathbf{l}_1, \ldots, \mathbf{l}_{n-r}\} = \emptyset$.

Korollar 2.12. *Hat das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ mehr Unbekannte als Gleichungen $(n > p)$ und besitzt es überhaupt Lösungen, so ist die Lösung nicht eindeutig. Insbesondere gilt: Ein homogenes lineares Gleichungssystem $A\mathbf{x} = \mathbf{0}$ mit $A \in M(p \times n, K)$ und $n > p$ hat nichttriviale Lösungen.*

Korollar 2.13. *Ein inhomogenes lineares Gleichungssystem $A\mathbf{x} = \mathbf{b}$ mit $A \in M(n \times n, K)$, dessen zugehöriges homogenes Gleichungssystem nur die triviale Lösung hat, besitzt (bei beliebigem $\mathbf{b} \in K^n$) eine eindeutige Lösung.*

Korollar 2.14. *Das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ mit $A \in M(p \times n, K)$ und $p > n$ ist nicht für alle $\mathbf{b} \in K^p$ lösbar.*

Bemerkung.

a) Die durch das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ zunächst implizit gegebene Lösungsmenge $\mathcal{L} = \{\mathbf{x} \in K^n \mid A\mathbf{x} = \mathbf{b}\}$ wird durch die Darstellung in Satz 2.11 parametrisiert, d.h., wir erhalten ihre Elemente als Bild der bijektiven Abbildung

$$(t_1, \ldots, t_{n-r}) \longmapsto \mathbf{x}_0 + t_1\mathbf{l}_1 + \cdots + t_{n-r}\mathbf{l}_{n-r}$$

von K^{n-r} nach L.

b) Wir werden später sehen, daß die Zahl r nicht davon abhängt, auf welchem Weg man die Matrix in Zeilenstufenform überführt.

c) Die elementaren Umformungen definieren Abbildungen

$$u: \ M(p \times n, K) \longrightarrow M(p \times n, K),$$

und zu jeder elementaren Umformung entsteht auch die Umkehrabbildung

$$u^{-1}: \ M(p \times n, K) \longrightarrow M(p \times n, K)$$

durch elementare Umformungen.

3. Vektorräume und lineare Abbildungen

Wollten wir nur ein effektives Verfahren gewinnen, lineare Gleichungssysteme zu lösen, wäre die Vorlesung an dieser Stelle beendet.
Wir wollen jetzt die Ergebnisse, die wir über die Lösungsmenge eines linearen Gleichungssystems gewonnen haben, in den Zusammenhang der Theorie der Vektorräume und ihrer strukturerhaltenden Abbildungen, der linearen Abbildungen, stellen.

Definition 3.1. *Sei $(K, +, \cdot)$ ein Körper, V eine Menge mit einem ausgezeichneten Element $\mathbf{0}_V = \mathbf{0}$ (das der* Nullvektor *von V genannt wird) und einer Verknüpfung $+_V = + : V \times V \to V$ (Addition in V, Vektoraddition) sowie einer Verknüpfung $\cdot_V = \cdot : K \times V \to V$ (*Skalarmultiplikation*).*
*V mit diesen Verknüpfungen heißt ein K-*Vektorraum*, falls: Für die Vektoraddition gilt:*

VA1 *(*Assoziativgesetz *der Vektoraddition): Für alle $u, v, w \in V$ gilt $(u + v) + w = u + (v + w)$.*
VA2 *(*Kommutativgesetz *der Vektoraddition): Für alle $u, v \in V$ gilt $u + v = v + u$.*
VA3 *(*Neutrales Element *der Vektoraddition): Für alle $v \in V$ gilt $v + \mathbf{0} = \mathbf{0} + v = v$.*
VA4 *(*Inverses Element *der Vektoraddition): Für alle $v \in V$ gibt es genau ein mit $-v$ bezeichnetes Element von V, für das $v + (-v) = (-v) + v = \mathbf{0}$ gilt.*

Für die Skalarmultiplikation gilt:

SM1 $1 \cdot v = v$ *für alle* $v \in V$
SM2 $(a + b)v = av + bv$ *für alle* $a, b \in K,\ v \in V$
SM3 $a(v + w) = av + aw$ *für alle* $a \in K,\ v, w \in V$
SM4 $a(bv) = (ab)v$ *für alle* $a, b \in K,\ v \in V$.

Eine Teilmenge W von V heißt Teilraum *(*Untervektorraum, Unterraum*), falls gilt:*

a) $\mathbf{0} \in W$
b) *(Abgeschlossenheit unter Addition) Für alle $w_1, w_2 \in W$ gilt $w_1 + w_2 \in W$*
c) *(Abgeschlossenheit unter Skalarmultiplikation) Für alle $\lambda \in K, w \in W$ ist $\lambda w \in W$.*

Bemerkung 3.2. Die Eigenschaften **VA1** bis **SM4** heißen die *Vektorraumaxiome.*

Beispiele:

- $K^n = \underbrace{K \times \cdots \times K}_{n-\text{mal}} = \left\{ \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \mid a_i \in K \text{ für } 1 \leq i \leq n \right\}$

  mit Addition $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}$

  und Skalarmultiplikation

  $$\lambda \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \vdots \\ \lambda a_n \end{pmatrix}.$$

  Die Elemente von K^n schreiben wir in der Regel wie oben als *Spaltenvektoren* $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. Den *Zeilenvektor* $(a_1, \ldots, a_n)$ schreibt man auch als ${}^t\mathbf{a}$. Wenn dadurch keine Verwirrung entstehen kann, schreibt man ihn einfach (eigentlich inkorrekt) ebenfalls als $\mathbf{a}$.
- $\mathbb{R}$ ist (mit der gewöhnlichen Multiplikation als Skalarmultiplikation) ein $\mathbb{Q}$-Vektorraum. Allgemeiner gilt:
  Ist L ein Körper, $K \subseteq L$ ein Teilkörper (also K eine Teilmenge, die bezüglich $+$ und $\cdot$ selbst ein Körper ist), so ist L ein K-Vektorraum.
- Ist K ein Körper, M eine Menge, so ist

  $$V := K^M := \{f : M \longrightarrow K \mid f \text{ ist Abbildung}\}$$

  mit den Verknüpfungen:

  $$\begin{array}{rcll} f_1 + f_2 & = & g \text{ mit } g(a) = f_1(a) + f_2(a) & \text{für alle } a \in M \\ \lambda f & = & h \text{ mit } h(a) = \lambda \cdot f(a) & \text{für alle } a \in M \end{array}$$

  ein K-Vektorraum (Übung).
  Man überlege sich als weitere Übung, dass diese Konstruktion für $M = \{1, 2, \ldots, n\} \subseteq \mathbb{N}$ erneut den Vektorraum K^n (in leichter Verkleidung) liefert.
  Allgemeiner können wir in dieser Definition auch K durch einen K-Vektorraum W ersetzen und erhalten wieder einen K-Vektorraum.
  Dagegen ist (mit den gleichen Verknüpfungen)

  $$\{f : M \longrightarrow \mathbb{R} \mid f(x) \geq 0 \text{ für alle } x \in M\}$$

  bzw.

  $$\{f : M \longrightarrow \mathbb{R} \mid f \text{ ist injektiv}\}$$

  kein $\mathbb{R}$-Vektorraum (letzteres falls $|M| > 1$).

- $G_1 := \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 + x_2 = 0 \right\}$ ist ein Teilraum von $\mathbb{R}^2$, dagegen ist $G_2 := \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 + x_2 = 1 \right\}$ kein Teilraum von $\mathbb{R}^2$.
  Geometrisch gesehen ist G_1 die Menge der Ortsvektoren der Punkte einer Geraden durch den Ursprung, während G_2 eine Gerade beschreibt, die nicht durch den Ursprung geht.
- Das vorige Beispiel lässt sich offensichtlich verallgemeinern: Im (mit dem 3-dimensionalen Anschaungsraum identifizierten) Vektorraum $\mathbb{R}^3$ sind die Geraden und Ebenen durch den Ursprung Untervektorräume, während Geraden bzw. Ebenen, die nicht durch den Ursprung gehen, keine Untervektorräume sind. Als Übung rechne man das einmal mit Hilfe der Geraden- bzw. Ebenengleichungen nach, zum anderen mit der *Parameterdarstellung* (zur Erinnerung: sind $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$ drei Punkte der Ebene E, die nicht auf einer gemeinsamen Geraden liegen, und setzt man $\mathbf{a}_1 := \mathbf{x}_1 - \mathbf{x}_0, \mathbf{a}_2 := \mathbf{x}_2 - \mathbf{x}_0$, so ist
  $$E = \{\mathbf{x}_0 + t_1\mathbf{a}_1 + t_2\mathbf{a}_2 \mid t_1, t_2 \in \mathbb{R}\},$$
  die Darstellung der Punkte der Ebene in dieser Form mit den freien Variablen (*Parametern*) t_1, t_2 nennt man eine Parameterdarstellung der Ebene. Analog (aber natürlich nur mit *einem* Parameter) ist die Parameterdarstellung einer Geraden definiert).
- Der Nullraum $\{\mathbf{0}\}$ ist ein Teilraum des Vektorraums V (V ein beliebiger K-Vektorraum, $\mathbf{0}$ der Nullvektor in V).

Die Ähnlichkeit der Axiome für die Vektoraddition und für die Addition im Körper benutzen wir zu einer weiteren Abstraktion:

Definition 3.3. *Sei G eine Menge mit einem ausgezeichneten Element $e \in G$ und einer Verknüpfung $\circ : G \times G \to G$.*
$(G, e, \circ)$ (oder einfach nur G) heißt eine Gruppe*, wenn gilt:*

G1 *(Assoziativität): Für alle $g_1, g_2, g_3 \in G$ gilt $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$.*
G2 *(linksneutrales Element) Für alle $g \in G$ gilt $e \circ g = g$.*
G3 *(linksinverses Element) Zu jedem $g \in G$ gibt es ein $g' \in G$ mit $g' \circ g = e$.*

Gilt zusätzlich

AG *(Kommutativität) Für alle $g_{,}g_2 \in G$ gilt $g_1 \circ g_2 = g_2 \circ g_1$. ,*

so heißt die Gruppe kommutativ oder abelsch.

Beispiel:

a) Ist K ein Körper und $K^\times := K \setminus \{0\}$, so sind $(K, +, 0)$ und $(K^\times, \cdot, 1)$ abelsche Gruppen, ist V ein Vektorraum, so ist $(V, +, \mathbf{0})$ eine abelsche Gruppe.

b) Sei $X \neq \emptyset$ eine nichtleere Menge, $\mathrm{Perm}(X) := S_X := \{f : X \to X \mid f \text{ ist bijektiv}\}$ die Menge der *Permutationen* (bijektiven Selbstabbildungen) von X. Mit der Komposition von Abbildungen als Verknüpfung und der identischen Abbildung Id_X als neutralem Element ist dann S_X eine Gruppe. Diese ist nicht kommutativ, wenn X mehr als zwei Elemente hat. Ist $X = \{1, \ldots, n\}$, so schreibt man auch S_n und nennt diese Gruppe die *symmetrische Gruppe* auf n Elementen.

Lemma 3.4. *Sei $(G, \circ, e)$ eine Gruppe. Dann gilt:*

a) *$g \circ e = g$ für alle $g \in G$ (e ist auch rechtsneutral).*
b) *Ist $e' \in G$ mit $e' \circ g = g$ für alle $g \in G$, so ist $e' = e$ (das neutrale Element ist eindeutig bestimmt).*
c) *Für alle $g \in G$ ist $g \circ g^{-1} = e$ (g^{-1} ist auch rechtsinvers).*
d) *Sind $g, h \in G$ mit $g \circ h = e$, so ist $h = g^{-1}$ (das inverse Element ist eindeutig bestimmt).*

Beweis. Übung. □

Bemerkung. In der Definition des Gruppenbegriffs kann man also äquivalent G2, G3 ersetzen durch:
(G2'): Für alle $g \in G$ gilt $e \circ g = g \circ e = g$ und e ist durch diese Eigenschft eindeutig bestimmt.
(G3'): Für jedes $g \in G$ gibt es genau ein Element $g^{-1} \in G$ mit $g \circ g^{-1} = g^{-1} \circ g = e$.

Definition 3.5. *Eine Teilmenge H der Gruppe G heißt* Untergruppe *von G, wenn gilt:*

a) *$e \in H$*
b) *Für alle $h_1, h_2 \in H$ ist $h_1 \circ h_2 \in H$ (Abgeschlossenheit unter der Verknüpfung).*
c) *Für alle $h \in H$ ist $h^{-1} \in H$ (Abgeschlossenheit unter Inversenbildung).*

Lemma 3.6. *Sei K ein Körper, V ein K-Vektorraum. Dann gilt:*

a) *Für alle $v, w \in V$ gibt es genau ein $x \in V$ mit $v + x = w$ (die Gleichung $v + x = w$ ist eindeutig lösbar). Insbesondere ist der Nullvektor $\mathbf{0}$ das einzige neutrale Element der Vektoraddition.*
b) *$a \cdot \mathbf{0} = 0 \cdot v = \mathbf{0}$ für alle $a \in K$, $v \in V$*
c) *$a \cdot (-v) = (-a) \cdot v = -(a \cdot v)$ für alle $a \in K$, $v \in V$*
d) *$(-a)(-v) = av$ für alle $a \in K$, $v \in V$*
e) *$a(v - w) = av - aw$ für alle $a \in K$, $v \in V$*
f) *$(a - b)v = av - bv$ für alle $a, b \in K$, $v \in V$.*

Beweis. Übung □

Lemma 3.7. *Sei V ein K-Vektorraum, $W \subseteq V$ eine Teilmenge.*

a) *Ist W ein Unterraum von V, so ist zu jedem $w \in W$ auch $-w \in W$.*
b) *Ist W ein Unterraum von V, so liefert die Einschränkung von $+_V : V \times V \to V$ auf $W \times W$ eine Verknüpfung $+ = +_W : W \times W \to W$ und die Einschränkung der Skalarmultiplikation $\cdot : K \times V \to V$ auf $K \times W$ liefert eine Verknüpfung $\cdot_W = \cdot : K \times W \to W$. Bezüglich dieser von V her induzierten Verknüpfungen ist dann W ein K-Vektorraum.*
c) *W ist genau dann ein Unterraum von V, wenn gilt:*
   i) *$W \neq \emptyset$*
   ii) *Für alle $w_1, w_2 \in W, \lambda \in K$ ist $w_1 + \lambda w_2 \in W$.*

Beweis. Übung □

Bemerkung. Zu b) und c) analoge Aussagen gelten für Untergruppen: eine Untergruppe ist mit der auf sie eingeschränkten Verknüpfung selbst eine Gruppe, und $H \subseteq G$ ist genau dann Untergruppe der Gruppe G, wenn gilt:

a) $H \neq \emptyset$
b) Für alle $h_1, h_2 \in H$ ist $h_1^{-1} h_2 \in H$.

Wir betrachten jetzt die strukturerhaltenden Abbildungen für Vektorräume:

Definition 3.8. *Seien V, W Vektorräume über dem Körper K. Eine Abbildung $f : V \longrightarrow W$ heißt* linear *(oder auch* Homomorphismus von Vektorräumen*), falls gilt:*

a) *$f(v + w) = f(v) + f(w)$ für alle $v, w \in V$*
b) *$f(av) = af(v)$ für alle $a \in K$, $v \in V$.*

Ist $f : V \longrightarrow W$ linear, so heißt

$$\mathrm{Ker}(f) := \{v \in V \mid f(v) = \mathbf{0}_W\}$$

der Kern von f.

Lemma 3.9. *Seien V, W Vektorräume über dem Körper K.*
Die Abbildung $f : V \longrightarrow W$ ist genau dann linear, wenn gilt:
Für alle $v_1, v_2 \in V, \lambda \in K$ ist $f(v_1 + \lambda v_2) = f(v_1) + \lambda f(v_2)$.

Beweis. Klar. □

Lemma 3.10. *Seien V, W Vektorräume über dem Körper K, sei $f : V \longrightarrow W$ eine lineare Abbildung. Dann gilt:*

a) *$f(\mathbf{0}_V) = \mathbf{0}_W$.*
b) *Für alle $v \in V$ ist $f(-v) = -f(v)$.*
c) *Für alle $v_1, v_2 \in V$ ist $f(v_1 - v_2) = f(v_1) - f(v_2)$.*

Beweis. a) $f(\mathbf{0}_V) = f(\mathbf{0}_V + \mathbf{0}_V) = f(\mathbf{0}_V) + f(\mathbf{0}_V)$, es folgt $f(\mathbf{0}_V) = \mathbf{0}_W$.

b) $f(-v) + f(v) = f(v + (-v) = f(\mathbf{0}_V) = \mathbf{0}_W$ (die letzte Gleichung gilt wegen a)), also ist $f(-v) = -f(v)$.
c) Übung.

$\square$

Beispiel:

- $K = \mathbb{R} = V$, $f(x) = ax$ ($a \in \mathbb{R}$ beliebig fest) ist linear.
- $K = \mathbb{R} = V$, $f(x) = 2x - 3$ ist nicht linear.
- $K = \mathbb{R}$, $V = \mathbb{R}^3$, $f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 + x_3 \\ x_2 \\ 0 \end{pmatrix}$ ist linear.
- $K = \mathbb{R}$, $V = \{f : \mathbb{R} \longrightarrow \mathbb{R} \mid f$ ist in ganz $\mathbb{R}$ differenzierbar$\}$
  $W = \{f : \mathbb{R} \longrightarrow \mathbb{R}\}$
  $D : V \longrightarrow W$ gegeben durch
  $D(f) = f'$ (Ableitung)
  ist eine lineare Abbildung.
- $K = \mathbb{R}$, $V = W = \{f : [0,1] \longrightarrow \mathbb{R} \mid f$ ist stetig in $[0,1]\}$
  $I : V \longrightarrow W$ gegeben durch
  $I(f)(x) := \int_0^x f(t)dt$
  ist linear.

Lemma 3.11. *V, W seien K-Vektorräume, $f : V \longrightarrow W$ eine lineare Abbildung.*
Dann sind

$$\begin{aligned} \mathrm{Kern}(f) &:= \{v \in V \mid f(v) = \mathbf{0}_W\} \text{ und} \\ \mathrm{Im}(f) &:= \{f(v) \mid v \in V\} =: f(V) \end{aligned}$$

Unterräume von V bzw. W.
f ist genau dann injektiv, wenn $\mathrm{Ker}(f) = \{\mathbf{0}_V\}$ gilt.

Beweis. Wegen $f(\mathbf{0}_V) = \mathbf{0}_W$ ist $\mathbf{0}_V \in \mathrm{Ker}(f), \mathbf{0}_W \in \mathrm{Im}(f)$. Beide Mengen sind also nichtleer.
Man rechnet mit Hilfe der in Definition 3.8 angegebenen Eigenschaften einer linearen Abbildung nach, dass beide Mengen unter den Vektorraumverknüpfungen abgeschlossen sind.
Sei jetzt f injektiv, $v \in \mathrm{Ker}(f)$, also $f(v) = \mathbf{0}_W = f(\mathbf{0}_V)$.
Weil f injektiv ist, muss dann $v = \mathbf{0}_V$ gelten, also ist $\mathrm{Ker}(f) = \{\mathbf{0}_V\}$.
Ist umgekehrt $\mathrm{Ker}(f) = \{\mathbf{0}_V\}$ und sind $v_1, v_2 \in V$ mit $f(v_1) = f(v_2)$, so ist $\mathbf{0} = f(v_1) - f(v_2) = f(v_1 - v_2)$, also ist $v_1 - v_2 \in \mathrm{Ker}(f) = \{\mathbf{0}_V\}$.
Damit folgt $v_1 - v_2 = \mathbf{0}_V$, also $v_1 = v_2$. $\square$

Bemerkung. Man überlege sich als Übung für $V = W = \mathbb{R}^n$ mit $n = 2$ oder $n = 3$, dass eine lineare Abbildung $f : V \longrightarrow V$ genau dann linear ist, wenn sie $\mathbf{0}$ auf $\mathbf{0}$ abbildet und Geraden auf Geraden oder auf einen Punkt abbildet. Dies erklärt die Bezeichnung „linear“ (lateinisch: linea = Gerade).

Definition und Lemma 3.12. *Sei*

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{p1} & \dots & a_{pn} \end{pmatrix} = (a_{ij})_{\substack{1\le i\le p \\ 1\le j\le n}} \in M(p \times n, K).$$

Dann definiert A durch

$$L_A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$$

mit $y_i = \sum_{j=1}^{n} a_{ij}x_j \quad (1 \le i \le p)$ *eine Abbildung*

$$L_A : K^n \longrightarrow K^p,$$

die zu A gehörige lineare Abbildung *von K^n nach K^p. Man schreibt auch $L_A(\mathbf{x}) =: A\mathbf{x}$.*

Beweis. Das rechnet man leicht nach. □

Wir können jetzt auch den soeben eingeführten Begriff der linearen Abbildung mit dem Problem verbinden, lineare Gleichungssysteme zu lösen:

Satz 3.13. *Sei $A\mathbf{x} = \mathbf{b}$ ein lineares Gleichungssystem mit Koeffizienten in K (mit $A \in M(p \times n, K)$, $\mathbf{b} \in K^p$), $\mathcal{L} = \mathcal{L}(A, \mathbf{b}) \subseteq K^n$ die Lösungsmenge, $\mathcal{L}_0 = \mathcal{L}(A, \mathbf{0})$ die Lösungsmenge des zugehörigen homogenen Systems $A\mathbf{x} = \mathbf{0}$. Dann gilt:*

a) *Das System $A\mathbf{x} = \mathbf{b}$ ist genau dann lösbar, wenn $\mathbf{b} \in \mathrm{Im}(L_A)$ gilt; die Lösungsmenge $\mathcal{L}$ ist das Urbild $L_A^{-1}(\{\mathbf{b}\}) = \{\mathbf{x} \in K^n \mid L_A(\mathbf{x}) = \mathbf{b}\}$.*
b) *Insbesondere gilt für die Lösungsmenge $\mathcal{L}_0$ des homogenen Gleichungssystems*
$$\mathcal{L}_0 = \mathrm{Ker}(L_A)$$
ist ein Untervektorraum des K^n.
c) *Für jedes $\mathbf{x}_0 \in \mathcal{L}(A, \mathbf{b})$ ist*
$$\mathcal{L}(A, \mathbf{b}) = \mathbf{x}_0 + \mathcal{L}_0 := \{\mathbf{x}_0 + \mathbf{y} \mid \mathbf{y} \in \mathcal{L}_0\}.$$

Beweis. Klar. □

Den Zusammenhang zwischen dem Vektorraumbegriff und linearen Gleichungssystemen, der sich hieraus ergibt, beschreiben wir noch einmal gesondert in dem folgenden Satz:

Satz 3.14. *Sei $A\mathbf{x} = \mathbf{b}$ ein lineares Gleichungssystem mit Koeffizienten in K (mit $A \in M(p \times n, K)$, $\mathbf{b} \in K^p$), $\mathcal{L} = \mathcal{L}(A, \mathbf{b}) \subseteq K^n$ die Lösungsmenge, $\mathcal{L}_0 = \mathcal{L}(A, \mathbf{0})$ die Lösungsmenge des zugehörigen homogenen Systems $A\mathbf{x} = \mathbf{0}$. Dann gilt:*

a) $\mathcal{L}_0$ *ist ein Untervektorraum von* K^n.
b) *Sind* $\mathbf{x}, \mathbf{y} \in \mathcal{L}$ *Lösungen des Gleichungssystems, so ist die Differenz* $\mathbf{x} - \mathbf{y}$ *eine Lösung des zugehörigen homogenen Gleichungssystems (also* $\mathbf{x} - \mathbf{y} \in \mathcal{L}_0$*).*
c) *Für* $\mathbf{x}_0 \in \mathcal{L}$ *ist*

$$\mathcal{L} = \{\mathbf{x}_0 + \mathbf{y} \mid \mathbf{y} \in \mathcal{L}_0\} =: \mathbf{x}_0 + \mathcal{L}_0$$

(hat man eine spezielle Lösung $\mathbf{x}_0$ *des inhomogenen Systems, so erhält man* alle Lösungen *des inhomogenen Systems, indem man alle Lösungen des zugehörigen homogenen Systems zu der speziellen Lösung* $\mathbf{x}_0$ *hinzuaddiert).*
Insbesondere gilt: Genau dann besitzt $A\mathbf{x} = \mathbf{b}$ *für jedes* $\mathbf{b} \in K^p$ *höchstens eine Lösung, wenn* $A\mathbf{x} = \mathbf{0}$ *nur die* triviale Lösung $\mathbf{x} = \mathbf{0}$ *hat.*

Beweis. Das folgt aus dem vorigen Satz. □

4. Basis und Dimension

Wir haben in Abschnitt 2 gesehen, dass man im Lösungsraum eines homogenen linearen Gleichungssystems eine Menge von Vektoren $\mathbf{l}_1, \ldots, \mathbf{l}_{n-r}$ finden kann (ein System von Fundamentallösungen), die die Eigenschaft hat, dass sich jeder Lösungsvektor in eindeutiger Weise als Linearkombination $\sum_{j=1}^{n-r} t_j \mathbf{l}_j$ mit Koeffizienten $t_j \in K$ schreiben lässt. Noch nicht völlig klar ist, ob die Anzahl der Fundamentallösungen nur vom Gleichungssystem, nicht aber von den durchgeführten Rechenschritten abhängt.
In diesem Abschnitt soll gezeigt werden, dass wir ein (*Basis* genanntes) Vektorsystem mit ähnlichen Eigenschaften in jedem Vektorraum finden können, und dass die Elementanzahl eines solchen Systems, wenn es endlich ist, eine feste nur vom betrachteten Vektorraum abhängige Zahl ist (die *Dimension* des Vektorraums).
Zunächst stellen wir ein paar einfache Eigenschaften und Bezeichnungen zusammen.
Vorab aber noch eine

Bemerkung. Wir haben schon wiederholt ohne Beweis benutzt, dass man für jede assoziative Verknüpfung $(a, b) \mapsto a \circ b$ in einem Produkt von $n \geq 3$ Elementen Klammern beliebig verschieben kann, dass also etwa

$$(a_1 \circ (a_2 \circ a_3)) \circ a_4 = (a_1 \circ a_2) \circ (a_3 \circ a_4) = a_1 \circ ((a_2 \circ a_3) \circ a_4)$$

gilt; man lässt dann in längeren Produkten die Klammern auch ganz fort und schreibt

$$a_1 \circ a_2 \circ \ldots \circ a_n$$

für jeden der Ausdrücke aus obiger Gleichung. Der Beweis dieser Tatsache ist eine (verhältnismäßig langweilige) Routineübung im sauberen Aufschreiben von Induktionsbeweisen.
Genauso zeigt man mit vollständiger Induktion, dass man bei einer kommutativen Verknüpfung $(a, b) \mapsto a \circ b$ in einem Produkt von n Elementen die Faktoren beliebig anordnen kann. Bei einer additiv geschriebenen Verknüpfung schreibt man daher (ohne Rücksicht auf die Anordnung)

$$\sum_{i=1}^{n} a_i = \sum_{i \in \{1,\ldots,n\}} a_i = a_1 + a_2 + \ldots + a_n,$$

bei einer multiplikativ geschriebenen Verknüpfung

$$\prod_{i=1}^{n} a_i = \prod_{i \in \{1,\ldots,n\}} a_i = a_1 \cdot a_2 \cdot \ldots \cdot a_n.$$

Nun also zu den angekündigten Eigenschaften und Bezeichnungen für Vektorräume:

Lemma 4.1. *Seien U_1, U_2 Unterräume des K-Vektorraums V.*

a) *$U_1 \cap U_2$ ist Unterraum von V.*
Allgemeiner gilt: Ist $X \subseteq \mathfrak{P}(V)$ eine Menge von Unterräumen von V, so ist $\bigcap_{U \in X} U$ ein Unterraum von V.

b) *$U_1 + U_2 := \{u_1 + u_2 \in V \mid u_1 \in U_1,\ u_2 \in U_2\}$ (die* Summe *von U_1, U_2) ist ein Unterraum von V.*
Allgemeiner gilt: Sind $U_1, \ldots U_n$ Unterräume von V so ist

$$U_1 + \ldots + U_n := \{u_1 + \ldots + u_n \in V \mid u_i \in U_i,\ (1 \leq i \leq n)\}$$

ein Unterraum von V.

Beweis. a) Zunächst ist $U_1 \cap U_2 \neq \emptyset$, da $\mathbf{0} \in U_j$ für $j = 1, 2$ gilt. Sind $v, w \in U_1 \cap U_2$, so ist $v + w \in U_1$, da $v \in U_1, w \in U_1$ gilt und U_1 ein Unterraum ist. Genauso ist $v + w \in U_2$, also gilt $v + w \in U_1 \cap U_2$. Ist $\lambda \in K$, so ist $\lambda v \in U_1, \lambda v \in U_2$, da $v \in U_1, v \in U_2$ gilt und U_1 und U_2 Unterräume von V sind. Damit ist gezeigt, dass $U_1 \cap U_2$ ein Unterraum ist. Genauso zeigt man die allgemeinere Aussage.

b) $U_1 + U_2$ ist wegen $\mathbf{0} \in U_1 + U_2$ nicht die leere Menge. Sind $v = v_1 + v_2$ und $w = w_1 + w_2$ mit $v_1, w_1 \in U_1, v_2, w_2 \in U_2$ Vektoren in $U_1 + U_2$, so ist $v + w = (v_1 + w_1) + (v_2 + w_2)$ wegen des Kommutativ- und des Assoziativgesetzes für die Addition in V, und da U_1 und U_2 Unterräume sind, ist $v_1 + w_1 \in U_1, v_2 + w_2 \in U_2$ und es folgt $v + w \in U_1 + U_2$. Genauso rechnet man nach, dass $\lambda v \in U_1 + U_2$ für $\lambda \in K$ gilt.

□

Bemerkung. Analog gilt auch: Ist G eine Gruppe mit Untergruppen H_1, H_2, so ist $H_1 \cap H_2$ ebenfalls eine Untergruppe von G.
Ist $(G, +)$ eine abelsche Gruppe mit Untergruppen H_1, H_2 und ist $H_1 + H_2$ wie oben definiert, so ist $H_1 + H_2$ eine Untergruppe von G.
Eine ähnliche Aussage für nicht kommutative Gruppen gilt im allgemeinen nicht.

Definition 4.2. *Sei $M = \{v_1, \ldots, v_n\} \subseteq V$ eine endliche Teilmenge des K - Vektorraums V. Ein Element $v \in V$ heißt* Linearkombination *der Elemente von M, wenn es $t_1, \ldots, t_n \in K$ gibt, so dass*

$$v = \sum_{i=1}^{n} t_i v_i = t_1 v_1 + \ldots + t_n v_n$$

gilt.
Ist $M \subseteq V$ eine beliebige (möglicherweise unendliche) Teilmenge von V, so heißt $v \in V$ eine Linearkombination der Elemente von M, wenn es $n \in \mathbb{N}$ und $v_1, \ldots, v_n \in M$ gibt, so dass

$$v = \sum_{i=1}^{n} t_i v_i = t_1 v_1 + \ldots + t_n v_n$$

mit geeigneten Koeffizienten $t_1, \ldots, t_n \in K$ *gilt (anders gesagt: Wenn* v *Linearkombination endlich vieler Elemente von* M *ist).*
Die lineare Hülle $\mathrm{Lin}(M)$ *(oder* $\mathrm{Span}(M)$*) ist die Menge aller Linearkombinationen von Elementen von* M *(mit der Konvention* $\mathrm{Lin}(\emptyset) = \{\mathbf{0}\}$*. Man schreibt auch* $\langle M \rangle := \mathrm{Lin}(M)$ *oder für endliche Mengen* $M = \{v_1, \ldots, v_n\}$ *auch* $\langle M \rangle =: \langle v_1, \ldots, v_n \rangle$*.* $\mathrm{Lin}(M)$ *heißt auch der von* M *erzeugte* (aufgespannte) Teilraum *und* M *ein* Erzeugendensystem *von* $\mathrm{Lin}(M)$*.*

Beispiel:

- Sind $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in K^3$, so ist $\langle v_1, v_2, v_3 \rangle = K^3$.
- Sind $v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$, $v_3 = \begin{pmatrix} 5 \\ 2 \\ 5 \end{pmatrix} \in \mathbb{R}^3$, so ist $v_3 \in \langle v_1, v_2 \rangle$.
  Dagegen ist $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \notin \langle v_1, v_2 \rangle$.

Satz 4.3. *Sei* V *ein* K*-Vektorraum. Dann ist* $\mathrm{Lin}(M)$ *für jede Menge* $M \subseteq V$ *ein Unterraum von* V*. Ferner gilt:*

a) $M \subseteq V$ *ist genau dann ein Unterraum, wenn* $M = \mathrm{Lin}(M)$ *gilt.*
b) $\mathrm{Lin}(M)$ *ist der Durchschnitt aller Unterräume* U *von* V *mit* $U \supseteq M$*.*
c) *Für* $M \subseteq M'$ *ist* $M \subseteq \mathrm{Lin}(M) \subseteq \mathrm{Lin}(M')$*.*
d) $\mathrm{Lin}(\mathrm{Lin}(M)) = \mathrm{Lin}(M)$*.*
e) *Für Unterräume* U_1, U_2 *von* V *ist* $\mathrm{Lin}(U_1 \cup U_2) = U_1 + U_2$*.*

Beweis. Zunächst ist $\mathbf{0} \in \mathrm{Lin}(M)$, denn für $M = \emptyset$ gilt das per definitionem und sonst ist $\mathbf{0} = 0 \cdot v$ für ein beliebiges $v \in M$. Dass für Vektoren $v, w \in \mathrm{Lin}(M)$ und $\lambda \in K$ auch $\lambda \cdot v + w = \lambda \cdot v + 1 \cdot w$ eine Linearkombination der Elemente von M ist, sieht man sofort, also ist $\mathrm{Lin}(M)$ in der Tat ein Unterraum von V.
Zu den weiteren Punkten:

a) Wegen $1 \cdot v = v$ gilt stets $M \subseteq \mathrm{Lin}(M)$. Ist M ein Unterraum, so gilt $\sum_{i=1}^n t_i v_i \in M$ für beliebige $v_1, \ldots, v_n \in M$ und $t_1, \ldots, t_n \in K$, wie man durch Induktion nach n sofort sieht, also hat man dann auch $\mathrm{Lin}(M) \subseteq M$. Ist umgekehrt $M = \mathrm{Lin}(M)$, so ist M ein Unterraum, weil das für $\mathrm{Lin}(M)$ gilt.
b) Ist $U \supseteq M$ ein Unterraum von V, so sind alle Linearkombinationen der Elemente von M in U, also ist $\mathrm{Lin}(M) \subseteq U$, damit ist $\mathrm{Lin}(M)$ im Durchschnitt aller Unterräume enthalten, die M als Teilmenge enthalten. Umgekehrt ist $\mathrm{Lin}(M)$ selbst ein Unterraum, der M enthält, nimmt also an der Durchschnittsbildung

teil und enthält daher den Durchschnitt aller Unterräume von V, die M als Teilmenge enthalten.

c) ist klar

d) folgt aus a) und der Tatsache, dass $\mathrm{Lin}(M)$ ein Unterraum von V ist.

e) $U_1 + U_2$ ist ein Unterraum von V, der $U_1 \cup U_2$ enthält, enthält also nach b) den Unterraum $\mathrm{Lin}(U_1 \cup U_2)$. Andererseits ist klar, dass $U_1 + U_2$ in $\mathrm{Lin}(U_1 \cup U_2)$ enthalten ist.

□

Bemerkung.

a) Im allgemeinen ist $U_1 \cup U_2$ selbst für Unterräume U_1, U_2 kein Unterraum, wie man etwa am Beispiel $U_1 = \{\binom{\lambda}{0} \mid \lambda \in K\} \subseteq K^2$, $U_2 = \{\binom{0}{\lambda} \mid \lambda \in K\} \subseteq K^2$ sieht. Man kann (leicht) zeigen: $U_1 \cup U_2$ ist genau dann ein Unterraum, wenn $U_1 \subseteq U_2$ oder $U_2 \subseteq U_1$ gilt.

b) Sei $A \in M(p \times n, K)$ eine $p \times n$-Matrix mit Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n \in K^p$. Dann ist das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ genau dann lösbar, wenn $\mathbf{b} \in \langle \mathbf{s}_1, \ldots, \mathbf{s}_n \rangle$ gilt.

Definition und Satz 4.4. *Sei V ein K-Vektorraum, $M \subseteq V$ eine Teilmenge. M heißt* linear unabhängig, *wenn die folgenden äquivalenten Bedingungen erfüllt sind:*

a) *Für jedes $u \in M$ ist $u \notin \mathrm{Lin}(M \setminus \{u\})$*

b) *Sind $u_1, \ldots, u_n \in M$ beliebige paarweise verschiedene Elemente von M ($n \in \mathbb{N}$) und $a_1, \ldots, a_n \in K$ mit $\sum_{i=1}^{n} a_i u_i = \mathbf{0}$, so ist $a_1 = \cdots = a_n = 0$.*
($\mathbf{0}$ läßt sich nur auf die triviale Weise als Linearkombination der Elemente von M darstellen.)
Ist M nicht linear unabhängig, so heißt M linear abhängig.

Beweis von Satz 4.4. Wir müssen zeigen, dass die Bedingungen a) und b) zueinander äquivalent sind.
Gilt b) nicht, so seien $u_1, \ldots, u_n \in M$ paarweise verschiedene Vektoren und $a_1, \ldots, a_n \in K$ mit $\sum_{i=1}^{n} a_i u_i = 0$ und $a_i \neq 0$ für ein i, ohne Einschränkung sei $i = 1$, also $a_1 \neq 0$. Dann ist $u_1 = -a_1^{-1} \sum_{i=2}^{n} a_i u_i = \sum_{i=2}^{n} (-a_1^{-1} a_i) u_i \in \mathrm{Lin}(M \setminus \{u_1\}$, also gilt auch a) nicht.
Gilt andererseits a) nicht, so sei $u_1 \in M$ ein Vektor mit $u_1 \in \mathrm{Lin}(M \setminus \{u_1\})$. Es gibt dann also paarweise und von u_1 verschiedene Vektoren $u_2, \ldots, u_n \in M$ und $a_2, \ldots, a_n \in K$, so dass $u_1 = \sum_{i=2}^{n} a_i u_i$ gilt.
Mit $a_1 = -1$ ist dann aber

$$\sum_{i=1}^{n} a_i u_i = -u_1 + \sum_{i=2}^{n} a_i u_i = \mathbf{0}$$

eine nicht triviale Linearkombination des Nullvektors aus paarweise verschiedenen Vektoren von M, also gilt auch b) nicht. $\square$

Bemerkung. a) M ist genau dann linear unabhängig, wenn für jeden Vektor $u \in M$ gilt: u lässt sich nicht als Linearkombination der Vektoren in $M \setminus \{u\}$ darstellen (ist also unabhängig von den anderen Vektoren in M).

b) Ist $M \subseteq V$ linear unabhängig, so ist M ein *minimales Erzeugendensystem* von $\mathrm{Lin}(M)$: Entfernt man einen der Vektoren u von M, so ist die verbleibende Teilmenge $M \setminus \{u\}$ kein Erzeugendensystem von $\mathrm{Lin}(M)$ mehr. Anders formuliert (Kontraposition): Ist $M' \subseteq M$ eine Teilmenge von M mit $\mathrm{Lin}(M') = \mathrm{Lin}(M)$, so ist $M' = M$.

Beispiel:

a) Die Vektoren $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}$ in $\mathbb{R}^3$ sind linear unabhängig.

Dagegen sind $\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\2\\0\end{pmatrix}, \begin{pmatrix}5\\2\\5\end{pmatrix}$ linear abhängig.

b) Die Ortsvektoren von 3 Punkten im $\mathbb{R}^3$ sind genau dann linear abhängig, wenn die Punkte auf einer gemeinsamen Ebene durch den Ursprung liegen.
Äquivalent: Die Richtungsvektoren dreier Geraden durch den Ursprung sind genau dann linear abhängig, wenn diese Geraden in einer gemeinsamen Ebene liegen.

Die folgende Notation verallgemeinert den Begriff des n-Tupels:

Definition 4.5. *Seien X, I Mengen, $X^I := Abb(I, X)$ die Menge aller Abbildungen von I nach X. Die Elemente von X^I heißen auch durch I indizierte* Familien *von Elementen von X; für die Abbildung $f : I \longrightarrow X$ mit $f(i) =: x_i \in X (i \in I)$ wird auch $(f(i))_{i\in I} =: (x_i)_{i\in I}$ geschrieben.*

Bemerkung. Ist $I = \{1, \ldots, n\}$, so wird X^I durch $f \longmapsto (f(1), \ldots, f(n))$ bijektiv auf $X^n = \underbrace{X \times \cdots \times X}_{n-\text{mal}}$ abgebildet.

Definition 4.6. *Eine Familie $(v_i)_{i\in I} \in V^I$ von Vektoren in V heißt linear unabhängig, wenn die v_i paarweise verschieden sind und $\{v_i \mid i \in I\}$ als Menge linear unabhängig ist.*
Insbesondere haben wir also für n-Tupel $(v_1, \ldots, v_n) \in V^n$ von Vektoren in V: Das n-Tupel $(v_1, \ldots, v_n) \in V^n$ heißt linear unabhängig, wenn die v_i paarweise verschieden sind und die Menge $\{v_1, \ldots, v_n\}$ linear unabhängig ist.

Beispiel: Das Tripel $\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}\right)$ ist linear abhängig, dagegen ist die Menge $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}\right\} = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$ linear unabhängig.

Lemma 4.7. *Sei V ein K-Vektorraum und $(v_1, \ldots, v_n) \in V^n$ ein n-Tupel von Vektoren in V.*
Dann wird durch

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \mapsto f(\mathbf{b}) = \sum_{i=1}^{n} b_i v_i \in V$$

eine lineare Abbildung $f : K^n \mapsto V$ definiert.
Es gilt $\mathrm{Im}(f) =< v_1, \ldots, v_n >= \mathrm{Lin}(\{v_1, \ldots, v_n\})$, und f ist genau dann injektiv, wenn das n-Tupel $(v_1, \ldots, v_n)$ linear unabhängig ist.

Beweis. Übung. □

Definition 4.8. *Sei V ein K-Vektorraum, $M = \{v_1, \ldots, v_n\} \subseteq V$ eine endliche Teilmenge.*
M heißt Basis *von V, wenn gilt:*
Für jedes $v \in V$ gibt es genau ein n-Tupel $(a_1, \ldots, a_n) \in K^n$ mit

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

(jedes $v \in V$ lässt sich in eindeutiger Weise als Linearkombination der v_i darstellen).
Ist allgemeiner $M \subseteq V$ eine beliebige Teilmenge, so heißt M Basis *von V, wenn gilt:*
Zu jedem $v \in V$ gibt es genau ein $a \in K^{(M)} := \{f : M \to K \mid f(m) \neq 0$ für höchstens endlich viele $m \in M\}$ mit $v = \sum_{m \in M} a(m)m$.
(Da $a(m) \neq 0$ nur für höchstens endlich viele $m \in M$ gilt, enthält die Summe $\sum_{m \in M} a(m)m$ nur endlich viele von $\mathbf{0}$ verschiedene Summanden und ist daher sinnvoll.)
Äquivalent ist: Jedes $v \in V$ lässt sich in eindeutiger Weise als Linearkombination der Elemente von M schreiben.
Eine beliebige Familie $(v_i)_{i \in I} \in V^I$ (auch Vektorsystem genannt) heißt Basis von V, wenn die v_i paarweise verschieden sind und $\{v_i \mid i \in I\}$ eine Basis von V ist.
Ist $V = \{\mathbf{0}\}$, so heißt $\emptyset$ eine Basis von V.

Beispiel:

a) Ist $V = K^n$ und $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$, so heißt $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ die *Standardbasis* (oft auch in leicht missbräuchlicher Weise *kanonische Basis* genannt) von K^n.

b) Die in Satz 2.7 und Satz 2.11 angegebenen Fundamentallösungen des homogenen linearen Gleichungssystems $A\mathbf{x} = \mathbf{0}$ bilden eine Basis des Lösungsraums.

c) Ist V ein K-Vektorraum und $\{v_1, \ldots, v_n\}$ eine Basis von V, so erhält man durch

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \longmapsto \sum_{i=1}^{n} a_i v_i \in V$$

eine bijektive lineare Abbildung $K^n \longrightarrow V$: Man sagt, V und K^n seien zueinander isomorphe K-Vektorräume.

Korollar 4.9. a) *M ist genau dann linear unabhängig, wenn M eine Basis von* $\mathrm{Lin}(M)$ *ist, wenn sich also jeder Vektor in* $\mathrm{Lin}(M)$ *auf genau eine Weise als Linearkombination der Elemente von M schreiben läßt.*

b) *Sei M eine linear unabhängige Teilmenge von V, $v \in V$. Es gilt genau dann $v \in \mathrm{Lin}(M)$, wenn $v \in M$ ist oder $M \cup \{v\}$ linear abhängig ist.*

Beweis. Zu a): Ist M eine Basis von $\mathrm{Lin}(M)$, so ist $\mathbf{0} = \sum_{v \in M} 0 \cdot v$ die eindeutige Darstellung des Nullvektors $\mathbf{0}$ als Linearkombination der Elemente von M, also ist M nach a) linear unabhängig. Ist umgekehrt M linear unabhängig, so lässt sich zunächst nach Definition von $\mathrm{Lin}(M)$ jeder Vektor von $\mathrm{Lin}(M)$ als Linearkombination der Elemente von M schreiben. Sind $a, b \in K^{(M)}$ mit

$$\sum_{v \in M} a(v)v = \sum_{v \in M} b(v)v,$$

so ist $\sum_{v \in M}(a(v) - b(v))v = \mathbf{0}$, wegen der linearen Unabhängigkeit von M also $a(v) - b(v) = 0$ für alle $v \in M$, d.h, $a = b$, die Schreibweise ist also eindeutig und M ist eine Basis von $\mathrm{Lin}(M)$.
Zu b): Ist $v \in \mathrm{Lin}(M)$, so ist offenbar $\mathrm{Lin}(M) = \mathrm{Lin}(M \cup \{v\})$. Ist dann $v \notin M$, so ist $M \cup \{v\}$ nach Teil a) der Definition linear abhängig, da dann $M = (M \cup \{v\}) \setminus \{v\}$ gilt.
Ist umgekehrt $v \in M$, so ist offenbar $v \in \mathrm{Lin}(M)$. Ist $M \cup \{v\}$ linear abhängig, so gibt es (paarweise verschiedene) Vektoren $v_1, \ldots, v_r \in M$ und $a_1, \ldots, a_{r+1} \in K$, die nicht alle 0 sind, so dass $a_{r+1}v + \sum_{j=1}^{r} a_j v_j = \mathbf{0}$ gilt. Wäre $a_{r+1} = 0$, so wäre diese Gleichung bereits eine nicht triviale

lineare Relation zwischen den Vektoren $v_1, \ldots, v_r \in M$ im Widerspruch zur linearen Unabhängigkeit von M. Also ist $a_{r+1} \neq 0$ und daher

$$v = -a_{r+1}^{-1} \sum_{j=1}^{r} a_j v_j,$$

also ist $v \in \mathrm{Lin}(M)$. □

Der folgende Satz liefert drei immer wieder benutzte Charakterisierungen einer Basis:

Satz 4.10. *Sei V ein K-Vektorraum, $M \subseteq V$. Dann sind äquivalent:*

a) *M ist eine Basis von V*
b) *M ist linear unabhängig und $\mathrm{Lin}(M) = V$ (M ist linear unabhängiges Erzeugendensystem)*
c) *M ist minimales Erzeugendensystem von V, d.h., es gilt $\mathrm{Lin}(M) = V$, und ist $M' \subsetneq M$, so ist $\mathrm{Lin}(M') \neq V$.*
d) *M ist* maximale linear unabhängige Teilmenge *von V (d.h., M ist linear unabhängig, und ist $M' \subseteq V$ mit $M' \supsetneq M$, so ist M' nicht linear unabhängig).*

Beweis. a) $\Rightarrow$ b): Ist M eine Basis von V, so ist $V = \mathrm{Lin}(M)$ und damit ist nach Korollar 4.9 die Menge M als Basis von $\mathrm{Lin}(M)$ linear unabhängig.
b) $\Rightarrow$ c): Das ist Teil b) der Bemerkung nach Definition und Satz 4.4
c)$\Rightarrow$ d): M ist linear unabhängig nach Teil a) der Definition der linearen Unabhängigkeit. Ist nämlich $v \in M$ und $M' := M \setminus \{v\}$, so würde aus $v \in \mathrm{Lin}(M')$ folgen, dass $M' \cup \{v\} \subseteq \mathrm{Lin}(M')$ und daher

$$V = \mathrm{Lin}(M) = \mathrm{Lin}(M' \cup \{v\}) \subseteq \mathrm{Lin}(\mathrm{Lin}(M')) = \mathrm{Lin}(M')$$

gilt, was wegen der vorausgesetzten Minimalität von M nicht möglich ist.
Die Maximalität von M als linear unabhängige Teilmenge sehen wir so:
Für jedes $v \in V \setminus M$ ist

$$V = \mathrm{Lin}(M) = \mathrm{Lin}((M \cup \{v\}) \setminus \{v\}) \subseteq \mathrm{Lin}(M \cup \{v\}) \subseteq V,$$

also herrscht hier überall Gleichheit, und nach Teil a) der Definition der linearen Unabhängigkeit kann eine echte Obermenge von M nicht mehr linear unabhängig sein.
d) $\Rightarrow$ a) Da M eine Basis von $\mathrm{Lin}(M)$ ist, müssen wir nur noch $\mathrm{Lin}(M) = V$ zeigen. Sei $v \in V$. Ist $v \in M$, so ist $v \in \mathrm{Lin}(M)$, ist $v \notin M$, so ist $M \cup \{v\} \supsetneq M$ nach Voraussetzung (Maximalität von M) linear abhängig. Nach b) von Korollar 4.9 ist dann ebenfalls $v \in \mathrm{Lin}(M)$. □

Satz 4.11. *Jeder K-Vektorraum V hat eine Basis. Ist V endlich erzeugt (d.h., $V = \mathrm{Lin}(M)$ für eine endliche Menge $M \subseteq V$), so hat V eine endliche Basis.*

Beweis. Im endlich erzeugten Fall sei M ein endliches Erzeugendensystem von V, etwa $|M| = r$. Ist M minimal, so sind wir fertig. Andernfalls entfernen wir so lange Vektoren aus M, bis wir bei einem minimalen Erzeugendensystem angekommen sind (das ist spätestens nach r Schritten der Fall), dieses ist dann nach dem vorigen Satz eine Basis.
Der Beweis im nicht endlich erzeugten Fall ist etwas schwieriger,wir skizzieren ihn hier nur: Ist M nicht endlich erzeugt, so kann man mit Hilfe eines mengentheoretischen Satzes, des Zorn'schen Lemmas, zeigen, dass jedes Erzeugendensystem ein minimales Erzeugendensystem enthält. Da insbesondere ganz V ein Erzeugendensystem ist, zeigt das die Existenz einer Basis, allerdings in höchst nicht konstruktiver Weise. Mit dem gleichen Satz kann man alternativ zeigen, dass jede linear unabhängige Menge in einer maximalen linear unabhängigen Menge enthalten ist. Wir gehen darauf hier nicht näher ein, dieses Thema wird in der Vorlesung Lineare Algebra II behandelt. □

Bemerkung. Beim Beweis im endlich erzeugten Fall stellt man (ohne Zorn'sches Lemma) ebenfalls fest, daß jedes endliche Erzeugendensystem ein minimales und damit eine Basis enthält.

Definition und Lemma 4.12. *Sei $(v_1, \ldots, v_n) \in V^n$ ein n-Tupel (auch Vektorsystem genannt).* Elementare Umformungen *von $(v_1, \ldots, v_n)$ sind*

i) *die Ersetzung von v_i durch $v_i + \lambda v_j$ $(i \neq j,\ \lambda \in K)$*
ii) *die Ersetzung von v_i durch λv_i $(0 \neq \lambda \in K)$*
iii) *die Vertauschung von v_i und v_j.*

Sei $(v'_i, \ldots, v'_n)$ ein Vektorsystem, das aus $(v_1, \ldots, v_n)$ durch wiederholtes Anwenden elementarer Umformungen hervorgeht. Dann gilt:

a) *$(v'_1, \ldots, v'_n)$ ist genau dann linear unabhängig, wenn $(v_1, \ldots, v_n)$ linear unabhängig ist.*
b) $\langle v_1, \ldots, v_n \rangle = \langle v'_1, \ldots, v'_n \rangle$
c) *$(v'_1, \ldots, v'_n)$ ist genau dann Basis von V, wenn $(v_1, \ldots, v_n)$ Basis von V ist.*

Beweis. Zunächst stellen wir fest: Geht $(v'_1, \ldots, v'_n)$ aus $(v_1, \ldots, v_n)$ durch elementare Umformungen hervor, so auch $(v_1, \ldots, v_n)$ aus $(v'_1, \ldots, v'_n)$. Wir müssen daher jeweils nur eine Richtung zeigen.
a): Seien $a_1, \ldots, a_n \in K$ mit $a_1 v'_1 + \ldots + a_n v'_n = \mathbf{0}$ und etwa $v'_1 = v_1 + \lambda v_2$. Dann erhält man durch Einsetzen $a_1 v_1 + (a_2 + \lambda a_1) v_2 + a_3 v_3 + \ldots + a_n v_n$. Ist $(v_1, \ldots, v_n)$ linear unabhängig, so folgt zunächst $a_1 = a_3 = \ldots = a_n = 0$ und dann auch $a_2 = 0$, also folgt die lineare Unabhängigkeit von $(v'_1, \ldots, v'_n)$. Genauso rechnet man die anderen Aussagen nach. □

Bemerkung. Lemma 4.12 gilt genauso für beliebige (nicht notwendig endliche) Familien von Vektoren.

Beispiel: Sind ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_p$ die Zeilen der $(p \times n)$-Matrix A über K, ${}^t\mathbf{z}'_1, \ldots, {}^t\mathbf{z}'_p$ die der Matrix A', die aus A durch elementare Umformungen hervorgeht, so geht (natürlich) $(\mathbf{z}'_1, \ldots, \mathbf{z}'_p)$ aus $(\mathbf{z}_1, \ldots, \mathbf{z}_p)$ durch elementare Umformungen hervor.

Lemma 4.13. *Sei V ein K-Vektorraum, $\emptyset \neq M \subseteq V$ eine Basis von V, $\mathbf{0} \neq w \in V$, $w = \sum_{u \in M} a(u)u$ mit $a \in K^{(M)}$ und $u_0 \in M$ mit $a_0 := a(u_0) \neq 0$.*
Dann ist $M' := (M \setminus \{u_0\}) \cup \{w\}$ eine Basis von V.

Beweis. Man ersetze in M zunächst den Vektor u_0 durch $u'_0 := a_0 u_0$ und anschließend (durch endlich viele Umformungen vom Typ i)) den Vektor u'_0 durch $w = \sum_{u \in M} a(u)u$. Insgesamt kann man also durch eine Abfolge elementarer Umformungen u_0 durch w ersetzen, also von M zu M' übergehen. Nach dem vorigen Lemma ist dann ebenso wie M auch M' eine Basis von V. □

Satz 4.14. (Austauschsatz) *Sei V ein K-Vektorraum, M eine Basis von V, $(w_1, \ldots, w_p) \in V^p$ ein linear unabhängiges Vektorsystem. Dann gibt es Elemente $u_1, \ldots, u_p$ von M, so daß $(M \setminus \{u_1, \ldots, u_p\}) \cup \{w_1, \ldots, w_p\}$ eine Basis von V ist.*
Insbesondere gilt:

a) $p \leq |M|$.
b) *Ist M (beliebiges) endliches Erzeugendensystem, so ist $|M|$ eine obere Schranke für die Elementanzahl linear unabhängiger Teilmengen.*
c) *(Basisergänzungssatz) Jede linear unabhängige Teilmenge von V lässt sich zu einer Basis erweitern.*

Beweis. Wir beweisen die erste Behauptung durch vollständige Induktion nach p:
Induktionsanfang: Für $p = 1$ folgt die Aussage direkt aus dem Lemma.
Induktionsannahme: Sei $p > 1$ und die Aussage für Systeme von $p - 1$ Vektoren bewiesen.
Induktionsschritt: Wir benutzen die Induktionsannahme um geeignete Vektoren $u_1, \ldots, u_{p-1} \in M$ durch $w_1, \ldots, w_{p-1}$ zu ersetzen, wir erhalten so eine neue Basis M' von V, der die Vektoren $w_1, \ldots, w_{p-1}$ angehören. Der Vektor w_p besitzt eine (eindeutige) Darstellung

$$w_p = \sum_{u \in M'} a(u)u$$

als Linearkombination der Vektoren der Basis M', und es gibt wenigstens ein $u =: u_p \in M' \setminus \{w_1, \ldots, w_{p-1}\} = M \setminus \{u_1, \ldots, u_{p-1}\}$, für das $a(u_p) \neq 0$ ist, da w_p wegen der vorausgesetzten linearen Unabhängigkeit von $w_1, \ldots, w_p$ nicht Linearkombination von $w_1, \ldots, w_{p-1}$ ist. Nach dem Lemma können wir u_p durch w_p ersetzen und erhalten die gesuchte Basis M'' mit $w_1, \ldots, w_p \in M''$.

Die restlichen Aussagen folgen direkt hieraus. Bei c) beachte man, dass wir im Moment die Aussage nur für eine endliche Menge linear unabhängiger Vektoren (bzw., wenn man den Begriff eingeführt hat, auch für abzählbar unendliche Mengen) beweisen können, im allgemeinen Fall benötigt man wieder das oben erwähnte Zorn'sche Lemma. □

Beispiel: Im $\mathbb{R}^3$ bilden die Vektoren

$$u_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \; u_2 = \begin{pmatrix} -1 \\ -1 \\ 4 \end{pmatrix}, \; u_3 = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}$$

eine Basis (das rechnen wir später nach). Wir wollen zwei der Vektoren durch

$$w_1 = \begin{pmatrix} 3 \\ 5 \\ -8 \end{pmatrix}, w_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

ersetzen.
Zunächst stellen wir w_1 als Linearkombination der Basis dar:

$$w_1 = 2u_1 - u_2$$

(dazu löst man ein lineares Gleichungssystem $a_1u_1 + a_2u_2 + a_3u_3 = w_1$). Als nächster Schritt wähle man ein $i \in \{1, 2, 3\}$, so dass der Koeffizient a_i bei u_i nicht Null ist, also $i = 1$ oder $i = 2$. Nach Lemma 4.13 ist dann

$$\{w_1, u_2, u_3\} \quad \text{und ebenso} \quad \{w_1, u_1, u_3\}$$

eine Basis von $\mathbb{R}^3$, wir fahren mit der erstgenannten fort, wollen jetzt also einen der Vektoren u_2, u_3 durch w_2 ersetzen.
Dazu stellen wir w_2 in der Basis $\{w_1, u_2, u_3\}$ dar:

$$w_2 = \frac{1}{2}w_1 + \frac{3}{2}u_2$$

Der Koeffizient bei (wenigstens) einem der verbleibenden u_i ist nicht Null, in diesem Fall bei u_2. Nach Lemma 4.13 ist dann

$$\{w_1, w_2, u_3\}$$

eine Basis des $\mathbb{R}^3$.

Bemerkung. Das in den Beispielen praktizierte Verfahren für den Austausch lässt sich wie folgt als algorithmisches Verfahren formulieren, das den Austausch explizit macht (Gauß-Elimination).
Sei die Basis $M = \{u_1, \ldots, u_n\}$ gegeben, schreibe

$$w_i = \sum_{j=1}^{n} a_{ij}u_j \quad (1 \leq i \leq p)$$

mit Koeffizienten $a_{ij} \in K$; die a_{ij} bilden eine $p \times n$-Matrix $A = (a_{ij})$.
Elementare Umformungen des Vektorsystems $(w_1, \ldots, w_p)$ entsprechen elementaren Zeilenumformungen der Matrix: Ersetzt man etwa w_i durch

$w_i + \lambda w_k (i \neq k, \lambda \in K)$, so ersetzt man in der Matrix die i-te Zeile durch die Summe der i-ten Zeile mit der mit λ multiplizierten k-ten Zeile, analog für die anderen Umformungstypen.
Man bringe nun die Matrix A durch elementare Zeilenumformungen auf reduzierte Zeilenstufenform $A' = (a'_{ij})$. Diese Umformungen bringen die Gleichungen

$$w_i = \sum_{j=1}^{n} a_{ij} u_j$$

in die Gestalt

$$w'_i = \sum_{j=1}^{n} a'_{ij} u_j,$$

wo das p-Tupel $(w'_1, \ldots, w'_p)$ aus $(w_1, \ldots, w_p)$ durch elementare Umformungen hervorgeht. Spaltenvertauschungen in der Matrix entsprechen hier Umnummerierungen der u_j; daher können wir die Matrix (a'_{ij}) nach geeignetem Umnummerieren der u_j in die Gestalt bringen:

$$\left(\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & * \\ 0 & & 1 & \\ \hline 0 & & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & & \cdots & 0 \end{array}\right)$$

Das liefert das umgeformte Gleichungssystem

$$\begin{aligned} w'_i &= u_i + \textstyle\sum_{j=i+1}^{n} a'_{ij} u_i && (1 \leq i \leq r) \\ w'_i &= \mathbf{0} && (r < i \leq p), \end{aligned}$$

also $r = p$ wegen der linearen Unabhängigkeit der w_i und damit der w'_i.
Man sieht jetzt direkt, dass man $u_1, \ldots, u_p$ durch $w'_1, \ldots, w'_p$ ersetzen kann, dass also auch $(w'_1, \ldots, w'_p, u_{p+1}, \ldots, u_n)$ eine Basis ist.
Da $(w'_1, \ldots, w'_p)$ aus $(w_1, \ldots, w_p)$ durch elementare Umformungen hervorgeht, ist auch $(w_1, \ldots, w_p, u_{p+1}, \ldots, u_n)$ eine Basis.
(Die Auswahl der Auszutauschenden unter den u_j versteckt sich hier in den Spaltenvertauschungen.)

Der folgende Satz ist eine direkte Folgerung:

Satz 4.15. *Sei V ein K-Vektorraum. Dann gilt: Ist V endlich erzeugt, so sind alle Basen von V endlich und haben die gleiche Anzahl von Elementen.*

Wir können jetzt die Dimension eines Vektorraums definieren:

Definition 4.16. *Sei V ein K-Vektorraum. Ist V endlich erzeugt, so heißt die Anzahl der Elemente einer Basis von V die* Dimension

dim $V = \dim_K V$ *von* V. *Ist* V *nicht endlich erzeugt, so heißt* V unendlich dimensional.

Bemerkung. a) Die Dimension des Nullraums $\{\mathbf{0}\}$ ist Null.

b) Im nicht endlich erzeugten Fall kann man zeigen, daß zwei Basen eines K-Vektorraumes V die gleiche Mächtigkeit haben, d.h., bijektiv aufeinander abgebildet werden können.

Korollar 4.17. *Sei* V *ein endlich erzeugter* K*-Vektorraum,* $U \subseteq V$ *ein Unterraum von* V. *Dann ist* $\dim\ U \leq \dim\ V$, *und genau dann gilt* $\dim\ U = \dim\ V$, *wenn* $U = V$ *ist.*

Beweis. Da eine Basis von U insbesondere linear unabhängig ist, kann sie nach dem Austauschsatz 4.14 zu einer Basis von V ergänzt werden. Daraus folgt sofort die Behauptung. □

Beispiel:

a) Der K-Vektorraum K^n über dem Körper K hat die Dimension n.

b) Im euklidischen Raum $\mathbb{R}^3$ gibt es Unterräume der Dimensionen 0 (Nullraum $\{\mathbf{0}\}$), 1 (Geraden durch den Ursprung), 2 (Ebenen durch den Ursprung) und 3 (der ganze Raum). Auch anschaulich ist hier klar, dass es zwischen Unterräumen der gleichen Dimension keine echten Inklusionen (d.h., Enthaltensein mit ausgeschlossener Gleichheit) geben kann.

c) Ist $A \in M(p \times n, K)$ eine Matrix mit p Zeilen und n Spalten, deren Zeilenstufenform genau $p - r$ Nullzeilen hat, so hat der Lösungsraum $\mathcal{L}_0$ des linearen Gleichungssystems $A\mathbf{x} = \mathbf{0}$ eine Basis aus $n - r$ Fundamentallösungen. Wir wissen jetzt, dass dann jede Basis aus $n - r$ Vektoren besteht, die Anzahl der Fundamentallösungen also unabhängig davon ist, wie man die Matrix auf Zeilenstufenform gebracht hat.

d) Der K-Vektorraum $K^{(\mathbb{N}_0)}$ hat keine endliche Basis, denn sind $f_1, \ldots, f_r$ endlich viele Vektoren aus $K^{(\mathbb{N}_0)}$, so gibt es ein $N \in \mathbb{N}_0$ mit $f_i(n) = 0$ für alle $n > N$ und $1 \leq i \leq r$; ein Element f mit (zum Beispiel) $f(N + 1) = 1$ kann also nicht Linearkombination dieser f_i sein.

Die (unendlich vielen) Vektoren

$$h_i : \mathbb{N}_0 \longrightarrow K \quad \text{mit } h_i(j) = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

($i \in \mathbb{N}_0$ beliebig) bilden eine Basis:
Ist $f \in K^{(\mathbb{N}_0)}$, so gibt es nach Definition von $K^{(\mathbb{N}_0)}$ ein $n_0 \in \mathbb{N}_0$, so dass $f(n) = 0$ für alle $n > n_0$ gilt (man setze $n_0 = \max(\{n \in \mathbb{N}_0 \mid f(n) \neq 0\}$, dieses Maximum existiert, weil die Menge $\{n \in \mathbb{N}_0 \mid f(n) \neq 0\}$ endlich ist).

Dann ist $f = \sum_{i=0}^{n} f(i)h_i$, und diese Darstellung von f als Linearkombination von endlich vielen der h_i ist offenbar eindeutig.

e) Auch der Vektorraum $K^{\mathbb{N}_0}$ hat keine endliche Basis. Man weiss zwar, dass er eine unendliche Basis besitzt, kann aber keine solche Basis explizit angeben.

Korollar 4.18. *Sei V ein K-Vektorraum der endlichen Dimension n. Dann gilt:*

a) *Eine linear unabhängige Teilmenge von V ist genau dann eine Basis von V, wenn sie n Elemente hat.*
b) *Ein Erzeugendensystem von V ist genau dann eine Basis von V, wenn es n Elemente hat.*
c) *n ist die maximale Anzahl linear unabhängiger Vektoren in V und die minimale Elementanzahl von Erzeugendensystemen in V.*

Beweis. Übung! □

Bemerkung. Mit dem Korollar wird es im Prinzip leichter, nachzuprüfen, ob ein gegebenes n-Tupel $(v_1, \ldots, v_n)$ von Vektoren eines K-Vektorraums V der Dimension n eine Basis ist, da man nur eine der Eigenschaften „lineare Unabhängigkeit“ und „Erzeugendensystem“ nachprüfen muss. Am Beispiel des K^n sieht man allerdings, dass man nicht wirklich Arbeit spart: Um etwa im K^n lineare Unabhängigkeit der Vektoren

$$\mathbf{s}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \ldots, \mathbf{s}_n = \begin{pmatrix} a_{n1} \\ \vdots \\ a_{nn} \end{pmatrix}$$

nachzuprüfen, muss man prüfen, ob das homogene lineare Gleichungssystem $A\mathbf{x} = \mathbf{0}$ eine nichttriviale Lösung hat, um nachzuprüfen, ob es ein Erzeugendensystem bildet, muss man nachprüfen, ob das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ für alle $\mathbf{b}$ lösbar ist. Beide Aufgaben lassen sich mit dem Gauß'schen Eliminationsverfahren leicht lösen und liefern das gleiche Ergebnis: Beide Bedingungen sind genau dann erfüllt, wenn in der Zeilenstufenform keine Nullzeile vorkommt ($r = n$ gilt).

Der Dimensionsbegriff lässt sich auf lineare Gleichungssysteme anwenden:

Definition 4.19. *Sei $A \in M(p \times n, K)$ eine $(p \times n)$-Matrix mit Zeilen ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_p \in K^n$ und Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n \in K^p$. Dann ist $\dim\langle \mathbf{z}_1, \ldots, \mathbf{z}_p \rangle =: \mathrm{rg}(A)$ der* Rang *(*Zeilenrang*) von A, $\dim\langle \mathbf{s}_1, \ldots, \mathbf{s}_n \rangle$ der* Spaltenrang *von A.*

Korollar 4.20. a) *Ist $A \in M(p \times n, K)$ und $r = r(\tilde{A})$ die Anzahl der Zeilen, die in einer Zeilenstufenform $\tilde{A}$ von A nicht identisch 0 sind, so ist $r = \mathrm{rg}(A)$. Insbesondere ist r unabhängig von der Art und Weise, in der die Zeilenstufenform erreicht wurde.*

b) *Der Lösungsraum des homogenen Gleichungssystems* $A\mathbf{x} = \mathbf{0}$ *hat Dimension* $n - \mathrm{rg}(A)$.

Bemerkung. Ist die Matrix $\tilde{A}$ in Zeilenstufenform, so hat der Unterraum

$$\{\mathbf{b} \in K^p \mid \tilde{A}\mathbf{x} = \mathbf{b} \text{ hat Lösung}\}$$

die Basis $(\mathbf{e}_1, \ldots, \mathbf{e}_r)$, wo die $\mathbf{e}_i$ die Standardbasis des K^p bilden, er hat also Dimension r. Hier ist also Zeilenrang = Spaltenrang. Wir werden gleich sehen, dass diese Gleichheit allgemein gilt.

Lemma 4.21. *Sei* $f : V \longrightarrow W$ *eine lineare Abbildung von* K*-Vektorräumen. Dann gilt:*

a) *Ist* f *bijektiv, so ist auch die Umkehrabbildung* f^{-1} *linear.* f *und* f^{-1} *heißen dann Isomorphismen von* K*-Vektorräumen und man sagt, die* K*-Vektorräume* V *und* W *seien isomorph. Ist* $V = W$*, so spricht man auch von* Automorphismen.
b) f *ist genau dann ein Isomorphismus, wenn* f *Basen von* V *bijektiv auf Basen von* W *abbildet.*
c) *Ist* f *ein Isomorphismus, so ist* $\dim(V) = \dim(W)$.
d) *Ist* $\dim(V) = \dim(W)$ *endlich, so sind* V *und* W *isomorph.*

Beweis. a) Seien $w_1 = f(v_1), w_2 = f(v_2) \in W, \lambda \in K$. Dann ist

$$\begin{aligned} f^{-1}(w_1 + \lambda w_2) &= f^{-1}(f(v_1) + \lambda f(v_2)) \\ &= f^{-1}(f(v_1 + \lambda v_2)) \\ &= v_1 + \lambda v_2 \\ &= f^{-1}(w_1) + \lambda f^{-1}(w_2). \end{aligned}$$

b) Sei f bijektiv, M eine Basis von V und $N := f(M)$ ihr Bild unter f.
Ist dann $w \in W$, so gibt es, weil f surjektiv ist, ein $v \in V$ mit $f(v) = w$. Drückt man v als Linearkombination

$$v = \sum_{u \in M} a(u)u$$

der Vektoren der Basis M aus, so ist $w = f(v) = \sum_{u \in M} a(u)f(u)$. Der (beliebig gewählte) Vektor $w \in W$ ist also Linearkombination der Vektoren $f(u) \in N$, wir sehen also, dass N den Vektorraum V erzeugt.
Sind $w_1 = f(u_1), \ldots, w_r = f(u_r)$ paarweise verschiedene Vektoren aus N und $a_1, \ldots, a_r \in K$ mit $\sum_{i=1}^r a_i w_i = \mathbf{0}$, so ist $\sum_{i=1}^r a_i u_i \in \mathrm{Ker}(f) = \{\mathbf{0}\}$, und wegen der linearen Unabhängigkeit von M gilt $a_1 = \ldots = a_r = 0$. Also ist N auch linear unabhängig und damit eine Basis.
Ist umgekehrt M eine Basis von V, die durch f bijektiv auf die Basis $f(M) = N$ von W abgebildet wird, so ist f insgesamt bijektiv:

Ist $w = \sum_{n\in N} a(n)n = \sum_{m\in M} a(f(m))f(m) \in W$, so setzt man $b(m) := a(f(m)$ für die $m \in M$ und hat $w = f(\sum_{m\in M} b(m)m) \in f(V)$, die Abbildung f ist also surjektiv.
Ist $v = \sum_{m\in M} a(m)m \in \mathrm{Ker}(f)$, so ist $\mathbf{0} = \sum_{m\in M} a(m)f(m)$, und da die $f(m)$ paarweise verschiedene Vektoren der linear unabhängigen Menge $N = f(M)$ sind, ist $a(m) = 0$ für alle $m \in M$ und daher $v = \mathbf{0}$. Die Abbildung f ist also auch injektiv.

c) folgt direkt aus b).

d) Ist $(v_1, \ldots, v_n)$ eine Basis von V und $(w_1, \ldots, w_n)$ eine Basis von W, so definiert man eine Abbildung $f : V \to W$ durch $f(\sum_{i=1}^n a_i v_i) = \sum_{i=1}^n a_i w_i$. Man rechnet leicht nach, dass f linear und bijektiv ist.

□

Bemerkung. Eine leichte Modifikation dieses Beweises zeigt ein klein wenig mehr:

a) Ist f surjektiv, so gilt für jedes Erzeugendensystem M von V, dass $f(M)$ ein Erzeugendensystem von W ist.

b) Ist f injektiv so gilt für jedes linear unabhängige Vektorsystem $(v_1, \ldots, v_r) \in V^r$, in V, dass auch $(f(v_1), \ldots f(v_r)$ linear unabhängig ist.

c) Gibt es eine Basis $(v_1, \ldots, v_n)$ von V, für die $(f(v_1), \ldots f(v_n))$ linear unabhängig ist, so ist f injektiv.

d) Gibt es ein Erzeugendensystem M von V, für das $f(M)$ ein Erzeugendensystem von W ist, so ist f surjektiv.

Mit a) und b) folgt also in c) aus der Gültigkeit der Bedingung für *eine* Basis ihre Gültigkeit für *alle* Basen und in d) aus der Gültigkeit der Bedingung für *ein* Erzeugendensystem ihre Gültigkeit für *alle* Erzeugendensysteme.

Korollar 4.22. *Sei $A \in M(p \times n, K)$ eine $p \times n$-Matrix mit Einträgen aus dem Körper K. Dann ist der Zeilenrang von A gleich dem Spaltenrang von A.*

Beweis. Jede elementare Zeilenumformung (und damit auch jede Abfolge elementarer Zeilenumformungen) definiert, angewendet auf Vektoren $\mathbf{x} \in K^p$, eine umkehrbare Abbildung $f : K^p \longrightarrow K^p$, deren Linearität man sofort nachprüft; die Abbildung f ist also ein Isomorphismus von K^p auf sich.
Formt man A durch elementare Umformungen in eine Matrix A' in reduzierter Zeilenstufenform mit $r = \mathrm{rg}(A) = \mathrm{rg}(A')$ um, so ist, wie oben bemerkt, r die Dimension von

$$U' := \{\mathbf{b}' \in K^p \mid A'\mathbf{x} = \mathbf{b}' \text{ hat eine Lösung}\}.$$

Ist $f : K^p \longrightarrow K^p$ der zu der Abfolge elementarer Umformungen gehörige Isomorphismus, so ist offenbar $U' = f(U)$ mit

$$U = \{\mathbf{b} \in K^p \mid A\mathbf{x} = \mathbf{b} \text{ hat eine Lösung}\},$$

der Unterraum $U \subseteq K^p$ hat also die gleiche Dimension $r = \mathrm{rg}(A)$ wie U'. Da die Dimension von U der Spaltenrang von A ist, folgt die Behauptung. $\square$

Bemerkung. Die Bezeichnungen „Zeilenrang“ und „Spaltenrang“ haben sich damit als überflüssig erwiesen und werden im Weiteren nicht mehr vorkommen, man spricht nur noch vom Rang einer Matrix, der dann gleichzeitig die Maximalanzahl linear unabhängiger Zeilen und die Maximalanzahl linear unabhängiger Spalten der Matrix ist.

Wir halten noch ein häufig benutztes Kriterium für die Lösbarkeit eines linearen Gleichungssystems fest:

Korollar 4.23. *Sei K ein Körper, seien $A \in M(p \times n, K), \mathbf{b} \in K^p$. Dann gilt: Das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ ist genau dann lösbar, wenn die Matrix A und die erweiterte Matrix $(A \mid \mathbf{b}) \in M(p \times (n+1), K)$ den gleichen Rang haben.*

Beweis. Das Gleichungssystem ist genau dann lösbar, wenn $\mathbf{b}$ zur linearen Hülle der Spaltenvektoren von A gehört, wenn also der von den Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$ von A erzeugte Unterraum $< \mathbf{s}_1, \ldots, \mathbf{s}_n >$ des K^p gleich $< \mathbf{s}_1, \ldots, \mathbf{s}_n, \mathbf{b} >$ ist, Das ist aber äquivalent zur Gleichheit der Dimensionen und damit zur Gleichheit der Ränge von A und der erweiterten Matrix $(A \mid \mathbf{b})$. $\square$

Definition und Satz 4.24. *Sei V ein K-Vektorraum der endlichen Dimension n, $\mathcal{B} = (v_1, \ldots, v_n)$ eine (geordnete) Basis von V. Dann sind die Abbildungen*

$$\begin{array}{rccc} f_{\mathcal{B}}: & K^n & \longrightarrow & V \\ & \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} & \longmapsto & \sum_{j=1}^n a_j v_j \end{array}$$

und

$$\begin{array}{rccc} c_{\mathcal{B}}: & V & \longrightarrow & K^n \\ & \sum_{j=1}^n a_j v_j & \longmapsto & \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \end{array}$$

zueinander inverse bijektive lineare Abbildungen (Isomorphismen von K-Vektorräumen). $c_{\mathcal{B}}$ heißt die Koordinatenabbildung *bezüglich der Basis $\mathcal{B}$.*

Beweis. Die Abbildung $c_{\mathcal{B}}$ ist wohldefiniert, da $\mathcal{B}$ eine Basis von V ist und daher jeder Vektor $v \in V$ eine eindeutige Darstellung $v = \sum_{j=1}^n a_j v_j$ hat. Dass $f_{\mathcal{B}}$ und $c_{\mathcal{B}}$ linear und zueinander invers sind, rechnet man sofort nach. $\square$

Bemerkung. a) Ist $V = K^n$ und $\mathcal{B}$ die Standardbasis, so sind $c_{\mathcal{B}}$ und $f_{\mathcal{B}}$ beide die identische Abbildung Id_V.

b) Wir wissen bereits, dass jeder endlich dimensionale K-Vektorraum für $n = \dim_K(V)$ zu K^n isomorph ist. Der Satz liefert für jede Basis $\mathcal{B}$ durch $f_{\mathcal{B}}$ eine Parametrisierung von V. Der intuitive Begriff „die Dimension ist die Anzahl der freien Parameter (Freiheitsgrade)“ wird hierdurch gerechtfertigt.

c) Umgekehrt kann man bei Vorgabe einer Basis $\mathcal{B}$ in V durch $c_{\mathcal{B}}$ Koordinaten bezüglich dieser Basis einführen. Die Basis braucht dabei keine zusätzlichen Eigenschaften wie Rechtwinkligkeit (die ja auch in einem beliebigen Vektorraum gar nicht definiert ist) zu erfüllen. Auch im K^n kann man von der Standardbasis abweichende Basen angeben (das wird in den Übungen geschehen) und dann die Koordinaten eines Vektors $\mathbf{x}$ bezüglich dieser Basis einführen, diese werden im allgemeinen natürlich von seinen Koordinaten x_i bezüglich der Standardbasis abweichen.

d) Analog kann man zeigen, dass ein nicht endlich erzeugter Vektorraum mit Basis $\mathcal{B}$ isomorph zu $K^{(\mathcal{B})}$ ist und dass der Isomorphietyp dieses Vektorraums nur von der Mächtigkeit von $\mathcal{B}$ abhängt.

Satz 4.25. (Dimensionsformel für Unterräume) *Sei V ein K-Vektorraum, seien U_1, U_2 zwei Unterräume von V. Dann gilt*

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim\ U_1 + \dim\ U_2.$$

Beweis. Wir beweisen das im Fall, dass V endlich erzeugt ist.
Sei $\{u_1, \ldots, u_r\}$ eine Basis von $U_1 \cap U_2$. Wir egänzen sie durch Vektoren $v_1, \ldots, v_s$ zu einer Basis von U_1 und durch Vektoren $w_1, \ldots, w_t$ zu einer Basis von U_2, haben also $r+s = \dim(U_1), r+t = \dim(U_2)$. Die Vektoren $u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t$ erzeugen offenbar den Raum $U_1 + U_2$. Sie sind linear unabhängig, denn ist

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{s} b_j v_j + \sum_{k=1}^{t} c_k w_k = \mathbf{0}$$

(mit $a_i, b_j, c_k \in K$), so ist

$$v := \sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{s} b_j v_j = -\sum_{k=1}^{t} c_k w_k \in U_1 \cap U_2,$$

kann also als Linearkombination $\sum_{i=1}^{r} a_i' u_i$ geschrieben werden. Da aber $(u_1, \ldots, u_r, v_1, \ldots, v_s)$ eine Basis von U_1 ist, müssen die beiden Schreibweisen $v = \sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{s} b_j v_j$ und $v = \sum_{i=1}^{r} a_i' u_i$ übereinstimmen, das heißt, die b_j sind gleich Null und $a_i = a_i'$ gilt für alle i. Damit ist dann aber

$$\sum_{i=1}^{r} a_i u_i + \sum_{k=1}^{t} c_k w_k = \mathbf{0}$$

eine lineare Relation zwischen den Basisvektoren $u_1, \ldots, u_r, , w_1, \ldots, w_t$ von U_2, also sind auch alle a_i und alle c_k gleich Null, was die behauptete lineare Unabhängigkeit zeigt.
Die

$$r + s + t = (r + s) + (r + t) - r = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2)$$

Vektoren $u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t$ bilden also eine Basis von $U_1 + U_2$, was die Behauptung beweist. □

Beispiel:

a) Eine Ursprungsgerade U_1 im $\mathbb{R}^3$, die nicht in der Ursprungsebene U_2 liegt ($U_1 \cap U_2 = \{\mathbf{0}\}$), spannt mit ihr zusammen den $\mathbb{R}^3$ auf ($\dim(U_1 + U_2) = 3$).
b) Zwei verschiedene Ebenen U_1 und U_2 durch den Ursprung im $\mathbb{R}^3$ schneiden sich in einer Geraden ($\dim(U_1 \cap U_2) = 1$) und spannen zusammen den $\mathbb{R}^3$ auf ($\dim(U_1 + U_2) = 3$).

Definition und Lemma 4.26. *Sind U_1, U_2 Unterräume des K-Vektoraums V mit $U_1 \cap U_2 = \{\mathbf{0}\}$, so heißt ihre Summe $U_1 + U_2$ auch eine* direkte Summe*; man schreibt $U_1 \oplus U_2$.*
Es gilt dann $\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2$, und jeder Vektor $v \in U_1 \oplus U_2$ lässt sich auf genau eine Weise als $v = u_1 + u_2$ mit $u_1 \in U_1$, $u_2 \in U_2$ schreiben.
Ferner gilt in diesem Fall: Ist $(v_1, \ldots, v_r)$ eine Basis von U_1, $(w_1, \ldots, w_s)$ eine Basis von U_2, so ist $(v_1, \ldots, v_r, w_1, \ldots, w_s)$ eine Basis von $U_1 \oplus U_2$.
Ist $U_1 \oplus U_2 = V$, so heißen U_1 und U_2 auch zueinander komplementär.

Beweis. Übung. □

Lemma 4.27. *Ist $U \subseteq V$ ein Unterraum des K-Vektorraums V, so gibt es einen zu U komplementären Unterraum U' von V.*

Beweis. Man ergänze (im endlich erzeugten Fall) eine Basis $\{u_1, \ldots, u_r\}$ von U durch Vektoren $u'_1, \ldots, u'_s$ zu einer Basis von V. Der Raum $U' := < u'_1, \ldots, u'_s >$ ist dann ein zu U komplementärer Unterraum. □

Bemerkung. Im allgemeinen gibt es viele verschiedene Möglichkeiten, einen komplementären Unterraum zu einem gegebenen Unterraum $U \subseteq V$ zu finden.
Ist etwa $U \subseteq \mathbb{R}^3$ eine Ebene durch den Ursprung, so sind alle Geraden durch den Ursprung, die nicht in dieser Ebene liegen, komplementäre Unterräume (und nicht etwa nur die eine, die senkrecht auf der Ebene steht).

Definition und Lemma 4.28. *Sei V ein K-Vektorraum mit Unterräumen U_1, U_2, so dass $V = U_1 \oplus U_2$ gilt.*
Dann werden durch

$$p_1(u_1 + u_2) = u_1, \quad p_2(u_1 + u_2) = u_2$$

lineare Abbildungen $p_1, p_2 : U \to V$ *definiert mit* $\mathrm{Im}(p_i) = U_i$ *(für* $i = 1, 2$*) und* $p_i \circ p_i = p_i \quad (i = 1, 2)$.
Die Abbildungen p_1, p_2 *heißen die Projektionen von* V *längs (entlang)* U_2 *auf* U_1 *bzw. längs (entlang)* U_1 *auf* U_2.

Beweis. Da jeder Vektor $v \in V$ wegen $V = U_1 \oplus U_2$ eindeutig als $v = u_1 + u_2$ mit $u_1 \in U_1, u_2 \in U_2$ geschrieben werden kann, sind p_1, p_2 wohldefinierte Abbildungen. Die Linearität rechnet man leicht nach: Für $i = 1, 2$ gilt

$$\begin{aligned} p_i((u_1 + u_2) + \lambda(u_1' + u_2')) &= p_i((u_1 + \lambda u_1') + (u_2 + \lambda u_2')) \\ &= u_i + \lambda u_i' \\ &= p_i(u_1 + u_2) + \lambda p_i(u_1' + u_2'). \end{aligned}$$

Wegen $p_i(p_i(u_1 + u_2)) = p_i(u_i + \mathbf{0}) = u_i = p_i(u_1 + u_2)$ gilt auch $p_i \circ p_i = p_i$. □

Bemerkung. Wählt man einen anderen zu (fest gehaltenem) U_1 komplementären Unterraum U_2' (also $V = U_1 \oplus U_2 = U_1 \oplus U_2'$ mit $U_2' \neq U_2$), so ändern sich *beide* Projektionen p_1 und p_2 (und nicht etwa nur p_2). Man prüfe das etwa am Beispiel $V = \mathbb{R}^2$, $U_1 = \{\left(\begin{smallmatrix} a \\ 0 \end{smallmatrix}\right) \mid a \in \mathbb{R}\}$ $U_2 = \{\left(\begin{smallmatrix} a \\ a \end{smallmatrix}\right) \mid a \in \mathbb{R}\}$, $U_2' = \{\left(\begin{smallmatrix} a \\ -a \end{smallmatrix}\right) \mid a \in \mathbb{R}\}$ nach.

Satz 4.29 (Dimensionsformel für Kern und Bild)**.** *Seien* V, W *Vektorräume über dem Körper* K*, sei* $f : V \to W$ *eine lineare Abbildung und* $U \subseteq V$ *ein zu* $\mathrm{Ker}(f)$ *komplementärer Unterraum von* V.
Dann ist $f|_U : U \to f(V) \subseteq W$ *ein Isomorphismus.*
Insbesondere gilt

$$\begin{aligned} \dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f)) &= \dim(V), \\ \dim(\mathrm{Im}(f)) &\leq \dim(V). \end{aligned}$$

Beweis. Da $f(\mathrm{Ker}(f)) = \{\mathbf{0}\}$ gilt, folgt $f(U) = f(V)$ aus $U \oplus \mathrm{Ker}(f) = V$, damit ist $f|_U : U \to f(V)$ surjektiv. Ferner ist $\mathrm{Ker}(f|_U) = \mathrm{Ker}(f) \cap U = \{\mathbf{0}\}$, also ist $f|_U$ injektiv.
Die Aussage über die Dimensionen ist dann klar. □

Bemerkung. Als Übung zeige man, dass man im endlich erzeugten Fall einen zu $\mathrm{Ker}(f)$ komplementären Unterraum U wie im Satz konstruieren kann, indem man eine Basis $(w_1, \ldots, w_r)$ von $\mathrm{Im}(f)$ wählt, die w_i als $w_i = f(u_i)$ mit $u_i \in V$ schreibt und $U =< u_1, \ldots, u_r >= \mathrm{Lin}(\{u_1, \ldots, u_r\}$ als die Menge aller Linearkombinationen der u_i definiert. Im allgemeinen Fall geht man genauso vor, nur schreibt man dann die Basis von $\mathrm{Im}(f)$ als Familie $(w_i)_{i \in I}$ mit einer Indexmenge I und $U = \mathrm{Lin}(\{u_i \mid i \in I\}$.

Bemerkung. Da Isomorphismen lineare Unabhängigkeit ebenso erhalten, wie die Eigenschaft, ein Erzeugendensystem zu sein, können wir die Grundaufgaben der linearen Algebra in beliebigen endlich dimensionalen Vektorräumen mit Hilfe der Koordinatenabbildung $c_{\mathcal{B}}$ und

ihrer Umkehrung, der Parametrisierung $f_{\mathcal{B}}$, nach Wahl einer Basis in den „Standardraum“ K^n übertragen.
Ein Beispiel für dieses Verfahren haben wir bereits im Beispiel und der Bemerkung nach Satz 4.14 gesehen: Gegeben seien eine Basis $(v_1, \ldots, v_n)$ von V und linear unabhängige Vektoren $u_1, \ldots, u_r$.
Man finde Indizes $j_1, \ldots, j_r$, so dass $(v_1, \ldots, v_n)$ bei Austausch von $v_{j_1}, \ldots, v_{j_r}$ durch $u_1, \ldots, u_r$ eine neue Basis von V liefert.
Dort hatten wir gezeigt, wie man dieses Problem in eine Aufgabe über Matrizen umwandelt und dieses mit dem Gauß-Algorithmus löst.

Wir stellen jetzt noch einmal zusammen, wie man im K^n den Gauß-Algorithmus, mit dessen Hilfe man eine Matrix bzw. ein lineares Gleichungssystem auf Zeilenstufenform bringt, zur Lösung einiger Grundaufgaben in algorithmischer Weise benutzen kann:

a) Gegeben seien Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_r \in K^n$.
Man finde eine Basis von $\langle \mathbf{a}_1, \ldots, \mathbf{a}_r \rangle$!
Lösung: Man betrachte die Matrix mit den Zeilen ${}^t\mathbf{a}_1, \ldots, {}^t\mathbf{a}_r$. Man forme sie in Zeilenstufenform um, die nicht verschwindenden Zeilen bilden eine Basis.

b) Gegeben Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_n \in K^p$, $U = \langle \mathbf{a}_1, \ldots, \mathbf{a}_n \rangle$.
Man finde ein lineares Gleichungssystem, dessen Lösungsraum U ist.
Lösung: Man betrachte die Matrix $A \in M(p \times n, K)$, deren Spalten $\mathbf{a}_1, \ldots, \mathbf{a}_n$ sind. U ist dann die Menge aller $\mathbf{b} \in K^p$, für die das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ lösbar ist.
Man bringe die erweiterte Matrix $(A|\mathbf{b}) \in M(p \times (n+1), K)$ durch elementare Zeilenumformungen auf Zeilenstufenform $(A'|\mathbf{b}')$, bei der alle Zeilen von A' ab der $(r+1)$-ten nur Nullen enthalten ($r = r(A) =$ Rang von A). Man hat dann

$$b'_i = \sum_{j=1}^{p} c_{ij} b_j \quad \text{für } 1 \leq i \leq p$$

mit gewissen $c_{ij} \in K$, und $A\mathbf{x} = \mathbf{b}$ ist genau dann lösbar, wenn $b'_{r+1} = \cdots = b'_p = 0$ ist.
Das System aus den linearen Gleichungen

$$\sum_{j=1}^{p} c_{ij} b_j = 0 \quad (r < i \leq p)$$

in den Variablen $b_1, \ldots, b_p$ hat also genau den Lösungsraum U.

c) Gegeben seien Unterräume U_1, U_2 von $V = K^n$ (durch Angabe von Erzeugenden).
Man finde Basen von $U_1 \cap U_2$, $U_1 + U_2$.
Lösung: Nimmt man die im vorigen Punkt gewonnenen linearen Gleichungssysteme, deren Lösungsräume U_1 und U_2 sind, so ist

$U_1 \cap U_2$ genau die Menge aller Vektoren, die allen Gleichungen zusammen genügen. Das Gauß-Verfahren liefert eine Basis dieses Lösungsraumes. $U_1 + U_2$ ist der Raum, der von allen angegebenen Erzeugern zusammen aufgespannt wird; wie im ersten Punkt angegeben findet man eine Basis.

Anzumerken ist noch, dass es in Einzelfällen häufig schnellere und elegantere Lösungen als die hier angegebenen gibt.

Zusammenfassung. Dieser Abschnitt brachte die folgenden grundlegenden Definitionen in einem Vektorraum V über einem Körper K:

a) **Linearkombinationen und lineare Hülle, Erzeugendensystem:** Definition 4.2, Satz 4.3
b) **Linear unabhängig bzw. abhängig:** Definition und Satz 4.4
c) **Basis und Dimension:** Definition 4.8, Satz 4.10, Definition 4.16
d) **Koordinatenabbildung:** Definition und Satz 4.24
e) **Direkte Summe von Unterräumen** Definition und Lemma 4.26

Wichtige Sätze sind:

a) **Existenz einer Basis:** Satz 4.11
b) **Austauschsatz, alle Basen haben gleiche Elementanzahl:** Satz 4.14, Satz 4.15
c) **Linearität der Umkehrabbildung, Abbildung von Basen:** Korollar 4.21
d) **Gleichheit von Zeilenrang und Spaltenrang einer Matrix:** Korollar 4.22
e) **Rangkriterium für Lösbarkeit eines linearen Gleichungssystems** Korollar 4.23
f) **Dimensionsformeln für Summen von Unterräumen und für Kern und Bild einer linearen Abbildung** Satz 4.25, Satz 4.29

5. Lineare Abbildungen und Matrizen

Satz 5.1. *V und W seien Vektorräume über dem Körper K.*

a) *Ist $(u_1, \ldots, u_r)$ ein linear unabhängiges Vektorsystem in V und sind beliebige Vektoren $w_1, \ldots, w_r \in W$ gegeben, so gibt es wenigstens eine lineare Abbildung $f : V \longrightarrow W$ mit $f(u_i) = w_i$ für $1 \leq i \leq r$ (man nennt jede solche Abbildung* eine *lineare Fortsetzung der Vorgabe $u_i \mapsto w_i$).*

b) *Sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, $w_1, \ldots, w_n \in W$ beliebige Vektoren in W.*

Dann wird durch

$$f(\sum_{i=1}^{n} a_i v_i) = \sum_{i=1}^{n} a_i w_i, \quad (a_i \in K, 1 \leq i \leq n)$$

eine lineare Abbildung $f : V \longrightarrow W$ definiert; diese ist die eindeutig bestimmte lineare Abbildung $f : V \longrightarrow W$ mit $f(v_i) = w_i$ für $1 \leq i \leq n$, sie heißt die lineare Fortsetzung *der Vorgabe $v_i \mapsto w_i$ auf den Basisvektoren.*

c) *V, W seien K-Vektorräume, $f : V \longrightarrow W$ eine lineare Abbildung. Dann gilt:*

Ist $(v_1, \ldots, v_n) \in V^n$ linear abhängig, so ist auch $(f(v_1), \ldots, f(v_n))$ linear abhängig.

Genauer gilt: Sind $a_i \in K (1 \leq i \leq n)$ mit $\sum_{i=1}^{n} a_i v_i = 0$, so gibt es höchstens dann eine lineare Abbildung $f : V \longrightarrow W$ mit $f(v_i) = w_i$, wenn $\sum_{i=1}^{n} a_i w_i = 0$ gilt.

Beweis. Zu b): Wegen der eindeutigen Darstellbarkeit jedes Vektors aus V als Linearkombination der Basisvektoren wird durch

$$f(\sum_{j=1}^{n} a_j v_j) := \sum_{j=1}^{n} a_j w_j$$

eine wohldefinierte Abbildung $f : V \longrightarrow W$ angegeben, die offenbar die Bedingung $f(v_i) = w_i$ für $1 \leq i \leq r$ erfüllt und deren Linearität man leicht nachrechnet. Umgekehrt ist klar, dass $f(\sum_{j=1}^{n} a_j v_j) := \sum_{j=1}^{n} a_j w_j$ für jede lineare Abbildung $f : V \to W$ mit $f(v_i) = w_i$ für $1 \leq i \leq n$ gelten muss, es also auch nur eine lineare Abbildung mit dieser Eigenschaft gibt.

Zu a): Man ergänze $(u_1, \ldots, u_r)$ zu einer Basis von V. Ist V endlich dimensional, so kann man diese als $(u_1, \ldots, u_n)$ schreiben und findet nach b) eine lineare Abbildung $f : V \to W$ mit $f(u_i) = \begin{cases} w_i & 1 \leq i \leq r \\ \mathbf{0} & i > r \end{cases}$.

Im allgemeinen Fall schreibt man die Basis von V als $(u_i)_{i \in I}$ mit einer Indexmenge $I \supseteq \{1, \ldots, r\}$. Da sich b) ganz analog zum oben gegebenen Beweis auch für beliebige Basen zeigen lässt (Übung, siehe auch die nachfolgende Bemerkung), folgt dann auch a) genau wie oben.

c) folgt sofort aus $f(\sum_{i=0}^n a_i v_i) = \sum_{i=1}^n a_i w_i$, falls $f(v_i) = w_i$ gilt. $\square$

Beispiel: Ist $V = \mathbb{R}^2$, $W = \mathbb{R}^3$, $v_1, v_2, v_3 \in V$ beliebige Vektoren, (w_1, w_2, w_3) in W linear unabhängig, so gibt es keine lineare Abbildung $f : V \longrightarrow W$ mit $f(v_i) = w_i$ für $1 \leq i \leq 3$ (da die v_i und damit nach dem Lemma für lineares f auch die $f(v_i)$ notwendig linear abhängig sind).

Bemerkung. Die Aussagen des Satzes gelten genauso für beliebige, eventuell unendliche, Vektorsysteme bzw. Basen, mit bis auf die Notation unveränderten Beweisen.

Definition und Satz 5.2. *V und W seien K-Vektorräume (K ein Körper), $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, $\mathcal{B}' = (w_1, \ldots, w_p)$ eine Basis von W.*

a) *Ist $A \in M(p \times n, K)$ eine $p \times n$-Matrix, so gibt es genau eine lineare Abbildung $f = f_{\mathcal{B}'}^{\mathcal{B}}(A) : V \longrightarrow W$ mit*

$$f(v_j) = \sum_{i=1}^{p} a_{ij} w_i \quad \text{für } 1 \leq j \leq n.$$

Die Abbildung $f = f_{\mathcal{B}'}^{\mathcal{B}}(A)$ heißt die bezüglich der Basen $\mathcal{B}$ von V und $\mathcal{B}'$ von W zu A gehörige lineare Abbildung, sie entsteht durch lineare Fortsetzung *der Vorgabe $f(v_j) = \sum_i a_{ij} w_i$ für $1 \leq j \leq n$.*

Ist $V = W$ und $\mathcal{B} = \mathcal{B}'$, so schreibt man auch $f_{\mathcal{B}}(A)$.

b) *Ist umgekehrt $f : V \to W$ eine lineare Abbildung und sind*

$$f(v_j) = \sum_{i=1}^{p} a_{ij} w_i \quad (1 \leq j \leq n)$$

die Darstellungen der Bilder $f(v_j)$ der Vektoren v_j aus der Basis $\mathcal{B}$ von V als Linearkombinationen der Vektoren w_i aus der Basis $\mathcal{B}'$ von W, so heißt die aus den a_{ij} gebildete Matrix $A = (a_{ij}) \in M(p \times n, K)$ die Matrix $A = M_{\mathcal{B}'}^{\mathcal{B}}(f)$ von f bezüglich der Basen $\mathcal{B}$ von V und $\mathcal{B}'$ von W.

Ist $V = W$ und $\mathcal{B} = \mathcal{B}'$, so schreibt man auch $M_{\mathcal{B}}(f)$.

Beweis. Die Aussage in a) folgt direkt aus Satz 5.1 über die lineare Fortsetzung von Abbildungen, der Rest ist Definition. $\square$

Bemerkung. a) Man hat also (nach Wahl von Basen in Bild- und Urbildraum) eine einfache Beschreibung für lineare Abbildungen: Diese entsprechen genau den pn-Tupeln von Elementen im Grundkörper K, die als Einträge der zugeordneten Matrix auftreten. Allgemeine Abbildungen, die nicht die Linearitätsbedingung erfüllen, sind erheblich schwieriger zu beschreiben.

b) Die Korrespondenz zwischen linearen Abbildungen auf der einen Seite und Matrizen auf der anderen Seite hängt wesentlich von der Auswahl der Basen $\mathcal{B}, \mathcal{B}'$ in V, W ab, wir werden im nächsten

Abschnitt untersuchen, wie sie sich ändert, wenn man zu anderen Basen übergeht.

c) Die j-te Spalte der der Abbildung f zugeordneten Matrix ist der Koordinatenvektor von $f(v_j)$ bezüglich der Basis $\mathcal{B}'$.

d) Ist $V = K^n, W = K^p$ und sind $\mathcal{B}, \mathcal{B}'$ die jeweiligen Standardbasen, so hat man

$$f_{\mathcal{B}'}^{\mathcal{B}}(A) = L_A, \qquad M_{\mathcal{B}'}^{\mathcal{B}}(L_A) = A.$$

Beispiel:

a) $V = W = \mathbb{R}^2$, $\mathcal{B} = \mathcal{B}'$ beliebige Basis von V, $f(v) = rv$ $(0 < r \in \mathbb{R})$ die *Streckung* um den Faktor r:

$$M_{\mathcal{B}'}^{\mathcal{B}}(f) = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$$

(unabhängig von $\mathcal{B} = \mathcal{B}'$).

b) $V = W = \mathbb{R}^2$, $f = D_\varphi =$ *Drehung* um den Winkel φ im Gegenuhrzeigersinn, $\mathcal{B} = \mathcal{B}' =$ Standardbasis aus $\binom{1}{0} = \mathbf{e}_1$ und $\binom{0}{1} = \mathbf{e}_2$.
Man hat

$$M_{\mathcal{B}'}^{\mathcal{B}}(D_\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

(*Drehmatrix*).
Bei gleichzeitiger Streckung um den Faktor r hat man die Matrix

$$\begin{pmatrix} r\cos\varphi & -r\sin\varphi \\ r\sin\varphi & r\cos\varphi \end{pmatrix}$$

für die *Drehstreckung*.

c) $V = W = \mathbb{R}^2$, f sei die lineare Abbildung mit $f(\mathbf{e}_1) = r\mathbf{e}_1$, $f(\mathbf{e}_2) = \mathbf{e}_2$ (Streckung um den Faktor r in x-Richtung, *Scherung*) $(0 < r \in \mathbb{R})$. Bezüglich der Standardbasen hat f die Matrix $\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}$.

d) $V = W = \mathbb{R}^2$, $0 < r \in \mathbb{R}$, f sei die lineare Abbildung mit $f(\mathbf{e}_1 + \mathbf{e}_2) = r(\mathbf{e}_1 + \mathbf{e}_2)$, $f(\mathbf{e}_1 - \mathbf{e}_2) = \mathbf{e}_1 - \mathbf{e}_2$ (Streckung in Richtung der Geraden $y = x$ um den Faktor r).
Ist $\mathcal{B}_1 = \mathcal{B}_1' = \{\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_2\}$, so ist

$$M_{\mathcal{B}_1'}^{\mathcal{B}_1}(f) = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}.$$

Bezüglich der Standardbasis $\mathcal{B}_2 = \mathcal{B}_2' = \{\mathbf{e}_1, \mathbf{e}_2\}$ ist

$$M_{\mathcal{B}_2'}^{\mathcal{B}_2}(f) = \begin{pmatrix} \frac{r+1}{2} & \frac{r-1}{2} \\ \frac{r-1}{2} & \frac{r+1}{2} \end{pmatrix}.$$

In der Matrixbeschreibung bezüglich $\mathcal{B}_2$ sieht man die geometrischen Eigenschaften von f wesentlich schlechter als in der Beschreibung bezüglich $\mathcal{B}_1$.

e) Wie in Aufgabe 4 von Blatt 7 sei $K[X]$ der Vektorraum $K^{(\mathbb{N}_0)}$, in dem man für $i \in \mathbb{N}_0$ die Bezeichnung X^i für die Abbildung von $\mathbb{N}_0$ in K verwendet, die durch

$$X^i(k) = \delta_{ik} := \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$$

definiert wird. Die X^i bilden dann also eine Basis von $K[X]$, und wir können jeden Vektor $P \in K[X]$ eindeutig als $P = \sum_{i \in \mathbb{N}_0} a_i X^i$ schreiben, mit $a_i \in K$, so dass $\{i \in \mathbb{N}_0 \mid a_i \neq 0\}$ endlich ist. Sind alle $a_i = 0$ so schreiben wir auch $P = 0$.
Eine andere (gleichwertige und ebenfalls eindeutige) Schreibweise ist

$$K[X] = \{0\} \cup \{P = \sum_{i=0}^{d} a_i X^i \mid d \in \mathbb{N}_0, a_d \neq 0\}.$$

Hat P die Darstellung $P = \sum_{i=0}^{d} a_i X^i$ mit $a_d \neq 0$, so heißt $d =: \deg(P)$ der Grad von P , man setzt dazu $\deg(0) = -\infty$. Wir schreiben $K[X]_d$ für den Unterraum der $P \in K[X]$ mit $\deg(P) \leq d$ (wobei $-\infty < n$ für alle $n \in \mathbb{N}_0$ gesetzt wird). Man hat dann $\dim(K[X]_d) = d + 1$ für alle $d \in \mathbb{N}_0$, und die X^i mit $i \leq d$ bilden eine Basis von $K[X]_d$, die wir die Standardbasis dieses Vektorraums nennen.
Wir betrachten jetzt die durch

$$M_X(\sum_{i=0}^{d} a_i X^i) = \sum_{i=1}^{d+1} a_{i-1} X^i$$

definierte Abbildung $M_X : K[X]_d \to K[X]_{d+1}$. Bezüglich der Standardbasen beider Räume hat sie die Matrix $A = (a_{ij}) \in M((d+1) \times d, K)$ mit $a_{ij} = \begin{cases} 1 & i = j+1 \\ 0 & i \neq j+1 \end{cases}$, also

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Definition und Satz 5.3. *Sei K ein Körper.*

a) *Die Menge* $\mathrm{Hom}_K(V, W)$ *der linearen Abbildungen des K-Vektorraums V in den K-Vektorraum W ist ein Unterraum des Vektorraums* $\mathrm{Abb}(V, W)$ *aller Abbildungen von V nach W.*

Ist $V = W$, so schreibt man $\mathrm{End}_K(V) := \mathrm{Hom}_K(V, W)$*, die linearen Selbstabbildungen von V nennt man auch* Endomorphismen.

b) *Die Menge $M(p \times n, K)$ ist mit komponentenweise definierter Addition und Multiplikation mit Skalaren $\lambda \in K$ ein zu K^{pn} isomorpher K-Vektorraum.*

c) *Seien $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{B}' = (w_1, \ldots, w_p)$ Basen der K-Vektorräume V bzw. W. Dann werden durch die in Definition und Satz 5.2 definierten Abbildungen*

$$\begin{array}{rcl} f & \longmapsto & M^{\mathcal{B}}_{\mathcal{B}'}(f) \\ A & \longmapsto & f^{\mathcal{B}}_{\mathcal{B}'}(A) \end{array}$$

zueinander inverse Isomorphismen von K-Vektorräumen zwischen $\mathrm{Hom}_K(V, W)$ *und $M(p \times n, K)$ gegeben.*

Beweis. a) In den Übungen wurde für $W = K$ gezeigt, dass $\mathrm{Abb}(V, W)$ ein K-Vektorraum ist; für allgemeines W zeigt man es genauso.
Dass $\mathrm{Hom}_K(V, W)$ ein Unterraum dieses Vektorraums ist, muss man nachrechnen:
Der Nullvektor von $\mathrm{Abb}(V, W)$, Nullabbildung 0, ist linear, gehört also zu $\mathrm{Hom}_K(V, W)$.
Sind $f, g \in \mathrm{Hom}_K(V, W), \lambda \in K$, so gilt für alle $v_1, v_2 \in V, a \in K$:

$$\begin{aligned} (f + \lambda g)(v_1 + av_2) &= f(v_1 + av_2) + \lambda g(v_1 + av_2) \\ &= f(v_1) + af(v_2) + \lambda(g(v_1) + ag(v_2)) \\ &= (f + \lambda g)(v_1) + a((f + \lambda g)(v_2)). \end{aligned}$$

b) ist klar.
Zu c): Dass die beiden in Definition und Satz 5.2 angegebenen Abbildungen $f \longmapsto M^{\mathcal{B}}_{\mathcal{B}'}(f)$ und $A \longmapsto f^{\mathcal{B}}_{\mathcal{B}'}(A)$ zueinander invers sind, ist klar, dass sie linear sind, rechnet man nach. □

Wir betrachten jetzt den Zusammenhang zwischen Matrizen und linearen Abbildungen weiter, vor allem wollen wir sehen, was für die Matrizen der Komposition (Hintereinanderausführung) von linearen Abbildungen entspricht.

Definition 5.4. (Matrizenprodukt) *Sei K ein Körper. Für Matrizen $A \in M(r \times p, K)$ und $B \in M(p \times n, K)$ ist das Produkt $AB \in M(r \times n, K)$ definiert durch*

$$AB = C = (c_{ik}) \quad \textit{mit } c_{ik} = \sum_{j=1}^{p} a_{ij} b_{jk} \quad (1 \le i \le r,\ 1 \le k \le n).$$

(Für Matrizen A, B, bei denen die Zeilenanzahl von B nicht gleich der Spaltenanzahl von A ist, ist ein Matrizenprodukt AB nicht definiert.)

Satz 5.5. *U, V, W seien K-Vektorräume mit endlichen Basen $\mathcal{B}, \mathcal{B}', \mathcal{B}''$. Ferner seien lineare Abbildungen*

$$g : U \longrightarrow V, \quad f : V \longrightarrow W$$

gegeben. Dann ist

$$M_{\mathcal{B}''}^{\mathcal{B}}(f \circ g) = M_{\mathcal{B}''}^{\mathcal{B}'}(f) M_{\mathcal{B}'}^{\mathcal{B}}(g).$$

(Die Matrix der Komposition *$f \circ g$ ist das Produkt aus der Matrix von f und der Matrix von g.)*

Beweis. Auch hier hilft wieder nur stures Nachrechnen:
Sind $\mathcal{B} = \{u_1, \ldots, u_n\}, \mathcal{B}' = \{v_1, \ldots, v_p\}, \mathcal{B}'' = \{w_1, \ldots, w_r\}$ so hat man nach Definition der der Abbildung zugeordneten Matrix mit $M_{\mathcal{B}''}^{\mathcal{B}'}(f) = A = (a_{ij}), \quad M_{\mathcal{B}'}^{\mathcal{B}}(g) = B = (b_{jk})$

$$g(u_k) = \sum_{j=1}^{p} b_{jk} v_j, \quad f(v_j) = \sum_{i=1}^{r} a_{ij} w_i.$$

Wendet man f auf $g(u_k)$ an, so erhält man also

$$\begin{aligned}
(f \circ g)(u_k) &= f(\sum_{j=1}^{p} b_{jk} v_j) \\
&= \sum_{j=1}^{p} b_{jk} f(v_j) \\
&= \sum_{j=1}^{p} b_{jk} (\sum_{i=1}^{r} a_{ij} w_i) \\
&= \sum_{i=1}^{r} (\sum_{j=1}^{p} a_{ij} b_{jk}) w_i,
\end{aligned}$$

was die Behauptung zeigt, da $\sum_{j=1}^{p} a_{ij} b_{jk}$ der ik-Eintrag von AB ist. □

Bemerkung. Man sieht, dass die Matrizenmultiplikation genau so definiert wurde, dass der obige Satz gilt.

Beispiel. a) Seien $A = (a_1, \ldots, a_n) \in M(1 \times n, K)$ ein Zeilenvektor, $B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in M(n \times 1, K)$ ein Spaltenvektor. Dann ist $AB = (a_1 b_1 + \cdots + a_n b_n) \in M(1 \times 1, K)$.
(1×1-Matrizen identifiziert man meist mit Körperelementen). Das verallgemeinert das bekannte *Skalarprodukt* im $\mathbb{R}^3$. Dagegen ist $BA \in M(n \times n, K)$ die $n \times n$-Matrix mit ij-Eintrag $b_i a_j$.

Man identifiziert meist $n \times 1$-Matrizen mit als Spalten geschriebenen Vektoren, $1 \times n$-Matrizen nennt man dann *Zeilenvektoren* und schreibt den Zeilenvektor wie schon früher benutzt als

$$(a_1, \ldots, a_n) = {}^t\mathbf{a} \text{ mit } \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Das "Skalarprodukt" von $\mathbf{a}$ und $\mathbf{b}$ in K^n im obigen Sinne ist dann also ${}^t\mathbf{a} \cdot \mathbf{b}$.

b) Die Faustregel für das Ausrechnen von Matrixprodukten ist: Den ik-Eintrag von AB erhält man, indem man die i-te Zeile von A mit der k-ten Spalte von B multipliziert (als Skalarprodukt im obigen Sinne).
Damit das geht, muss natürlich die Länge einer Zeile von A (= Spaltenanzahl von A) gleich der Länge einer Spalte von B (= Zeilenanzahl von B) sein.

c) Sei $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$ in $M(2 \times 2, \mathbb{R})$.
Dann ist $AB = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$, $BA = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$.
Insbesondere ist $AB \neq BA$, die Matrixmultiplikation ist also selbst dann nicht kommutativ, wenn AB und BA das gleiche Format haben.

d) Sei $V = W = \mathbb{R}^2$, $\mathcal{B} = \mathcal{B}'$ die Standardbasis. Für Winkel α und β im positiven Sinn hat die Drehung D_α um α die Matrix $\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$, analog für D_β. Da $D_\alpha \circ D_\beta = D_{\alpha+\beta}$ gilt, ist die Aussage von Lemma 5.5 hier äquivalent zu den Additionstheoremen für Sinus und Cosinus:

$$\begin{aligned} \sin(\alpha+\beta) &= \sin\alpha \cdot \cos\beta + \cos\alpha \cdot \sin\beta \\ \cos(\alpha+\beta) &= \cos\alpha \cdot \cos\beta - \sin\alpha \cdot \sin\beta \ . \end{aligned}$$

Speziell hat die Drehung um 90° die Matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{mit } J^2 = -E_2$$

(= Matrix der Drehung um 180°).

Lemma 5.6. a) *Das Matrizenprodukt ist assoziativ und distributiv (also $(AB)C = A(BC)$, $A(B+C) = AB+AC$, $(A+B)C = AC+BC$), wenn die vorkommenden Produkte und Summen definiert sind, insbesondere ist jeweils die linke Seite genau dann definiert, wenn die rechte Seite definiert ist.*

b) *Die Komposition von linearen Abbildungen ist assoziativ und distributiv. Es gilt also $(f \circ g) \circ h = f \circ (g \circ h)$, $f \circ (g+h) = f \circ g + f \circ h$, $(f+g) \circ h = f \circ h + g \circ h$, wenn die vorkommenden Produkte und*

Summen definiert sind, insbesondere ist jeweils die linke Seite genau dann definiert, wenn die rechte Seite definiert ist.

Beweis. Für beliebige Abbildungen haben wir uns bereits von der Assoziativität überzeugt. Das Distributivgesetz $(f+g)\circ h = f\circ h + g\circ h$ rechnet man ebenfalls für beliebige Abbildungen sofort nach. Für das andere Distributivgesetz seien $g,h : U \to V$ und $f : V \to W$ lineare Abbildungen zwischen den K-Vektorräumen U, V, W. Für $u \in U$ ist dann

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

Die Aussage über Matrizen kann man dann entweder hieraus folgern, da die Bijektionen $A \mapsto f^{\mathcal{B}}_{\mathcal{B}'}(A),\ f \mapsto M^{\mathcal{B}}_{\mathcal{B}'}(f)$ zwischen Matrizen und linearen Abbildungen mit Summe und Produkt verträglich sind.
Alternativ rechnet man sie ebenfalls nach. Für die Assoziativität hat man etwa:
Seien $A \in M(s\times r, K), B \in M(r\times p, K), C \in M(p\times n, K)$.
Der il-Eintrag $(1 \le i \le s, 1 \le l \le n)$ der $s\times n$-Matrix $(AB)C$ ist

$$\sum_{k=1}^{p}(\sum_{j=1}^{r} a_{ij}b_{jk})c_{kl} = \sum_{j=1}^{r} a_{ij}(\sum_{k=1}^{p} b_{jk}c_{kl}),$$

also gleich dem il- Eintrag von $A(BC)$. □

Bemerkung: Die quadratische Matrix

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & & \dots & 0 & 1 \end{pmatrix} =: E_n \in M(n\times n, K) =: M_n(K),$$

deren Diagonaleinträge 1 sind und für die alle anderen Einträge 0 sind, heißt die $n\times n$-*Einheitsmatrix* ($E_n = (\delta_{ij})$, wo δ_{ij} das Kronecker-Delta ist).
Für diese gilt:

$$\begin{array}{rcll} E_nA & = & A & \text{für alle } A \in M(n\times r, K), \\ BE_n & = & B & \text{für alle } B \in M(p\times n, K). \end{array}$$

Insbesondere hat man in der Menge $M_n(K) = M(n\times n, K)$ mit $+$ und Matrizenprodukt zwei Verknüpfungen, für die alle Körperaxiome mit Ausnahme der Kommutativität der Multiplikation und der Existenz multiplikativer Inverser erfüllt sind. Man sagt (siehe die Bemerkung nach Definition 2.1), $M(n\times n, K) = M_n(K)$ sei ein *Ring* (mit Einselement E_n), der *Matrizenring* vom Grad oder von der Ordnung n. Dass multiplikative Inverse in $M_n(K)$ nicht immer existieren, sieht

man (warum?) etwa an

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

Definition und Lemma 5.7. *Sei K ein Körper*

a) *Für $A \in M(p \times n, K)$ sei ${}^tA$ die $n \times p$-Matrix ${}^tA = B = (b_{ij})$ mit $b_{ij} = a_{ji}$ für $1 \leq i \leq n, 1 \leq j \leq p$. Die Matrix ${}^tA$ heißt die zu A* transponierte *Matrix (seltener: gestürzte Matrix), ihre Zeilen sind die Spalten von A und umgekehrt.*
b) *Es gilt ${}^t({}^tA) = A$ für alle A.*
c) *Durch $A \mapsto {}^tA$ wird ein Vektorraumisomorphismus $M(p \times n, K) \longrightarrow M(n \times p, K)$ gegeben.*
d) *Seien $A \in M(r \times p, K), B \in M(p \times n, K)$. Dann gilt:*
$$ {}^t(AB) = {}^tB\ {}^tA. $$
e) *Ist $A \in M_n(K)$ eine quadratische Matrix, so heißt A symmetrisch, wenn ${}^tA = A$ gilt, schiefsymmetrisch, wenn ${}^tA = -A$ gilt.*

Beweis. Man rechnet das nach. □

Bemerkung. Da Zeilenvektoren der Länge n auch als $1 \times n$-Matrizen, Spaltenvektoren der Länge n als $n \times 1$-Matrizen aufgefasst werden können, verallgemeinert die Definition die Notation ${}^t\mathbf{z}$ für die Zeilenvektoren einer Matrix.

Lemma 5.8. *V und W seien K-Vektorräume mit Basen $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{B}' = (w_1, \ldots, w_p)$. Sei $f : V \longrightarrow W$ linear, $A = M^{\mathcal{B}}_{\mathcal{B}'}(f)$ die Matrix von A bezüglich der Basen $\mathcal{B}$ und $\mathcal{B}'$. Sei*

$$v = \sum_{j=1}^{n} x_i v_i \in V \qquad \text{mit } f(v) = w = \sum_{i=1}^{p} y_i w_i.$$

Dann ist

$$\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} .$$

Insbesondere ist die Abbildung $L_A : K^n \longrightarrow K^p$ aus Lemma 3.12 durch das Matrizenprodukt gegeben:

$$L_A(\mathbf{x}) = A \cdot \mathbf{x}$$

(wobei K^n mit $M(n \times 1, K)$ identifiziert wird).

Man hat mit den Koordinatenabbildungen $c_{\mathcal{B}}, c_{\mathcal{B}'}$ aus Definition und Satz 4.24 das kommutative Diagramm

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow c_{\mathcal{B}} & & \downarrow c_{\mathcal{B}'} \\ K^n & \xrightarrow{L_A} & K^p \end{array}$$

(d.h., $c_{\mathcal{B}'} \circ f = L_A \circ c_{\mathcal{B}}$*).*
Insbesondere sieht man: Der Kern von f *wird durch* $c_{\mathcal{B}}$ *isomorph auf den Kern von* L_A *abgebildet, das Bild von* f *wird durch* $c_{\mathcal{B}}$ *isomorph auf das Bild von* L_A *abgebildet.*

Beweis. Es gilt

$$\begin{aligned} f(\sum_{j=1}^{n} x_j v_j) &= \sum_{j=1}^{n} x_j f(v_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} w_i \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} x_j) w_i. \end{aligned}$$

Koeffizientenvergleich mit dem Ausdruck

$$f(v) = \sum_{i=1}^{m} y_i w_i$$

liefert also die erste Behauptung.
Die zweite Behauptung ist nach Definition von L_A eine direkte Folgerung aus der ersten, die letzte Behauptung sieht man direkt, wen man die Definition der Koordinatenabbildungen einsetzt. □

Ein wichtiger Spezialfall entsteht, wenn man speziell als Bildraum W den Grundkörper K betrachtet:

Definition und Korollar 5.9. *Sei* V *ein Vektorraum über dem Körper* K*.*
Der Vektorraum $\mathrm{Hom}_K(V, K) =: V^*$ *heißt der* Dualraum *von* V*.*
Hat V *die endliche Basis* $\mathcal{B} = (v_1, \ldots, v_n)$ *und ist* $\mathcal{B}'$ *die (einelementige) Basis* $\{1\} = \mathcal{B}'$ *des* K*-Vektorraums* K*, so wird durch die Abbildung*

$$f \mapsto \begin{pmatrix} f(v_1) \\ \vdots \\ f(v_n) \end{pmatrix} = {}^t(M^{\mathcal{B}}_{\mathcal{B}'}(f)) \tag{5.1}$$

ein Isomorphismus von V^* *auf* K^n *gegeben.*
Die Elemente des Dualraums heißen auch Linearformen *oder* lineare Funktionale*.*
Das Urbild des j*-ten Vektors* $\mathbf{e}_j$ *der Standardbasis von* K^n *unter diesem Isomorphismus wird mit* v_j^* *bezeichnet; es gilt*

$$v_j^*(v_i) = \delta_{ij} := \begin{cases} 1 & \textit{falls } i = j \\ 0 & \textit{sonst} \end{cases} \tag{5.2}$$

(das Symbol δ_{ij} *in obiger Bedeutung heißt* Kronecker-Delta *(Leopold Kronecker, 1823-1891)).*
$\{v_1^*, \ldots, v_n^*\}$ *heißt die zu* $\{v_1, \ldots, v_n\}$ duale Basis *von* V^**.*

Bemerkung. Da $v_i^*(\sum_{j=1}^n x_j v_j) = x_i$ gilt, heißen die v_i^* auch die *Koordinatenfunktionen* zur Basis $\mathcal{B} = (v_1, \ldots, v_n)$. Es gilt

$$c_{\mathcal{B}}(v) = \begin{pmatrix} v_1^*(v) \\ \vdots \\ v_n^*(v) \end{pmatrix} \qquad \text{für alle } v \in V;$$

man schreibt auch

$$c_{\mathcal{B}} = \begin{pmatrix} v_1^* \\ \vdots \\ v_n^* \end{pmatrix}.$$

Definition und Satz 5.10. *Sei V ein K-Vektorraum und V^* sein Dualraum. Für jede Teilmenge M von V sei*

$$M^\perp := \mathrm{Ann}(M) := \{\varphi \in V^* \mid \varphi(M) = \{\mathbf{0}\}\}$$

der Annullator *von M.*

a) $\mathrm{Ann}(M) = \mathrm{Ann}(\mathrm{Lin}(M))$ *für alle $M \subseteq V$.*
b) $\mathrm{Ann}(M)$ *ist ein Unterraum von V^*.*
c) *Ist $U \subseteq V$ ein Unterraum und $v \in V \setminus U$, so gibt es ein $\varphi \in V^*$ mit $\varphi(v) \neq 0$.*
d) *Ist $U \subseteq V$ ein Unterraum, so ist* $\dim(U) + \dim(\mathrm{Ann}(U)) = \dim(V)$.

Beweis. Übung. □

Definition und Satz 5.11. *Sei $f : V \longrightarrow W$ eine lineare Abbildung von K-Vektorräumen. Sei $f^* := {}^t f : W^* \to V^*$ die durch*

$${}^t f(\varphi) := \varphi \circ f \in V^* \quad \textit{für alle } \varphi \in W^*$$

definierte transponierte Abbildung.

a) *$f^* := {}^t f : W^* \to V^*$ ist linear.*
b) *Seien $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{B}' = (w_1, \ldots, w_p)$ Basen von V bzw. W und $\mathcal{B}^* = (v_1^*, \ldots, v_n^*)$, $(\mathcal{B}')^* = (w_1^*, \ldots, w_p^*)$ die hierzu dualen Basen der Dualräume V^* und W^*, sei $A = M_{\mathcal{B}'}^{\mathcal{B}}(f) \in M(p \times n, K)$ die Matrix von f bezüglich der Basen $\mathcal{B}, \mathcal{B}'$.*
Dann hat ${}^t f : W^* \to V^*$ bezüglich der Basen $(\mathcal{B}')^*$ von W^* und $\mathcal{B}^*$ von V^* die Matrix $M_{\mathcal{B}^*}^{(\mathcal{B}')^*}({}^t f) = {}^t A$.

Beweis. Die Linearität von f^* rechnen wir nach: Sind $\varphi_1, \varphi_2 \in W^*$ und $\lambda \in K$, so gilt

$$\begin{aligned} f^*(\varphi_1 + \lambda\varphi_2) &= (\varphi_1 + \lambda\varphi_2) \circ f \\ &= \varphi_1 \circ f + \lambda\varphi_2 \circ f \\ &= f^*(\varphi_1) + \lambda f^*(\varphi_2). \end{aligned}$$

Für b) betrachten wir erneut das Ergebnis aus Lemma 5.8: Wir haben mit den dortigen Notationen $y_i = w_i^*(f(v)) = (w_i^* \circ f)(v) = ({}^t f(w_i^*))(v)$ und $x_j = v_j^*(v)$ für alle $v \in V$.
Die dortige Gleichung $y_i = \sum_{j=1}^n a_{ij} x_j$ ergibt also

$$ {}^t f(w_i^*) = \sum_{j=1}^n a_{ij} v_j^* \quad \text{für } 1 \leq i \leq p. $$

Das heißt aber gerade, dass ${}^t A$ die Matrix von ${}^t f$ bezüglich der Basen $(\mathcal{B}')^*$ von W^* und $\mathcal{B}^*$ von V^* ist. □

Definition und Satz 5.12. *Eine Matrix $A \in M_n(K) = M(n \times n, K)$ heißt* invertierbar*, wenn es eine Matrix $B \in M(n \times n, K)$ gibt mit $AB = BA = E_n$ (man nennt invertierbare Matrizen auch* regulär*, nicht invertierbare Matrizen* singulär*).*
Für die K-Vektorräume V, W seien Basen $\mathcal{B} = (v_1, \ldots, v_n)$ bzw. $\mathcal{B}' = (w_1, \ldots, w_p)$ gegeben, $f : V \longrightarrow W$ sei linear mit $A = M_{\mathcal{B}'}^{\mathcal{B}}(f) \in M(p \times n, K)$.
Dann gilt:

a) *Es gibt genau dann $B \in M(n \times p, K)$ mit $AB = E_p$, wenn f surjektiv ist (oder äquivalent: wenn $\mathrm{rg}(A) = p$ gilt).*
b) *Es gibt genau dann $B' \in M(n \times p, K)$ mit $B'A = E_n$, wenn f injektiv ist (oder äquivalent: wenn $\mathrm{rg}(A) = n$ gilt).*
c) *$A \in M_n(K)$ ist genau dann invertierbar, wenn f bijektiv ist. Äquivalent dazu ist $\mathrm{rg}(A) = n$.*
d) *Die regulären Matrizen in $M_n(K) = M(n \times n, K)$ bilden eine Gruppe; diese wird mit $\mathrm{GL}_n(K)$ bezeichnet, sie heißt die* allgemeine lineare Gruppe *bzw. auf Englisch* general linear group.

Beweis. c) ist eine Folgerung aus a) und b), kann aber auch mit Hilfe der Korrespondenz zwischen Matrizen und linearen Abbildungen leicht direkt bewiesen werden (Übung). Für d) muss man nur nachprüfen, dass das Produkt zweier invertierbarer Matrizen invertierbar ist (Übung: Warum ist der Rest dieser Behauptung klar?); das sieht man daran, dass $B^{-1}A^{-1}$ zu AB invers ist (nachrechnen!).
a) und b) kann man, wenn man will, durch Lösen linearer Gleichungssysteme zeigen: Für a) überlegt man sich, dass die Lösbarkeit der Matrixgleichung $AB = E_p$ ist äquivalent ist zur Lösbarkeit von $A\mathbf{x} = \mathbf{b}$ für jedes $\mathbf{b}$ und damit zu $\mathrm{rg}(A) = p$, b) führt man auf a) zurück, indem man die Äquivalenz von $B'A = E_n$ mit ${}^t A {}^t B' = E_n$ feststellt und $\mathrm{rg}(A) = \mathrm{rg}({}^t A$ ausnutzt.
Alternativ (und konzeptioneller) kann man für a) so vorgehen: $AB = E_p$ ist äquivalent zu $L_A \circ L_B = Id_{K^p}$.
Mit Hilfe von Satz 5.13 (siehe unten, siehe auch die Übungen) folgen dann a) und b).

Die Aussage über den Rang ist dabei klar, wenn $f = L_A$ ist, da bekanntlich $\dim(\mathrm{Im}(L_A)) = \mathrm{rg}(A)$ und $\dim(\mathrm{Ker}(L_A)) = n - \dim(\mathrm{Im}(L_A))$ gilt, der allgemeine Fall folgt daraus, siehe Lemma 5.14 unten. $\square$

Beispiel: Als Übung zeige man, dass für $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M(2 \times 2, K)$ gilt:

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} &= \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} \\ &= (ad - bc)E_2. \end{aligned}$$

Es folgt, dass $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M(2 \times 2, K)$ genau dann invertierbar ist, wenn $ad - bc \neq 0$ gilt.

Satz 5.13. *Seien V, W Vektorräume über dem Körper K und $f : V \to W$ eine lineare Abbildung.*

a) *Die folgenden Aussagen sind äquivalent:*
   i) *f ist surjektiv.*
   ii) *Es gibt ein Erzeugendensystem M von V, für das $f(M)$ den Vektorraum W erzeugt.*
   iii) *Für alle Erzeugendensysteme M von V erzeugt $f(M)$ den Vektorraum W.*
   iv) *Es gibt eine lineare Abbildung $g : W \to V$ mit $f \circ g = \mathrm{Id}_W$.*

b) *Die folgenden Aussagen sind äquivalent:*
   i) *f ist injektiv.*
   ii) *Es gibt eine Basis $(v_i)_{i \in I}$ von V, für die $(f(v_i))_{i \in I}$ linear unabhängig in W ist.*
   iii) *Für jede linear unabhängige Familie $(v_i)_{i \in I}$ in V ist $(f(v_i))_{i \in I}$ linear unabhängig in W.*
   iv) *Es gibt eine lineare Abbildung $g : W \to V$ mit $g \circ f = \mathrm{Id}_V$.*

Beweis. Die Äquivalenz von i)-iii) sowohl in a) als auch in b) folgt aus dem Beweis von Lemma 4.21, siehe auch die Bemerkung nach diesem Lemma. Zur Äquivalenz von i) und iv) sowohl in a) als auch in b) betrachten wir zunächst a):
Ist $g : W \to V$ mit $f \circ g = \mathrm{Id}_W$ gegeben, so folgt aus $f(g(W)) = W$, dass erst recht $f(V) = W$ gilt, f also surjektiv ist.
Ist umgekehrt f surjektiv, so sei $(w_i)_{i \in I}$ eine Basis von W; da f surjektiv ist, kann man w_i für alle $i \in I$ als $f(v_i)$ mit geeigneten $v_i \in V$ schreiben (um genau zu sein: Hier braucht man für unendliches I das Auswahlaxiom der Mengenlehre). Nach Satz 5.1, b) gibt es dann aber eine lineare Abbildung $g : W \to V$ mit $g(w_i) = v_i$ für alle $i \in I$, und wegen $(f \circ g)(w_i) = w_i$ für alle Vektoren der Basis $(w_i)_{i \in I}$ muss $f \circ g = \mathrm{Id}_V$ gelten.
Für b) sei zunächst $g : W \to V$ mit $g \circ f = \mathrm{Id}_V$ gegeben. Sind $v_1, v_2 \in V$ mit $f(v_1) = f(v_2)$, so ist auch $v_1 = (g \circ f)(v_1) = (g \circ f)(v_2) = v_2$, also

ist f injektiv. Ist umgekehrt f injektiv, so sei $(v_i)_{i\in I}$ eine Basis von V. Weil f injektiv ist, sind die $w_i := f(v_i) \in W$ linear unabhängig. Nach Satz 5.1, a) gibt es dann eine lineare Abbildung $g : W \to V$ mit $g(w_i) = v_i$ für alle $i \in I$, und wir haben $(g \circ f)(v_i) = v_i$ für alle Basisvektoren v_i von V, also $g \circ f = \mathrm{Id}_V$. □

Definition und Lemma 5.14. *Sei $f : V \longrightarrow W$ eine lineare Abbildung von K-Vektorräumen. Dann ist der Rang von f definiert als*

$$\mathrm{rg}(f) := \dim(\mathrm{Im}(f)).$$

Sind $\mathcal{B}, \mathcal{B}'$ endliche Basen von V und W und $A = M^{\mathcal{B}}_{\mathcal{B}'}(f)$ die Matrix von f bezüglich $\mathcal{B}, \mathcal{B}'$, so ist der Rang von f gleich dem Rang der Matrix A.
Insbesondere gilt in diesem Fall:

a) *f ist genau dann surjektiv, wenn* $\mathrm{rg}(A) = \dim(W)$ *gilt.*
b) *f ist genau dann injektiv, wenn* $\mathrm{rg}(A) = \dim(V)$ *gilt.*

Beweis. Wir wissen bereits, dass der Rang der linearen Abbildung $L_A : K^n \longrightarrow K^m$ gleich $\mathrm{rg}(A)$ ist, und aus dem zweiten Teil von Lemma 5.8 folgt, dass L_A und f gleichen Rang haben, da das Bild von f durch die Koordinatenabbildung $c_{\mathcal{B}'}$ isomorph auf das von L_A abgebildet wird. Die Behauptung a) ist damit sofort klar. Für b) stellen wir zunächst fest, dass wiederum nach Lemma 5.8 die Abbildungen f und L_A gleichzeitig injektiv sind (d.h., die Injektivität von f ist äquivalent zur Injektivität von L_A). Da die Injektivität von L_A gleichwertig dazu ist, dass das Gleichungssystem $A\mathbf{x} = \mathbf{0}$ nur die triviale Lösung hat, folgt b) aus dem, was wir über lineare Gleichungssysteme gezeigt haben. □

Korollar 5.15. *Sei K ein Körper, $A \in M(p \times m, K)$, $B \in M(m \times n, K)$. Dann ist*

$$\mathrm{rg}(AB) \leq \min(\mathrm{rg}(A), \mathrm{rg}(B)).$$

Beweis. Da der Rang von AB gleich der Dimension des Bildes der linearen Abbildung $L_A \circ L_B$ ist und dieses Bild offenbar gleich $L_A(L_B(K^m)) \subseteq L_A(K^p)$ ist, ist sofort $\mathrm{rg}(AB) \leq \mathrm{rg}(A)$ klar. Da aber die Dimension von $L_A(L_B(K^m))$ nicht größer sein kann als die von $L_B(K^m)$, folgt auch $\mathrm{rg}(AB) \leq \mathrm{rg}(B)$. □

Satz 5.16. *Sei $\mathbb{C} := \langle E_2, J\rangle$ der von den Matrizen E_2 und $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ aufgespannte Unterraum von $M(2 \times 2, \mathbb{R})$.*

Dann gilt: $\mathbb{C}$ ist (bezüglich der Matrizenverknüpfungen) ein Körper mit Einselement $1 = E_2$; für das Element $i := J$ gilt $i^2 = -1$. Die Teilmenge $\{aE_2 \mid a \in \mathbb{R}\}$ ist ein zu $\mathbb{R}$ isomorpher Teilkörper, der im allgemeinen mit $\mathbb{R}$ identifiziert wird; man schreibt dann auch 1 für E_2 und a für aE_2.

Der Körper $\mathbb{C}$ heißt der Körper der komplexen Zahlen*, i die imaginäre*

Einheit. In $\mathbb{C}$ hat jedes Element z eine eindeutige Darstellung $z = a+bi$ mit $a, b \in \mathbb{R}$ und es gelten die Rechenregeln

$$\begin{aligned}(a+bi)(c+di) &= (ac-bd)+(ad+bc)i \\ (a+bi)^{-1} &= \frac{a-bi}{a^2+b^2} \quad \textit{falls } a+bi \neq 0 \textit{ ist.}\end{aligned}$$

Beweis. Man rechnet nach, dass $i^2 = -E_2$ gilt und folgert daraus sofort die angegebene Rechenregel und die multiplikative Abgeschlossenhheit von $\mathbb{C}$. Die Gültigkeit der Assoziativ- und Distributivgesetze folgt aus deren Gültigkeit für den Matrizenring, ebenso die Gruppeneigenschaft von $(\mathbb{C}, +)$ und die Tatsache, dass E_2 multiplikativ neutral ist. Dass die auf $\mathbb{C}$ eingeschränkte Matrizenmultiplikation kommutativ ist, rechnet man sofort nach, und dass jedes von Null verschiedene Element $a + bi = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ invertierbar ist, folgt aus dem Beispiel nach Definition und Korollar 5.12: Danach ist $a+bi = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ genau dann invertierbar, wenn $a^2 + b^2 \neq 0$ gilt, und dann gilt die angegebene Formel für das inverse Element. □

Bemerkung:

- Der Körper $\mathbb{C}$ ist *algebraisch abgeschlossen*, d.h., dass jedes Polynom $p(x) = a_0 + a_1x + \cdots + a_nx^n$ mit Koeffizienten $a_0, \ldots, a_n$ in $\mathbb{C}$ eine Nullstelle in $\mathbb{C}$ hat.
  Daraus folgt (siehe den späteren Paragraphen über Polynome), dass für jedes Polynom p wie oben (mit $a_n \neq 0$) gilt: Es gibt $z_1, \ldots, z_n \in \mathbb{C}$ mit
  $$p(x) = a_n(x - z_1) \ldots (x - z_n)$$
  (die z_i sind dabei nicht notwendig verschieden).
  Zum Beispiel gilt in $\mathbb{C}$:
  $$\begin{aligned} x^2+1 &= (x+i)(x-i) \\ (x^3-1) &= (x-1)(x+\tfrac{1}{2}-\tfrac{\sqrt{3}}{2}i)(x+\tfrac{1}{2}+\tfrac{\sqrt{3}}{2}i). \end{aligned}$$
- Den Körper $\mathbb{C}$ veranschaulicht man sich meist in der Gauß'schen Zahlenebene:
  Dem Element $z = a + ib$ wird der Punkt $\begin{pmatrix} a \\ b \end{pmatrix}$ in der Ebene zugeordnet, also gerade der Punkt, den man durch Anwendung der Matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ auf den Punkt $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ erhält.
  Setzt man $r := \sqrt{a^2+b^2}$, so kann man $a = r\cos\varphi, b = r\sin\varphi$ mit einem geeigneten Winkel φ schreiben, die Matrix $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ ist also die Matrix einer Drehstreckung, die sich aus einer Streckung um den Faktor r und einer Drehung um den Winkel φ zusammensetzt.

Man sieht, dass dann r gerade der Abstand des Punktes $\binom{a}{b}$ vom Ursprung und φ der Winkel des Ortsvektors mit der x-Achse ist.

Für die komplexe Zahl $\cos\varphi + i\sin\varphi$ ergibt sich in der Analysis die Darstellung

$$\cos\varphi + i\sin\varphi = \exp(i\varphi)$$

(Der Winkel φ wird dabei im Bogenmaß gemessen).

Zusammenfassung Matrizen und lineare Abbildungen

$f : V \longrightarrow W$ sei eine lineare Abbildung (V, W endlichdimensionale K-Vektorräume).

1. Eine lineare Abbildung $f : V \longrightarrow W$ ist bestimmt durch die Bilder der Basisvektoren von V, zu jeder Vorgabe von Bildern existiert genau eine lineare Abbildung (*lineare Fortsetzung* der Vorgabe auf den Basisvektoren).
   (auch gültig bei unendlicher Dimension)
   Entscheidender Punkt: Eindeutige Darstellung der Vektoren von V als Linearkombinationen der Basisvektoren.
2. Nach Punkt 1 wird f eindeutig charakterisiert durch Angabe der Koordinaten a_{ij} der $f(v_j)$ bezüglich der Basisvektoren w_i in W. Diese Koeffizienten werden in der Matrix $A = M^{\mathcal{B}}_{\mathcal{B}'}(f)$ zusammengefasst.
   Der Zusammenhang zwischen Matrix und Abbildung ist
   $$\begin{aligned} f(v_j) &= \sum a_{ij} w_i \\ f(\textstyle\sum_{j=1}^n x_j v_j) &= \textstyle\sum_{i=1}^p y_i w_i \quad \text{mit} \\ \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} &= A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} . \end{aligned}$$
3. Ist $f : V \longrightarrow W$ linear, so hat das Bild $f(V) = \mathrm{Im}(f)$ Dimension $\leq \dim V$. Wir werden im nächsten Abschnitt sehen, dass die Differenz gerade die Dimension des Kerns von f ist (Dimensionsformel).
   Also: Eine lineare Abbildung kann nicht die Dimension vergrößern.
4. Zur Hintereinanderausführung (Komposition) von Abbildungen gehört das Produkt der zugehörigen Matrizen. Der (ik)-Koeffizient des Produkts AB ist das "Skalarprodukt" aus i-ter Zeile von A und k-ter Spalte von B.
5. Die Matrix $A \in M(p \times n, K)$ gehört genau dann zu einer surjektiven Abbildung, wenn $\mathrm{rg}(A) = p$, genau dann zu einer injektiven Abbildung, wenn $\mathrm{rg}(A) = n$ gilt; in jedem Fall ist $\mathrm{rg}(A) \leq$

$\min(n, P)$.
Die Differenz $n - \mathrm{rg}(A)$ ist (siehe die Dimensionsformel im nächsten Abschnitt) die Dimension des Kerns der zugehörigen linearen Abbildung, sie heißt auch der *Defekt* von A.

6. Basiswechsel und Matrizen

Definition 6.1. *Sei V ein K-Vektorraum mit Basen $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{B}' = (v'_1, \ldots, v'_n)$. Es gelte*

$$v'_j = \sum_{i=1}^{n} s_{ij} v_i \quad \text{für } 1 \leq j \leq n \ .$$

Dann heißt die Matrix $S = (s_{ij}) \in M(n \times n, K)$ die Übergangsmatrix *von $\mathcal{B}$ zu $\mathcal{B}'$ (*Matrix des Basiswechsels *von $\mathcal{B}$ zu $\mathcal{B}'$).*
Die Übergangsmatrix drückt also die Vektoren der neuen Basis $\mathcal{B}'$ durch die Vektoren der alten Basis $\mathcal{B}$ aus, ihre Spalten sind die Koordinatenvektoren der neuen Basisvektoren bezüglich der alten Basis.

Beispiele:

- Sei $V = \mathbb{R}^2$ mit der Standardbasis $\mathcal{B} = (e_1, e_2)$. Die Basis $\mathcal{B}' = (e'_1, e'_2)$ gehe aus $\mathcal{B}$ durch Drehung um den Winkel φ im Gegenuhrzeigersinn hervor, also $e'_1 = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$, $e'_2 = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}$. Dann ist $S = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ die Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$
- Allgemeiner sei $\mathcal{B} = (e_1, \ldots, e_n)$ die Standardbasis des K^n und $\mathcal{B}' = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ eine weitere Basis des K^n. Dann ist die Matrix S mit den Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$ die Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$.

Lemma 6.2. a) *Mit den Bezeichnungen von Definition 6.1 ist die Übergangsmatrix S gleich $M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_V)$. Insbesondere ist S invertierbar und $S^{-1} = M_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{Id}_V)$ die Übergangsmatrix von $\mathcal{B}'$ zu $\mathcal{B}$.*

b) *Ist $f : V \longrightarrow V$ der lineare Isomorphismus mit $f(v_i) = v'_i$ $(1 \leq i \leq n)$, so ist*

$$S = M_{\mathcal{B}}^{\mathcal{B}}(f).$$

c) *Das Diagramm*

$$\begin{tikzcd} & V \arrow[d, "c_{\mathcal{B}'}"'] \arrow[dr, "c_{\mathcal{B}}"] & \\ & K^n \arrow[r, "L_S"] & K^n \end{tikzcd}$$

ist kommutativ.
Für $v = \sum_{i=1}^{n} x_i v_i = \sum_{j=1}^{n} y_j v'_j$ $(x_i,\ y_i \in K$ für $1 \leq i \leq n)$ gilt

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = S \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \ . \tag{6.1}$$

Beweis. Das Diagramm ist nur eine andere Schreibweise für die Gleichung (6.1), die wir durch Nachrechnen beweisen:

$$\begin{aligned}\sum_{j=1}^{n} y_j v'_j &= \sum_{j=1}^{n} y_j \sum_{i=1}^{n} s_{ij} v_i \\ &= \sum_{i=1}^{n} (\sum_{j=1}^{n} s_{ij} y_j) v_i \\ &= \sum_{i=1}^{n} x_i v_i,\end{aligned}$$

Vergleich der Koeffizienten bei v_i ergibt $x_i = \sum_{j=1}^{n} s_{ij} y_j$, also die Behauptung. □

Bemerkung: Die Matrix S drückt also einerseits die neuen Basisvektoren (Vektoren von $\mathcal{B}'$) durch die alten Basisvektoren (Vektoren von $\mathcal{B}$) aus, andererseits die Koordinaten $x_1, \ldots, x_n$ bezüglich der alten Basis durch die Koordinaten $y_1, \ldots, y_n$ bezüglich der neuen Basis. In der Bezeichnung "Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$" steckt daher eine gewisse Willkür, die oben bemerkte Überkreuzung, die viel Verwirrung hervorruft, liegt aber in der Natur der Sache, man kann nur wählen, in welcher Richtung man sie durchläuft.

Satz 6.3. *(Transformation der Koordinatenmatrix bei Basiswechsel)*
Sei $f : V \longrightarrow W$ linear. Seien $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{B}' = (v'_1, \ldots, v'_n)$ Basen von V und $S = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_V)$ die Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$, seien $\mathcal{C} = (w_1, \ldots, w_p)$ und $\mathcal{C}' = (w'_1, \ldots, w'_p)$ Basen von W mit Übergangsmatrix $T = M_{\mathcal{C}}^{\mathcal{C}'}(\mathrm{Id}_W)$. Dann gilt:
Ist $A = M_{\mathcal{C}}^{\mathcal{B}}(f)$, $A' = M_{\mathcal{C}'}^{\mathcal{B}'}(f)$, so ist $A' = T^{-1}AS$.

Beweis. Man liest die Gleichheit $A' = T^{-1}AS$ an dem kommutativen Diagramm

$$\begin{array}{ccccc} & K^n & \xrightarrow{A\cdot} & K^p & \\ c_{\mathcal{B}} \nearrow & & & & \searrow c_{\mathcal{C}}^{-1} \\ V & \uparrow S\cdot & & \downarrow T^{-1}\cdot & W \\ c_{\mathcal{B}'} \searrow & & & & \nearrow c_{\mathcal{C}'}^{-1} \\ & K^n & \xrightarrow{A'\cdot} & K^p & \end{array}$$

ab. In diesem stellen die mit $A\cdot, S\cdot, A'\cdot, T^{-1}\cdot$ bezeichneten Pfeile jeweils die Multiplikation von links mit der betreffenden Matrix dar, also die linearen Abbildungen L_A, L_S usw.. □

Den besonders häufigen Spezialfall $V = K^n$, $W = K^p$ mit Standardbasen $\mathcal{B}$, $\mathcal{C}$ notieren wir als Korollar:

Korollar 6.4. *Seien $\mathcal{B}' = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ und $\mathcal{C}' = (\mathbf{t}_1, \ldots, \mathbf{t}_p)$ Basen von K^n bzw. K^p, S bzw. T die Matrix mit Spalten $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ bzw.*

$(\mathbf{t}_1, \ldots, \mathbf{t}_p)$, *$A \in M(p \times n, K)$. Dann ist*

$$M_{\mathcal{C}'}^{\mathcal{B}'}(L_A) = T^{-1}AS$$

(wobei L_A wie üblich durch $L_A(\mathbf{x}) = A \cdot \mathbf{x}$ gegeben ist).

Korollar 6.5. *Sei V ein endlich dimensionaler K-Vektorraum und* $\mathrm{End}(V)$ *die Menge der linearen Abbildungen von V nach V; sei $\mathcal{B}$ eine Basis von V und $M_{\mathcal{B}}(f) = M_{\mathcal{B}}^{\mathcal{B}}(f)$ die Matrix von f bezüglich $\mathcal{B}$. Dann gilt:*
Ist $\mathcal{B}'$ eine weitere Basis von V, S die Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$, $A = M_{\mathcal{B}}(f)$, $A' = M_{\mathcal{B}'}(f)$, so ist $A' = S^{-1}AS$.

Korollar 6.6. *Ist $\mathcal{B} = (\mathbf{s}_1, \ldots, \mathbf{s}_n) \in K^n$, S die Matrix mit Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$, so hat L_A bezüglich $\mathcal{B}$ die Matrix $A' = S^{-1}AS$.*

Beweis. Für alle drei Korollare ist der Beweis klar. □

Definition 6.7. a) *Seien $A, A' \in M(p \times n, K)$. Dann heißen A und A'* äquivalent *($A \sim A'$), wenn es invertierbare Matrizen $S \in M(n \times n, K)$, $T \in M(p \times p, K)$ gibt, so dass $A' = T^{-1}AS$ gilt.*
b) *Seien $A, A' \in M(n \times n, K)$. Dann heißen A und A'* ähnlich *(oder* konjugiert*) ($A \approx A'$), wenn es eine invertierbare Matrix $S \in M(n \times n, K)$ gibt, so dass $A' = S^{-1}AS$ gilt.*

Bemerkung. a) Äquivalenz und Ähnlichkeit von Matrizen sind Äquivalenzrelationen, wie man leicht nachrechnet.
b) Ist G eine Gruppe, so heißen Elemente $x, x' \in G$ zueinander *konjugiert*, wenn es $g \in G$ gibt mit $x' = g^{-1}xg$.

Lemma 6.8. *Sei K ein Körper, $n \in \mathbb{N}$.*
a) *Die Matrizen $A, A' \in M(p \times n, K)$ sind genau dann äquivalent zueinander, wenn sie bezüglich geeigneter Basen von K^n, K^p die gleiche lineare Abbildung $f : K^n \longrightarrow K^p$ darstellen.*
b) *Die Matrizen $A, A' \in M(n \times n, K)$ sind genau dann konjugiert zueinander, wenn sie bezüglich geeigneter Basen von K^n den gleichen Endomorphismus von K^n darstellen.*

Beweis. Klar. □

Definition und Lemma 6.9. *In $M(n \times n, K)$ sei E^{ij} (für $1 \leq i, j \leq n$) die Matrix, deren ij-Eintrag gleich 1 ist und deren sonstige Einträge 0 sind. Dann gilt für eine Basis $(v_1, \ldots, v_n)$ von K:*
a) *Die elementare Basisumformung $v_j \longmapsto v_j' = v_j + \lambda v_i$ ($1 \leq i, j \leq n, i \neq j$) (mit $v_k \longmapsto v_k' = v_k$ für $k \neq j$) hat die Übergangsmatrix*

$$T_{ij}(\lambda) = E_n + \lambda E^{ij}.$$

Es gilt $T_{ij}(\lambda)T_{ij}(\lambda') = T_{ij}(\lambda + \lambda')$, insbesondere ist $T_{ij}(\lambda)$ invertierbar mit $(T_{ij}(\lambda))^{-1} = T_{ij}(-\lambda)$.
Die Matrizen $T_{ij}(\lambda)$ heißen Elementarmatrizen.

b) *Für $j \neq i$ hat die elementare Basisumformung*

$$v_j \longmapsto v_j' = v_i,\ v_i \longmapsto v_i' = v_j,\ v_k \longmapsto v_k' = v_k \quad \text{für } k \notin \{i,j\}$$

(Vertauschung von v_i und v_j) die Matrix

$$P_{ij} = E_n - E^{ii} - E^{jj} + E^{ij} + E^{ji}$$

(der $k\ell$-Eintrag von P_{ij} ist $\delta_{k\ell}$ für $i \neq k \neq j$, $i \neq \ell \neq j$, 0 für $k = \ell = i$ und für $k = \ell = j$, 1 für $k = i$, $\ell = j$ und für $k = j$, $\ell = i$). Die P_{ij} heißen elementare Permutationsmatrizen*, Produkte von Matrizen vom Typ P_{ij} heißen Permutationsmatrizen.*

c) *Für $1 \leq i \leq n$ und $\lambda \in K, \lambda \neq 0$ hat die elementare Basisumformung*

$$v_i \longmapsto v_i' := \lambda v_i,\ v_j \longmapsto v_j' := v_j \text{ für } j \neq i$$

die Matrix

$$D_i(\lambda) := \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \lambda & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix},$$

bei der das Diagonalelement in Position (i,i) gleich λ ist.

Beweis. Auch hier folgt der Beweis direkt aus der Definition der Matrix des Basiswechsels. □

Lemma 6.10. *Sei $A \in M(p \times n, K)$ eine Matrix mit Zeilen ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_p$ und Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Dann gilt:*

a) *$T_{ij}(\lambda) \cdot A$ geht aus A durch die Zeilenumformung $\mathbf{z}_i \longmapsto \mathbf{z}_i + \lambda \mathbf{z}_j$ hervor, $A \cdot T_{ij}(\lambda)$ durch die Spaltenumformung $\mathbf{s}_j \longmapsto \mathbf{s}_j + \lambda \mathbf{s}_i$. (Dabei ist einmal $T_{ij}(\lambda) \in M(p \times p, K)$, einmal $T_{ij}(\lambda) \in M(n \times n, K)$!)*

b) *$P_{ij} \cdot A$ geht aus A durch Vertauschen von i-ter und j-ter Zeile hervor, $A \cdot P_{ij}$ durch Vertauschen von i-ter Spalte und j-ter Spalte.*

Beweis. Man rechnet das nach. Zum Beispiel für $T_{ij}(\lambda) \cdot A$ bemerkt man zunächst, dass diese Matrix in allen Zeilen außer der i-ten mit A übereinstimmt, da für $k \neq i$ die k-te Zeile von $T_{ij}(\lambda)$ der k-te Standard-Einheitsvektor, also gleich der k-ten Zeile der Einheitsmatrix ist.
In der i-ten Zeile hat $T_{ij}(\lambda) \cdot A$ in der il-Position den Eintrag

$$1 \cdot a_{il} + \lambda \cdot a_{jl},$$

also den Eintrag, der durch Addition der mit λ multiplizierten j-ten Zeile zur i-ten Zeile entsteht.
Genauso rechnet man die anderen Behauptungen nach (Übung). □

Satz 6.11. *Sei $A \in M(n \times n, K)$ regulär ($\mathrm{rg}(A) = n$). Dann gibt es Matrizen $T_1, \ldots, T_r$, die alle von einem der Typen $T_{ij}(\lambda), P_{ij}, D_i(\lambda)$ sind, so dass*

$$T_r \cdots T_1 A = E_n$$

gilt.
Wendet man die entsprechenden elementaren Umformungen (in der gleichen Reihenfolge) auf E_n an, so erhält man die zu A inverse Matrix A^{-1}.
Beschränkt man die Matrizen T_k von oben auf Elementarmatrizen, so erreicht man immerhin noch, dass

$$T_r \cdots T_1 A = D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

eine Diagonalmatrix *mit $d_1 \cdots d_n \neq 0$ ist. Man kann in diesem Fall sogar noch erreichen, dass*

$$D = \begin{pmatrix} 1 & & 0 & \\ & \ddots & & \\ 0 & & 1 & \\ & & & d \end{pmatrix}$$

gilt. Wendet man die diesen Elementarmatrizen entsprechenden elementaren Umformungen vom Typ i) (in der gleichen Reihenfolge) auf E_n an, so erhält man $DA^{-1} =: B$, also $A^{-1} = D^{-1}B$.

Beweis. Wir haben gesehen, dass jede elementare Zeilenumformung der Matrix A durch Multiplikation der Matrix von links mit einer geeigneten Matrix T realisiert werden kann; dabei ist T entweder eine Elementarmatrix, eine Permutationsmatrix oder eine Diagonalmatrix $D_i(\lambda)$.
Da der Rang der Matrix A gleich ihrer Zeilenanzahl n ist, ist die reduzierte Zeilenstufenform die Einheitsmatrix (keine Zeile ist die Nullzeile, und es ist kein Platz da für Stufen, die um mehr als einen Index springen). Bringt man also die Matrix A durch elementare Umformungen, die Multiplikation von links mit Matrizen $T_1, \ldots, T_r$ entsprechen, auf reduzierte Zeilenstufenform, so erhält man wie behauptet $T_r \cdots T_1 A = E_n$.
Um die Behauptung für auf Elementarmatrizen beschränkte T_k zu zeigen, müssen wir noch einmal den Beweis für die Möglichkeit der Transformation einer beliebigen Matrix auf (reduzierte) Zeilenstufenform (Satz 2.10) durchlaufen und sehen, dass wir in der gegebenen speziellen Situation einer quadratischen $n \times n$- Matrix vom vollen Rang n nur die (durch Multiplikation mit Elementarmatrizen darstellbaren) Umformungen vom Typ i) benötigen (also auf die Transformationen der Typen ii) (Multiplikation einer Zeile mit $\lambda \neq 0$) und iii) (Vertauschen zweier Zeilen) verzichten können) wenn wir statt der reduzierten Zeilenstufenform E_n der Matrix nur die etwas allgemeinere Diagonalgestalt erreichen wollen.

Wegen der rekursiven Struktur des Beweises (bzw. des algorithmischen Verfahrens) müssen wir nur den Rekursionsschritt überprüfen, der das Problem auf das gleiche Problem mit um eins verminderter Zeilen- und Spaltenzahl zurückführt.
Eine Vertauschung zweier Zeilen nimmt man in diesem Schritt dann vor, wenn die erste Zeile im $1,1$- Eintrag eine Null hat. Da der Rang der Matrix n ist, ist irgendein a_{i1} ungleich Null, und durch Addition der i-ten Zeile zur ersten (Typ i)!) erreicht man auch $a_{11} \neq 0$. Eine Multiplikation einer Zeile mit $\lambda \neq 0$ benutzt man dann, wenn man den ersten von 0 verschiedenen Eintrag einer Zeile zu 1 machen will. Streben wir (wie im ersten Teil der Behauptung) nur Diagonalgestalt an, so können wir auf diesen Schritt ebenfalls verzichten.
Wir müssen uns jetzt nur noch überzeugen, dass wir mit ausschließlicher Benutzung der Umformungen vom Typ i) die Matrix $\left(\begin{smallmatrix} d_1 & 0 \\ 0 & d_2 \end{smallmatrix}\right)$ in die Gestalt $\left(\begin{smallmatrix} 1 & 0 \\ 0 & d_1 d_2 \end{smallmatrix}\right)$ umformen können. Dazu durchlaufen wir die folgenden Schritte:

- Addiere die erste Zeile zur zweiten
- Addiere die mit $\frac{1-d_1}{d_1}$ multiplizierte zweite Zeile zur ersten Zeile
- Addiere die mit $-d_1$ multiplizierte erste Zeile zur zweiten Zeile
- Die Matrix hat jetzt die Gestalt $\left(\begin{smallmatrix} 1 & c \\ 0 & d_1 d_2 \end{smallmatrix}\right)$; man addiere noch die mit $\frac{-c}{d_1 d_2}$ multiplizierte zweite Zeile zur ersten und hat die gewünschte Gestalt erreicht.

□

Korollar 6.12. *Ist A invertierbar, so erhält man die Inverse von A, indem man die Matrix $(A|E_n) \in M(n \times 2n, K)$ durch elementare Zeilenumformungen auf* reduzierte Zeilenstufenform (reduced row echelon form) *bringt; das Ergebnis ist dann $(E_n|A^{-1})$.*

Beweis. Schreiben wir wie im vorigen Satz

$$T_r \cdots T_1 A = E_n,$$

so folgt

$$A^{-1} = (T_r \cdots T_1) E_n,$$

was die Behauptung zeigt. □

Bemerkung: In Satz 6.11 und Korollar 6.12 kann man genauso gut mit Spaltenumformungen statt mit Zeilenumformungen arbeiten (A wird dann von rechts mit Elementarmatrizen multipliziert).

Beispiel: Siehe das Maple-Worksheet auf der Webseite der Vorlesung.

Definition und Lemma 6.13. *Eine Matrix $A = (a_{ij}) \in M_n(K) = M(n \times n, K)$ heißt ober Dreiecksmatrix, wenn $a_{ij} = 0$ für $i > j$ gilt, untere Dreicksmatrix, wenn $a_{ij} = 0$ für $i < j$ gilt.*

a) *Sind $A \in M(r \times p, K)$ und $B \in M(p \times n, K)$ obere Dreiecksmatrizen (untere Dreiecksmatrizen), so ist AB obere (bzw. untere) Dreiecksmatrix.*
b) *Eine Dreiecksmatrix $A \in M_n(K)$ ist genau dann invertierbar, wenn alle Diagonaleinträge a_{ii} von Null verschieden sind.*

Beweis. Man rechnet das nach. Seien etwa $A = (a_{ij}), B = (b_{jk})$ obere Dreiecksmatrizen, also $a_{ij} = 0$ falls $i > j$ und $b_{jk} = 0$ falls $j > k$.
Der ik-Eintrag c_{ik} von $C := AB$ ist $c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$. Ist $i > k$, so gibt es kein j mit $i \leq j, j \leq k$, also ist in obiger Summe stets $a_{ij} = 0$ oder $b_{jk} = 0$, also $c_{ik} = 0$, damit ist C ebenfalls eine obere Dreiecksmatrix.
Die Rechnung für untere Dreiecksmatrizen verläuft analog.
Für b) nehmen wir an, ein Diagonaleintrag der Dreiecksmatrix A sei gleich 0 und setzen $i_0 = \min\{1 \leq i \leq n \mid a_{ii} = 0\}$. Dann sieht man sofort, dass der von den ersten i_0 Spalten (bei oberer Dreiecksmatrix) bzw. Zeilen (bei unterer Dreiecksmatrix) aufgespannte Teilraum des K^n Dimension $i_0 - 1$ hat; die Matrix kann also nicht Rang n haben und kann damit nicht invertierbar sein.
Umgekehrt beweist man im Fall, dass alle Diagonaleinträge von 0 verschieden sind, leicht die lineare Unabhängigkeit der Spaltenvektoren (bzw. der Zeilenvektoren) der oberen (bzw. unteren) Dreiecksmatrix A. □

Satz 6.14. *Sei $A \in M(p \times n, K)$.*

a) *Es gibt eine invertierbare Matrix $T \in M(p \times p, K)$, so dass TA Zeilenstufenform hat. T kann als Produkt von Elementarmatrizen und Permutationsmatrizen gewählt werden.*
b) *A ist äquivalent zu $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ und zu $\begin{pmatrix} 0 & E_r \\ 0 & 0 \end{pmatrix}$ mit $r = \mathrm{rg}(A)$.*
c) *(LU-*Zerlegung*, auch LR-*Zerlegung* genannt): Es gibt eine Permutationsmatrix P, so dass man $PA = LU$ mit einer unteren Dreiecksmatrix (Englisch: lower triangular matrix) $L \in M(p \times p, K)$ und einer oberen Dreiecksmatrix (Englisch: upper triangular matrix) $U \in M(p \times n, K)$ schreiben kann. Man erhält diese Zerlegung, indem man PA durch Gauß-Elimination auf Zeilenstufenform U bringt (unter Verzicht auf die Bedingung $a_{k,s(i)} = 0$ für $1 \leq k < i \leq r$); L^{-1} ist das Produkt der zugehörigen Elementarmatrizen.*

Beweis. a) ist klar: Wie bei Satz 2.10 bringt man die Matrix A durch elementare Zeilenumformungen in Zeilenstufenform und findet T als das Produkt derjenigen Matrizen, die durch Linksmultiplikation die benutzten Umformungen liefern. Da wir hier auf die Normierungsbedingung verzichten, dass das erste von 0 verschiedene Element jeder Zeile gleich 1 ist, benötigen wir dabei keine Matrizen vom Typ $D_i(\lambda)$ und kommen mit Elementarmatrizen und Permutationsmatrizen aus.

Bei b) ist nach Definition der Äquivalenz von Matrizen zu zeigen, dass es invertierbare Matrizen T' und S gibt, so dass

$$T'AS = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

gilt. Wir wissen bereits, dass wir A durch eine Kombination elementarer Zeilenumformungen und elementarer Spaltenumformungen in diese Gestalt bringen können. Bezeichnet man mit T' das Produkt der Matrizen, die zu den benötigten elementaren Zeilenumformungen gehören, mit S das Produkt der Matrizen, die zu den benötigten elementaren Spaltenumformungen gehören, so hat man die gesuchten Matrizen.
Die Gestalt

$$\begin{pmatrix} 0 & E_r \\ 0 & 0 \end{pmatrix}$$

erreicht man durch weitere Spaltenvertauschungen.
Man kann b) aber auch ganz anders zeigen: Das Bild von L_A hat Dimension $r = \mathrm{rg}(A)$. Nach Satz 4.29 können wir eine Basis $(v_1, \ldots, v_n)$ von K^n finden, in der die ersten $n-r$ Vektoren eine Basis von $\mathrm{Ker}(L_A)$ bilden und die verbleibenden r Vektoren durch L_A auf eine Basis $(w_1, \ldots, w_r)$ von $\mathrm{Im}(L_A)$ abgebildet werden. Ergänzt man letztere noch zu einer Basis von K^p, so hat L_A bezüglich dieser Basen die Matrix $\begin{pmatrix} 0 & E_r \\ 0 & 0 \end{pmatrix}$, was nach Lemma 6.8 die Behauptung zeigt.
Für c) muss man erneut in den Ablauf des Gauß - Algorithmus einsteigen; wir skizzieren das hier nur.
Zunächst definiere man für $1 \leq i \leq r$ wie schon früher $s(i)$ als das kleinste j, für das die Teilmatrix aus den ersten j Spalten von A den Rang i hat. Man überlegt sich dann, dass man durch geeignete Zeilenvertauschungen erreichen kann, dass in der so umgeformten Matrix A' für alle $1 \leq i \leq r$ die Teilmatrix aus den ersten i Zeilen und den ersten $s(i)$ Spalten Rang i hat. Da Zeilenvertauschungen durch Linksmultiplikation mit elementaren Permutationsmatrizen erreicht werden, findet man also eine Permutationsmatrix P, für die $PA = A'$ diese Eigenschaft hat.
Durchläuft man nun den Gauß-Algorithmus, so sieht man, dass die einzige Umformung, die man benötigt, um A' auf (nicht reduzierte) Zeilenstufenform zu bringen, Umformungen vom folgenden Typ sind:
Addition der mit $\lambda \in K$ multiplizierten j-ten Zeile zur i-ten Zeile für ein Paar (i, j) mit $j < i$.
Diese Umformungen werden durch Linksmultiplikation mit Elementarmatrizen bewirkt, die untere Dreiecksmatrizen sind, ihr Produkt bezeichnet man mit L^{-1}. Die Matrix in Zeilenstufenform ist eine obere Dreiecksmatrix, die wir U nennen.
Damit haben wir

$$PA = LU$$

erreicht.
Der Verzicht auf die Reduziertheit der Zeilenstufenform von U im obigen Argument ist wesentlich: Will man reduzierte Zeilenstufenform erreichen, so muss man weiter unten stehende Zeilen zu weiter oben stehenden Zeilen addieren; dafür benötigt man Elementarmatrizen, die obere statt unterer Dreiecksmatrizen sind. □

Beispiel: Siehe das Maple-Worksheet auf der Webseite der Vorlesung.

Bemerkung. *Die LU-Zerlegung spielt in der numerischen linearen Algebra eine wichtige Rolle, Sie werden ihr in der Vorlesung „Praktische Mathematik“ wieder begegnen.*

Zusammenfassung:

Hat der K-Vektorraum V Basen $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{B}' = (v'_1, \ldots, v'_n)$ mit $v'_j = \sum_{i=1}^n s_{ij} v_i$, so ist $S = (s_{ij})$ die Matrix des Basiswechsels von $\mathcal{B}$ zu $\mathcal{B}'$ (Übergangsmatrix von $\mathcal{B}$ zu $\mathcal{B}'$), sie ist gleich $M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_V)$ sowie gleich $M_{\mathcal{B}}^{\mathcal{B}}(f)$, wo f die lineare Abbildung von V in sich mit $f(v_j) = v'_j$ $(1 \leq j \leq n)$ ist.

Ist $f : V \longrightarrow W$ linear, $A = M_{\mathcal{C}}^{\mathcal{B}}(f) \in M(p \times n, K)$ mit Basen $\mathcal{B}$ von V, $\mathcal{C}$ von W, so gehört $A' \in M(p \times n, K)$ genau dann zu f bezüglich Basen $\mathcal{B}'$ von V, $\mathcal{C}'$ von W, wenn $A' = T^{-1}AS$ mit $T \in \mathrm{GL}_n(K)$, $S \in \mathrm{GL}_n(K)$ gilt; S und T sind dabei die Übergangsmatrizen von $\mathcal{B}$ zu $\mathcal{B}'$ bzw. von $\mathcal{C}$ zu $\mathcal{C}'$; A und A' heißen in diesem Fall äquivalent. Ist $\mathrm{rg}(A) = r$, so ist A äquivalent zu $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ (wobei die Einträge 0 Nullmatrizen geeigneter Größe bezeichnen).

Ist A quadratisch $(p = n)$ und $A' = S^{-1}AS$ mit $S \in \mathrm{GL}_n(K)$, so heißen A und A' ähnlich (oder konjugiert); äquivalent ist, dass $A = M_{\mathcal{B}}^{\mathcal{B}}(f)$, $A' = M_{\mathcal{B}'}^{\mathcal{B}'}(f)$ für geeignete Basen eines n-dimensionalen K-Vektorraums V ist, $f \in \mathrm{End}(V)$.

Elementare Zeilenumformungen einer Matrix entsprechen speziellen Basiswechseln $(w_i \longmapsto w_i + \lambda w_j)$ im Bildraum, elementare Spaltenumformungen speziellen Basiswechseln im Urbildraum, sie können auch durch Links- bzw. Rechtsmultiplikation mit Elementarmatrizen realisiert werden. Dies führt auf ein Berechnungsverfahren zur Inversenberechnung mittels elementarer Umformungen (Korollar 6.12) sowie zu Zerlegungen von Matrizen in Produkte von Dreiecksmatrizen und Permutationsmatrizen (LU-Zerlegung, LR-Zerlegung).

7. Gruppen, Permutationen, Determinante

Ziel dieses Abschnitts ist eine Formel für das Volumen eines Parallelepipeds (Parallelotops) (= verallgemeinertes Parallelogramm, "schiefer Quader") im $\mathbb{R}^n$.

Im $\mathbb{R}^3$ weiß man: Spannen $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ das Parallelepiped (Spat) $P = \{x\mathfrak{a} + y\mathfrak{b} + z\mathfrak{c} \mid 0 \leq x, y, z \leq 1\}$ auf, so ist

$$\mathrm{vol}(P) = |\mathfrak{a} \cdot (\mathfrak{b} \times \mathfrak{c})| = |\det(\mathfrak{a}, \mathfrak{b}, \mathfrak{c})|,$$

dabei ist für $A = (a_{ij}) \in M_3(\mathbb{R})$ die Determinante $\det(A)$ durch die Regel von Sarrus gegeben:

$$\begin{aligned} \det A &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned}$$

$\det A$ hat die Eigenschaften

- $\det(A) = 0$, falls zwei Zeilen (zwei Spalten) gleich sind (es reicht sogar: die Spalten (bzw. Zeilen) von A sind linear abhängig).
- $\det(\mathfrak{a} + \lambda\mathfrak{a}', \mathfrak{b}, \mathfrak{c}) = \det(\mathfrak{a}, \mathfrak{b}, \mathfrak{c}) + \lambda \det(\mathfrak{a}', \mathfrak{b}, \mathfrak{c})$ (det ist linear in der ersten Spalte), genauso in den anderen Spalten bzw. in den Zeilen.

Definition 7.1. *Sei V ein K-Vektorraum, $n \in \mathbb{N}$.*

a) *$M : V^n \longrightarrow K$ heißt (n-fache) Multilinearform, wenn gilt:*

$$M(v_1, \ldots, v_{i-1}, v_i + \lambda v_i', v_{i+1}, \ldots, v_n) = \\ M(v_1, \ldots, v_n) + \lambda M(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_n) \ (1 \leq i \leq n)$$

(M ist linear in jeder Komponente, bei festgehaltenen restlichen Komponenten).

b) *Ist M aus a) multilinear, so heißt M alternierend, wenn gilt: Sind zwei der v_i gleich, so ist $M(v_1, \ldots, v_n) = 0$.*

c) *Ist $V = K^n$ und $M = d$ wie in a), b) eine alternierende n-fache Multilinearform mit $d(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$ ($(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ die Standardbasis), so heißt d Determinantenfunktion. Ist $A = (\mathbf{s}_1, \ldots, \mathbf{s}_n) \in M_n(K)$ eine Matrix mit Spaltenvektoren $\mathbf{s}_j$, so schreibt man*

$$d(A) = M(\mathbf{s}_1, \ldots, \mathbf{s}_n)$$

und fasst d auf diese Weise als Abbildung $d : M_n(K) \to K$ auf, die als Funktion der Spalten linear in jeder Spalte ist.

Am Beispiel $n = 3$ sieht man, dass Permutationen der Indizes und Verteilungen von Vorzeichen eine Rolle spielen. Dafür brauchen wir etwas Vorlauf.

Erinnerung: Definition Gruppe, charakterisiert durch Gruppenaxiome G1, G2, G3, siehe Definition 3.3, Lemma 3.4, $S_X = \mathrm{Perm}(X)$ die Gruppe der Permutationen der Menge X, speziell: $S_n := S_{\{1,\ldots,n\}}$, siehe das Beispiel nach Definition 3.3.

Definition 7.2. *Seien G, H Gruppen, $f : G \longrightarrow H$ eine Abbildung.*

a) *f heißt* Homomorphismus *(von Gruppen), falls*

$$f(g_1 \circ_G g_2) = f(g_1) \circ_H f(g_2)$$

für alle $g_1, g_2 \in G$ gilt. Ist f zusätzlich bijektiv, so heißt f ein Isomorphismus *von Gruppen und man sagt, G und H seien isomorph ($G \cong H$).*

b) *Ist $f : G \longrightarrow H$ Homomorphismus, so heißt*

$$\mathrm{Ker}(f) := \{g \in G \mid f(g) = e_H\}$$

der Kern von f.

Beispiel: $G = (V, +)$, $H = (W, +)$ seien die additiven Gruppen der K-Vektorräume V, W. Dann sind alle linearen Abbildungen $F : V \longrightarrow W$ erst recht Gruppenhomomorphismen.

Lemma 7.3. *Der Homomorphismus $f : G \longrightarrow H$ ist genau dann injektiv, wenn $\mathrm{Ker}(f) = \{e_G\}$ gilt.*

Beweis. Genau wie bei linearen Abbildungen: Übung. □

Definition 7.4. *Sei X eine Menge, G eine Gruppe. Eine Operation von G auf X ist gegeben durch eine Abbildung*

$$\begin{array}{ccc} G \times X & \longrightarrow & X \\ (g, x) & \longmapsto & g.x = g(x) = gx \end{array}$$

mit den Eigenschaften

a) *$(g_1 g_2).x = g_1.(g_2.x)$ für alle $g_1, g_2 \in G$, $x \in X$.*
b) *$e.x = x$ für alle $x \in X$.*

Beispiel:

a) S_X (und alle seine Untergruppen) operiert auf X durch

$$\varphi.x := \varphi(x).$$

b) Ist $f : G \longrightarrow S_X$ ein Homomorphismus, so operiert G auf X durch

$$g.x := (f(g))(x).$$

Definition und Lemma 7.5. *Die Gruppe G operiere auf der Menge X.*

a) *Durch $x_1 \sim x_2 \Leftrightarrow \exists g \in G$ mit $gx_1 = x_2$ wird eine Äquivalenzrelation gegeben.*
Die Äquivalenzklassen heißen die Bahnen (Orbits) der Operation, die Klasse von $x \in X$ heißt $B_x = O_x$. Falls es nur eine Klasse gibt, so heißt die Operation transitiv. Falls die Bahn nur ein Element hat, heißt sie trivial.

b) *Für jedes $x \in X$ ist $\mathrm{Stab}_G(x) := G_x := \{g \in G \mid gx = x\}$ eine Untergruppe von G, $G_x = \mathrm{Stab}_G(x)$ heißt Stabilisator (Fixgruppe, Standgruppe) von x (in G).*

Beweis. a) Klar (Nachrechnen).

b) Nachzuprüfen ist:

$e \in G_x$

$g \in G_x \Rightarrow g^{-1} \in G_x$

$g_1, g_2 \in G_x \Rightarrow g_1 g_2 \in G$.

Alle drei Aussagen rechnet man sofort nach. □

Beispiel:

a) $G = S_X$ operiert transitiv auf X, $\mathrm{Stab}_G(x) \cong S_{X \setminus \{x\}}$.

b) $\{e\} \subseteq S_X$ operiert auf X, alle Bahnen sind einelementig.

c) $\mathrm{GL}_n(K)$ operiert auf K^n mit zwei Bahnen: $\{\mathbf{0}\}$ und $K^n \setminus \{\mathbf{0}\}$.

Satz 7.6. *(Bahnformel) Sei G eine endliche Gruppe, X eine endliche Menge, auf der G operiert, $x \in X$.*
Dann ist $|O_x| \cdot |\mathrm{Stab}_G(x)| = |G|$.

Beweis. Sei $O_x = \{x = x_1, \ldots, x_r\}$, $r = |O_x|$.
Für $1 \leq i \leq r$ sei $M_i = \{g \in G \mid gx = x_i\}$ (die M_i sind die Äquivalenzklassen der Relation $g \sim g' \Leftrightarrow gx = g'x$ auf G), wir fixieren jeweils ein $g_i \in M_i$.
Wir haben dann $gx = g_i x \Leftrightarrow g_i^{-1} g \in \mathrm{Stab}_G(x) = \{e = h_1, h_2, \ldots, h_s\}$.
Also ist $M_i = \{g_i h_1, \ldots, g_i h_s\}$ und daher $|M_i| = s = |\mathrm{Stab}_G(x)|$ für alle i.
Da offenbar $G = \dot{\bigcup}_{i=1}^{r} M_i$ (disjunkte Vereinigung) gilt, folgt die Behauptung. □

Korollar 7.7. *Für $n \in \mathbb{N}$ ist $|S_n| = n! = 1 \cdot 2 \cdots n$.*

Beweis. S_n operiert transitiv (d.h., mit nur einer Bahn) auf $\{1, \ldots, n\}$, mit $\mathrm{Stab}_{S_n}(n) \cong S_{n-1}$ für $n \geq 1$.
Damit funktioniert ein Beweis durch vollständige Induktion: $n = 1$ ist trivial.
Die Behauptung sei richtig für n. Dann gilt:

$$|S_{n+1}| = \underbrace{(n+1)}_{|O_{n+1}|} \cdot |S_n| = (n+1)n! = (n+1)!$$

□

Definition 7.8. a) *Eine Permutation $\sigma \in S_n$ heißt eine Transposition, wenn es $i \neq j \in \{1, \ldots, n\}$ gibt mit $\sigma(i) = j$, $\sigma(j) = i$, $\sigma(k) = k$ für alle $k \neq i$, $k \neq j$. Man schreibt dann $\sigma = (i, j) = (ij)$*

b) *Eine Permutation $\sigma \in S_n$ heißt ein Zykel der Länge r (ein r-Zykel), wenn es paarweise verschiedene $i_k \in \{1, \ldots, n\}$ gibt ($1 \leq k \leq r$) mit $\sigma(i_k) = i_{k+1}$ ($1 \leq k < r$), $\sigma(i_r) = i_1$, $\sigma(j) = j$ für $j \notin \{i_1, \ldots, i_r\}$. Man schreibt dann $\sigma = (i_1, \ldots, i_r)$.*

Lemma 7.9. a) *Jede Permutation $\sigma \in S_n$ ($n \geq 2$) kann als Produkt $\sigma = \tau_1 \cdots \tau_r$ ($r \in \mathbb{N}$) von Transpositionen $\tau_1, \ldots, \tau_r$ geschrieben werden.*

b) *Jede Permutation kann eindeutig als Produkt elementfremder Zykel geschrieben werden.*

Beweis. a) Induktion nach n. $n = 2$ ist klar:
$\mathrm{Id} = (12) \circ (12)$, $(12) = (12)$.
Die Behauptung sei richtig für ein $n \geq 2$, sei $\sigma \in S_{n+1}$. Ist $\sigma(n+1) = n+1$, so schreibe $\sigma|\{1, \ldots, n\}$ nach Induktionsannahme als Produkt von Transpositionen, σ selbst ist dann das Produkt der gleichen Transpositionen (aufgefasst als Elemente von S_{n+1}, die $n+1$ fixieren).
Andernfalls sei $\sigma(n+1) = k \neq n+1$ und $\tau = (k, n+1)$. Dann ist $(\tau \circ \sigma)(n+1) = n+1$, also $\rho := \tau \circ \sigma$ Produkt von Transpositionen, also ist $\sigma = \tau \circ \rho$ ebenfalls ein Produkt von Transpositionen.

b) Ist σ ein r-Zykel und $I = \{i_1, \ldots, i_r\}$, so ist I eine Bahn der Untergruppe $\langle \sigma \rangle := \{\sigma^j \mid j \in \mathbb{Z}\}$ von S_n und $\sigma|_{\{1,\ldots,n\} \setminus I} = \mathrm{Id}$, d.h., $\langle \sigma \rangle$ hat nur eine nicht triviale Bahn, nämlich I.
Hat umgekehrt $\langle \sigma \rangle$ für ein $\sigma \in S_n$ nur die eine nicht triviale Bahn I und ist $i_1 \in I$, so sei $r = \min\{j \in \mathbb{N} \mid \sigma^j(i_1) = i_1\}$; r ist endlich, denn $\{\sigma^j(i_1)\}$ ist endlich, also gibt es $\ell > k$ mit $\sigma^k(i_1) = \sigma^\ell(i_1)$, also $\sigma^{\ell-k}(i_1) = i_1$. Die $\sigma^j(i_1)$ für $1 \leq j \leq r$ sind paarweise verschieden, denn sonst wäre $\sigma^{j'-j}(i_1) = i_i$ mit $0 < j' - j < r$, Widerspruch.
Dann ist offenbar $I = (i_1, \sigma(i_1), \ldots, \sigma^{r-1}(i_1)) = (i_1, \ldots, i_r)$.

Die Behauptung folgt jetzt durch Induktion:
Sei $\sigma \in S_n$, $I \subseteq \{1, \ldots, n\}$ eine Bahn von σ, ρ der zugehörige Zykel. Dann ist $\rho^{-1}\sigma|_I = I$. ρ^{-1} kann also als Permutation von $n - |I|$ Elementen eindeutig als Produkt von elementfremden Zykeln $\rho_1, \ldots, \rho_t$ geschrieben werden, in denen die Elemente von I nicht vorkommen.
Dann ist $\sigma = \rho\rho_1 \ldots \rho_t$ Produkt von Zykeln.
Die Eindeutigkeit zeige man als Übung.

□

Beispiel:

$$\begin{aligned}
\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 2 & 1 & 6 \end{pmatrix} &= (1\ 3\ 5)(2\ 4)(6) \text{ Zykelzerlegung} \\
&= (1\ 3)(3\ 5)(2\ 4) \text{ Transpositionen} \\
\mathrm{Id}_{\{1,\ldots,6\}} &= (1)(2)(3)(4)(5)(6).
\end{aligned}$$

Definition und Satz 7.10. *Für $\sigma \in S_n$ ist das Signum* $\mathrm{sgn}(\sigma)$ *durch*

$$\mathrm{sgn}(\sigma) := \prod_{i<j} \mathrm{sgn}(\sigma(j) - \sigma(i))$$

definiert, dabei ist $\operatorname{sgn}(k-l)$ *gleich* $+1$*, wenn* $k > l$*, gleich* -1*, wenn* $k < l$ *gilt.*
Es gilt:

a) sgn *ist multiplikativ, also* $\operatorname{sgn}(\sigma_1\sigma_2) = \operatorname{sgn}(\sigma_1)\operatorname{sgn}(\sigma_2)$.
b) $\operatorname{sgn}(\tau) = -1$ *für jede Transposition.*
c) $\operatorname{sgn}(\sigma) = 1 \Leftrightarrow \sigma$ *ist Produkt einer geraden Anzahl von Transpositionen.*
d) *Es gilt* $\operatorname{sgn}(\sigma) = (-1)^\alpha$*, wo* $\alpha = \#\{(i,j) \mid i < j$ *und* $\sigma(i) > \sigma(j)\}$ *die Anzahl der* Fehlstände *von* σ *ist.*

Eine Permutation σ *heißt* gerade*, wenn* $\operatorname{sgn}(\sigma) = +1$ *ist,* ungerade*, wenn* $\operatorname{sgn}(\sigma) = -1$ *ist.*
Die Menge $A_n = \{\sigma \in S_n \mid \operatorname{sgn}(\sigma) = +1\}$ *der geraden Permutationen heißt die* alternierende Gruppe.

Beweis. d) ist klar, c) folgt direkt aus a) und b).
Für b) können wir o.E. $\sigma = (1\ 2)$ annehmen.
Dann ist für $i < j$

$$\operatorname{sgn}(\sigma(j) - \sigma(i)) = \begin{cases} -1 & i = 1,\ j = 2 \\ +1 & \text{sonst.} \end{cases}$$

Bleibt a): Seien $\sigma, \rho \in S_n$ und

$$J = \{(\rho(i), \rho(j)) \mid i < j\}.$$

Gegenüber $\{(i,j) \mid i < j\}$ sind hier einige der Paare umgeordnet (nämlich die mit $\rho(i) > \rho(j)$), also ist $\prod_{i<j} \operatorname{sgn}(\sigma(j) - \sigma(i))$ i.a. von $\prod_{(k,\ell)\in J} \operatorname{sgn}(\sigma(\ell) - \sigma(k))$ verschieden.
Wir haben aber

$$\prod_{i<j} \operatorname{sgn}(\sigma(j) - \sigma(i)) = \prod_{(k,\ell)\in J} \operatorname{sgn}(\sigma(\ell) - \sigma(k))\operatorname{sgn}(\ell - k),$$

denn wenn $k > \ell$ in $(k,\ell) \in J$ gilt, so kommt $\operatorname{sgn}(\sigma(k) - \sigma(\ell))$ als Faktor auf der linken Seite vor, und es ist $\operatorname{sgn}(\sigma(\ell) - \sigma(k)) = -\operatorname{sgn}(\sigma(k) - \sigma(\ell)$

Damit:

$$\begin{aligned} \operatorname{sgn}(\sigma) &= \prod_{i<j} \operatorname{sgn}(\sigma(j) - \sigma(i)) \\ &= \prod_{i<j} \operatorname{sgn}(\sigma(\rho(j)) - \sigma(\rho(i)) \prod_{i<j} \operatorname{sgn}(\rho(j) - \rho(i)) \\ &= \operatorname{sgn}(\sigma\rho)\operatorname{sgn}(\rho) \end{aligned}$$

also die Behauptung. □

Definition und Satz 7.11 (Leibniz'sche Formel)**.** *Sei* K *ein Körper,* $n \in \mathbb{N}$*. Für* $A = (a_{ij}) \in M_n(K) = M(n \times n, K)$ *sei*

$$\det(A) := \sum_{\pi \in S_n} \operatorname{sgn}(\pi)\, a_{1,\pi(1)} a_{2,\pi(2)} \dots a_{n,\pi(n)}.$$

$\det(A)$ *heißt die* Determinante *von A. Es gilt:*

a) $\det(E_n) = 1$
b) *Als Funktion der Zeilen von A ist* det *eine alternierende n-fache Multilinearform auf K^n. Das Gleiche gilt für* det *als Funktion der Spalten von A.*

Die Determinante ist also eine Determinantenfunktion im Sinne von Definition 7.1.
Ferner gilt $\det({}^tA) = \det(A)$.

Beweis. Ist A eine beliebige Diagonalmatrix (also $a_{ij} = 0$ falls $i \neq j$), so sind in der Summe alle Terme außer dem für $\pi = \mathrm{Id}$ gleich 0, die Determinante von A also das Produkt der Diagonalelemente; insbesondere erhält man für $A = E_n$ den Wert 1.
Ist $a_{ij} = a'_{ij} + \lambda a''_{ij}$ (mit $\lambda \in K$) für festes i und $1 \leq j \leq n$ und A', bzw. A'' die Matrix, die aus A durch Ersetzen der Einträge der i-ten Zeile durch die a'_{ij} bzw. die a''_{ij} hervorgeht, so sieht man direkt durch Einsetzen in die Formel, dass $\det(A) = \det(A') + \lambda \det(A'')$ gilt; die Determinante ist also linear als Funktion jeder Zeile. Genauso sieht man, dass sie linear als Funktion jeder Spalte ist. Um zu sehen, dass sie alternierend ist, nehmen wir o. E. an, dass $a_{1j} = a_{2j}$ für alle j gilt. Ist dann $\tau = (12)$ die Transposition von 1 und 2, so heben sich in der definierenden Summe für jedes $\pi \in S_n$ die Terme für π und $\pi \circ \tau$ wegen $\mathrm{sgn}(\pi \circ \tau) = -\mathrm{sgn}(\pi)$ auf, also ist $\det(A) = 0$, die Multilinearform det also alternierend.
Dass sich die Determinante unter Transposition der Matrix nicht ändert, sieht man direkt an der Formel; man gehe von π zu π^{-1} über, was den Wert der Summe nicht ändert. □

Satz 7.12 (Entwicklungsformel von Laplace)**.** *Für $1 \leq i, j \leq n$ und $A \in M_n(K)$ sei A_{ij} die $(n-1) \times (n-1)$-Matrix, die aus A durch Streichen der i-ten Zeile und der j-ten Spalte entsteht* (Streichungsmatrix)*. Dann gilt:*

a) *Für $1 \leq i \leq n$ ist*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

(Entwicklung nach der i-ten Zeile).

b) *Für $1 \leq j \leq n$ ist*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

(Entwicklung nach der j-ten Spalte).

Beweis. Wir beweisen das nur für die Entwicklung nach der ersten Zeile, die anderen Fälle gehen analog bzw. lassen sich durch Transposition

(Entwicklung nach Spalten) oder Vertauschen von Zeilen bzw. Spalten auf diesen Fall zurückführen.
Der Eintrag a_{11} kommt genau bei den $\pi \in S_n$ vor, für die $\pi(1) = 1$ gilt; wenn wir diese π mit $\pi|_{\{2,\ldots,n\}} \in S_{n-1}$ identifizieren erhalten wir $a_{11}\det(A_{11})$ als Beitrag dieser π. Analog kommt der Eintrag a_{1i} genau bei den $\pi \in S_n$ vor, für die $\pi(1) = i$ gilt. Schreiben wir diese als $(1i) \circ \pi'$ mit $\pi' \in S_{n-1}$, so erhalten wir $-a_{1i}\det(A'_{1i})$ als Beitrag dieser π, wobei A'_{1i} aus der Streichungsmatrix A_{1i} hervorgeht, indem man die erste Spalte um $i-2$ Positionen nach rechts in die Position $i-1$ verschiebt. Da man das durch $i-2$ Vertauschungen benachbarter Spalten erreicht, ist $\det(A'_{1i}) = (-1)^i \det(A_{1i})$, und wir erhalten insgesamt $(-1)^{1+i}a_{1i}\det(A'_{1i})$ als Beitrag der $\pi \in S_n$, für die $\pi(1) = i$ gilt. $\square$

Um zu sehen, dass die oben definierte Determinante durch die Eigenschaften a), b) bereits eindeutig festgelegt ist und weitere Eigenschaften für sie zu beweisen, brauchen wir noch ein paar Aussagen über alternierende Multilinearformen:

Lemma 7.13. *Sei $f : V^r \longrightarrow K$ eine r-fache alternierende Multilinearform und $v_1, \ldots, v_r \in V$, $\lambda \in K$. Dann gilt:*

a) *Für $i \neq j$ ist*

$$f(v_1, \ldots, v_{i-1}, v_i + \lambda v_j, v_{i+1}, \ldots, v_r) = f(v_1, \ldots, v_r)$$

(elementare Umformungen vom Typ $v_i \longmapsto v_i + \lambda v_j$ des Vektorsystems $(v_1, \ldots, v_r)$ ändern die Determinante nicht).

b) $f(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_r) = \lambda f(v_1, \ldots, v_r)$

c) *Setzt man $v'_i = v_j$, $v'_j = v_i$ und $v'_k = v_k$ für $i \neq k \neq j$, so ist*

$$f(v'_1, \ldots, v'_r) = -f(v_1, \ldots, v_r)$$

(Vertauschen von zwei Vektoren ändert das Vorzeichen).

Insbesondere ändert sich die in Definition und Satz 7.11 definierte Determinante nicht, wenn man eine elementare Zeilen- oder Spaltenumformung vom Typ i) (Addition der mit λ multiplizierten $j$ten Spalte bzw. Zeile zur i-ten Spalte bzw. Zeile ($i \neq j$)) durchführt.

Beweis. Für a) hat man (falls etwa $j > i$ ist)

$$\begin{aligned} &f(v_1, \ldots, v_{i-1}, v_i + \lambda v_j, v_{i+1}, \ldots, v_j, \ldots, v_r) \\ &\quad = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_j, \ldots, v_r) \\ &\qquad\qquad + \lambda f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_j, \ldots, v_r) \\ &\quad = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_j, \ldots, v_r) \end{aligned}$$

wie behauptet.
b) folgt direkt aus der Definition der Multilinearität.

Für c) findet man (etwa für $j > i$) mit Hilfe von a) und b)

$$\begin{aligned}
&f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_r) \\
&\quad = f(v_1, \ldots, v_{i-1}, v_j + v_i, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_r) \\
&\quad = f(v_1, \ldots, v_{i-1}, v_j + v_i, v_{i+1}, \ldots, v_{j-1}, v_i - (v_j + v_i), v_{j+1}, \ldots, v_r) \\
&\quad = f(v_1, \ldots, v_{i-1}, v_j + v_i, v_{i+1}, \ldots, v_{j-1}, -v_j, v_{j+1}, \ldots, v_r) \\
&\quad = f(v_1, \ldots, v_{i-1}, v_j + v_i - v_j, v_{i+1}, \ldots, v_{j-1}, -v_j, v_{j+1}, \ldots, v_r) \\
&\quad = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, -v_j, v_{j+1}, \ldots, v_r) \\
&\quad = -f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_r)
\end{aligned}$$

□

Bemerkung: Eigenschaft c) begründet das Wort "alternierend". Ist $2 := 1 + 1 \neq 0$, so kann man aus c) die definierende Eigenschaft einer alternierenden Multilinearform zurückgewinnen, falls $1 + 1 = 0$ in K gilt (man sagt auch $\mathrm{char}(K) = 2$), ist das i.a. nicht möglich und c) wird schwächer als die definierende Eigenschaft.

Korollar 7.14. *Ist $d : M_n(K) \longrightarrow K$ eine Determinantenfunktion (siehe Definition 7.1) und $A \in M_n(K)$ mit $\mathrm{rg}(A) < n$, so ist $d(A) = 0$.*

Beweis. Ohne Einschränkung können wir annehmen, dass die erste Spalte $\mathbf{a}_1$ von A eine Linearkombination

$$\mathbf{a}_1 = \sum_{j=2}^{n} \lambda_j \mathbf{a}_j$$

ist. Die Multilinearität von d ergibt dann

$$d(\sum_{j=2}^{n} \lambda_j \mathbf{a}_j, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \sum_{j=2}^{n} \lambda_j d(\mathbf{a}_j, \mathbf{a}_2, \ldots, \mathbf{a}_n)$$

und weil d alternierend ist, sind hier alle Terme auf der rechten Seite gleich 0. □

Satz 7.15. *Die in Satz 7.11 definierte Determinante ist die einzige Determinantenfunktion $d : M_n(K) \longrightarrow K$.*

Beweis. Sei $d : M_n(K) \longrightarrow K$ eine Determinantenfunktion. Ist $\mathrm{rg}(A) < n$, so ist nach dem vorigen Korollar $d(A) = 0 = \det(A)$. Hat die Matrix A vollen Rang, so wissen wir aus dem Beweis von Satz 6.11, dass wir A durch elementare Zeilenumformungen vom Typ i) ($z_i \mapsto z_i + \lambda z_j$ mit $j \neq i$) in Diagonalgestalt

$$D = \begin{pmatrix} 1 & & 0 & \\ & \ddots & & \\ 0 & & 1 & \\ & & & \delta \end{pmatrix}$$

bringen kann. Genauso ist das natürlich durch elementare Spaltenumformungen vom Typ i) möglich, und von diesen wissen wir, dass sie den

Wert der alternierenden n-fachen Multilinearform d nicht ändern (Lemma 7.13 a)), es gilt also $d(A) = d(D)$ und genauso $\det(D) = \det(A)$. Lemma 7.13 b) impliziert dann, dass $d(D) = \delta d(E_n) = d = \det(D) = \det(A)$ ist, was zu zeigen war (ist d eine Determinantenfunktion, so ist nach Definition $d(E_n) = 1$). □

Satz 7.16. (Multiplikativität der Determinante)

a) *Für alle* $A, B \in M(n \times n, K)$ *gilt*

$$\det(AB) = \det(A)\ \det(B).$$

b) *Ist* $A \in M_n(K)$ *invertierbar, so ist* $\det(A) \neq 0$ *und es gilt:*

$$\det(A^{-1}) = (\det(A))^{-1}.$$

c) $\det : GL_n(K) \to K^\times = K \setminus \{0\}$ *ist ein Gruppenhomomorphismus.*

Beweis. a): Hat die Matrix B die Spalten $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$, so hat AB die Spalten $(A\mathbf{b}_1, \ldots, A\mathbf{b}_n)$. Die Abbildung $d_A : M_n(K) \longrightarrow K$, die durch $d_A(B) := \det(AB)$ gegeben ist, ist daher als Funktion der Spalten von B eine n-fache alternierende Multilinearform, wegen der Eindeutigkeit der Determinante (Satz 7.15) folgt also

$$\det(AB) = d_A(B) = d_A(E_n)\det(B) = \det(A)\det(B).$$

Alternativ kann man auch so vorgehen:
Ist $\mathrm{rg}(A) < n$ oder $\mathrm{rg}(B) < n$, so ist auch $\mathrm{rg}(AB) < n$ (Korollar 5.15) und beide Seiten der Gleichung sind 0.
Andernfalls lässt B sich durch elementare Spaltenumformungen vom Typ i) in die Gestalt

$$B' = \begin{pmatrix} 1 & & 0 & \\ & \ddots & & \\ 0 & & 1 & \\ & & & \delta \end{pmatrix}$$

mit $\delta = \det(B)$ bringen, und durch die gleichen Spaltenumformungen wird AB in AB' überführt.
AB und AB' haben also die gleiche Determinante, da Spaltenumformungen vom Typ i) (Addition der mit λ multiplizierten $j$ten Spalte zur i-ten Spalte $(i \neq j)$) bekanntlich die Determinante nicht ändern. Die Matrix AB' hat aber die gleichen ersten $n-1$ Spalten wie A und ihre n-te Spalte ist das δ-fache der n-ten Spalte von A. Die Multilinearität der Determinante impliziert also

$$\det(AB) = \det(AB') = \delta\det(A) = \det(B)\det(A)$$

wie behauptet.
Die Aussagen b) und c) folgen unmittelbar aus a). □

Korollar 7.17. a) *Sind* A *und* A' *in* $M_n(K)$ *zueinander ähnliche (konjugierte) Matrizen, so ist* $\det(A) = \det(A')$.

b) *Sei V ein endlich dimensionaler K-Vektorraum, $f \in \mathrm{End}(V)$ ein Endomorphismus von V. Dann ist die Determinante* $\det(M_{\mathcal{B}}(f))$ *von f bezüglich einer Basis $\mathcal{B}$ von V unabhängig von der Wahl der Basis $\mathcal{B}$.*

Beweis. a) ist klar, da ja A und A' genau dann ähnlich zueinander sind, wenn $A' = S^{-1}AS$ mit einer invertierbaren Matrix $S \in M_n(K)$ gilt, woraus wegen der Multiplikativität der Determinante sofort $\det(A) = \det(A')$ folgt.
Da zwei Matrizen genau dann den gleichen Endomorphismen bezüglich verschiedener Basen repräsentieren, wenn sie zueinander ähnlich sind, ist auch b) klar, da sie dann nach a) die gleiche Determinante haben. □

Definition 7.18. (Determinante eines Endomorphismus) *Sei V ein endlich dimensionaler K-Vektorraum und $f \in \mathrm{End}(V)$ ein Endomorphismus von V. Dann ist die Determinante* $\det(f)$ *von f definiert als* $\det(f) := \det(M_{\mathcal{B}}(f))$ *für eine beliebige Basis $\mathcal{B}$ von V.*

Bemerkung. Die Menge $\mathrm{SL}_n(K) := \{A \in M(n \times n, K) \mid \det(A) = 1\}$ ist eine Untergruppe von $\mathrm{GL}_n(K)$; sie heißt die *spezielle lineare Gruppe.*
$\mathrm{SL}_n(K)$ besteht genau aus den Matrizen, die sich als Produkt von Elementarmatrizen $T_{ij}(\lambda)$ schreiben lassen.
Zum Beweis überlegt man sich:
Die Multiplikativität der Determinante ist gleichwertig zu der Aussage, dass det ein Gruppenhomomorphismus von der Gruppe $GL_n(K)$ in die multiplikative Gruppe $K^\times$ des Körpers K ist. Der Kern dieses Gruppenhomomorphismus ist offenbar $SL_n(K)$, diese Menge ist also eine Untergruppe.
Elementarmatrizen T haben, wie man sofort sieht, Determinante 1, also haben auch alle Produkte von Elementarmatrizen Determinante 1. Umgekehrt folgt aus Satz 7.15, dass man für jede Matrix $A \in M_n(K)$ der Determinante 1 ein Produkt T von Elementarmatrizen finden kann, so dass $TA = E_n$ und damit $A = T^{-1}$ gilt. Da mit T auch T^{-1} ein Produkt von Elementarmatrizen ist, folgt auch die andere Richtung der Behauptung.

Korollar 7.19. a) *Sei $T = (t_{ij}) \in M(n \times n, K)$ eine Dreiecksmatrix (obere oder untere). Dann ist* $\det(T) = t_{11} \cdots t_{nn}$.

b) *Sei* $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \in M(n \times n, K)$ *eine* Blockmatrix *mit $A \in M(r \times r, K)$, $B \in M(r \times (n-r), K)$, $C \in M((n-r) \times (n-r), K)$. Dann ist* $\det\left(\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}\right) = \det(A)\ \det(C)$.

Beweis. Übung. Man benutze entweder die Formel von Leibniz oder geeignete elementare Umformungen. □

Satz 7.20. *Zu* $A = (a_{ij}) \in M(n \times n, K)$ *sei die* Komplementärmatrix $\tilde{A} = (\tilde{a}_{ij})$ *definiert durch*

$$\tilde{a}_{ij} = (-1)^{i+j} \det(A_{ji})$$

(A_{ji} *die Streichungsmatrix, die durch Streichen der* j*-ten Zeile und der* i*-ten Spalte von* A *entsteht). Dann gilt*

$$A\tilde{A} = \tilde{A}A = \det(A) \cdot E_n.$$

Insbesondere gilt: $A \in M_n(K)$ *ist genau dann invertierbar, wenn* $\det(A) \neq 0$ *gilt, und dann ist*

$$A^{-1} = \frac{1}{\det(A)} \tilde{A}.$$

Beweis. Man überlegt sich zunächst leicht mit Hilfe des Entwicklungssatzes, dass die Matrix B_{ji}, die aus A durch Ersetzen der i-ten Spalte durch den j-ten Standardeinheitsvektor $\mathbf{e}_j$ hervorgeht, Determinante $(-1)^{i+j} \det(A_{ji})$ hat (Entwickeln nach der i-ten Spalte).
Dann ist aber der ik-Eintrag von $\tilde{A}A$ gleich

$$\begin{aligned} \sum_{j=1}^{n} \tilde{a}_{ij} a_{jk} &= \sum_{j=1}^{n} (-1)^{i+j} \det(A_{ji}) a_{jk} \\ &= \sum_{j=1}^{n} a_{jk} \det(B_{ji}) \\ &= \det(C_i), \end{aligned}$$

wo C_i aus A hervorgeht, indem man die i-te Spalte von A durch $\sum_{j=1}^{n} a_{jk}\mathbf{e}_j$ ersetzt, also durch die k-te Spalte von A. Offensichtlich ist $C_k = A$ und, weil det alternierend ist, $\det(C_i) = 0$ für $i \neq k$, also haben wir $\tilde{A}A = \det(A)E_n$. Durch Übergang zur Transponierten berechnet man das Produkt in der umgekehrten Reihenfolge.
Der Rest der Behauptung ist klar (dass invertierbare Matrizen von 0 verschiedene Determinante haben, wissen wir bereits, siehe Satz 7.16) □

Bemerkung: Dieser Satz ist für die praktische Inversenberechnung weniger geeignet als das Verfahren mit Hilfe elementarer Umformungen aus dem vorigen Abschnitt. Man kann aber die Aussage des Satzes benutzen, um theoretische Aussagen (Abschätzungen, Differenzierbarkeit) über die Abhängigkeit der Inversen von den Einträgen der Matrix A zu beweisen.

Satz 7.21. (Cramer'sche Regel) *Sei* $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n) \in \mathrm{GL}_n(K)$, $\mathbf{b} \in K^n$.

Dann lässt sich die (eindeutig bestimmte) Lösung $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ *des linearen Gleichungssystems* $A\mathbf{x} = \mathbf{b}$ *durch*

$$x_j = \frac{\det A_j}{\det A} \quad (1 \leq j \leq n)$$

mit $A_j = (\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{b}, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n)$ *berechnen.*

Beweis. Wir haben $\mathbf{x} = A^{-1}\mathbf{b}$, also $\det(A)x_j = \sum_{k=1}^n \tilde{a}_{jk} b_k$, wo die $\tilde{a}_{jk}$ die Einträge der Komplementärmatrix sind. Wie im Beweis des vorigen Satzes erhalten wir $\det(A)x_j = \det(A_j)$. □

Bemerkung: Auch die Cramer'sche Regel ist für praktische Rechnung weniger effizient als die Berechnung durch den Gauß-Algorithmus. Sie erlaubt aber, die Abhängigkeit des Lösungsvektors von den Einträgen der Matrix A und dem Vektor $\mathbf{b}$ zu bestimmen, auch hier erhält man z.B. beliebig häufige Differenzierbarkeit.

Bemerkung: Bei den Beweisen dieses Abschnitts wurde nicht benutzt, dass Elemente $\neq 0$ in K invertierbar sind. Geht man die Sätze und Beweise durch, so sieht man daher, dass alle Aussagen genauso für Matrizen mit Einträgen aus einem beliebigen kommutativen Ring R mit Einselement gelten. Die Kommutativität der Multiplikation geht allerdings entscheidend ein.

Korollar 7.22. (Vandermonde-Determinante).
Seien $a_1, \ldots, a_n \in K$. *Dann ist*

$$\det \begin{pmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_n & & a_n^{n-1} \end{pmatrix} = \prod_{i<j} (a_j - a_i).$$

(Die Matrix in obiger Gleichung heißt Vandermonde-Matrix.)

Beweis. Wir beweisen das durch Induktion nach n. Die Behauptung ist klar für $n = 1$ und für $n = 2$. Sei $n > 2$ und die Behauptung bewiesen für die $(n-1) \times (n-1)$ Vandermonde-Matrix. Seien $\mathbf{s}_1, \ldots, \mathbf{s}_n$ die Spalten der Matrix. Wir führen nacheinander die Spaltentransformationen

$$\begin{array}{ccc} \mathbf{s}_n & \mapsto & \mathbf{s}_n - a_1 \mathbf{s}_{n-1} \\ \vdots & & \vdots \\ \mathbf{s}_2 & \mapsto & \mathbf{s}_2 - a_1 \mathbf{s}_1 \end{array}$$

durch (die die Determinante nicht verändern) und erhalten die Matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & a_2 - a_1 & \cdots & a_2^{n-1} - a_1 a_2^{n-2} \\ \vdots & & & \vdots \\ 1 & a_n - a_1 & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{pmatrix}.$$

Entwickeln wir diese Determinante nach der ersten Zeile, so erhalten wir

$$\det \begin{pmatrix} a_2 - a_1 & \cdots & a_2^{n-1} - a_1 a_2^{n-2} \\ \vdots & & \vdots \\ a_n - a_1 & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{pmatrix}.$$

Ziehen wir hier für $2 \leq i \leq n$ aus der $i-1$-ten Zeile den Faktor $a_i - a_1$ heraus, so erhalten wir

$$\det \begin{pmatrix} a_2 - a_1 & \cdots & a_2^{n-1} - a_1 a_2^{n-2} \\ \vdots & & \vdots \\ a_n - a_1 & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{pmatrix} = \prod_{i=2}^{n} (a_i - a_1) \det \begin{pmatrix} 1 & a_2 & \cdots & a_2^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & a_n & & a_n^{n-2} \end{pmatrix},$$

Da nach Induktionsannahme

$$\begin{pmatrix} 1 & a_2 & \cdots & a_2^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & a_n & & a_n^{n-2} \end{pmatrix} = \prod_{2 \leq i < j \leq n} (a_j - a_i)$$

gilt, folgt die Behauptung.

□

Zusammenfassung:

Die Determinante $\det(A)$, aufgefasst als Funktion der Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$ oder der Zeilen ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_n$ einer Matrix $A \in M(n \times n, K)$ ist eine alternierende n-fache Multilinearform mit $\det(E_n) = 1$; sie ist durch diese Eigenschaften eindeutig charakterisiert.
Elementare Umformungen vom Typ i) (Addition der mit λ multiplizierent $j$ten Zeile/Spalte zur i-ten Zeile/Spalte ($i \neq j$)) ändern den Wert der Determinante nicht.

Für die Determinante gilt die Rekursionsformel (Laplace'sche Entwicklungsformel)

$$\begin{aligned} \det(A) &= \textstyle\sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \\ &= \textstyle\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \end{aligned}$$

mit den Streichungsmatrizen A_{ij}.

Ferner gilt die Formel von Leibniz

$$\begin{aligned} \det(A) &= \textstyle\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \\ &= \textstyle\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}. \end{aligned}$$

Beide Formeln werden in der Regel nicht zur praktischen Rechnung benutzt (stattdessen: Gauß-Algorithmus).

Die Determinante von A ist genau dann 0, wenn die Matrix A singulär ist.

Die Determinante ist multiplikativ; ähnliche Matrizen haben die gleiche

Determinante.

Mit Hilfe der Determinante erhält man explizite Formeln für die Lösung eines linearen Gleichungssystems mit regulärer Matrix (Cramer'sche Regel) und für die Inverse einer Matrix ($A\tilde{A} = \tilde{A}A = (\det A) \cdot E_n$ mit $\tilde{a}_{ij} = (-1)^{i+j} \det(A_{ji})$).

Eine weitere wichtige Anwendung (Berechnung von Eigenwerten) wird im nächsten Abschnitt behandelt.

8. Eigenvektoren und Eigenwerte

Definition 8.1. *Sei $f : V \longrightarrow V$ lineare Abbildung von K-Vektorräumen. Ein Vektor $v \neq \mathbf{0}$ aus V heißt* Eigenvektor *von f, wenn es $\lambda \in K$ gibt mit $f(v) = \lambda v$. Die Zahl $\lambda \in K$ heißt dann der zugehörige* Eigenwert *von f.*
Ist λ Eigenwert von f, so heißt

$$V_\lambda(f) := V_\lambda := \{v \in V \mid f(v) = \lambda v\}$$

der Eigenraum *von f zum Eigenwert λ.*
Ist $A \in M_n(K)$, so heißt $\mathbf{x} \in K^n$ Eigenvektor zum Eigenwert λ von A, wenn $A\mathbf{x} = \lambda\mathbf{x}$ gilt, wenn also $\mathbf{x}$ Eigenvektor zum Eigenwert λ der zugehörigen linearen Abbildung $L_A : K^n \longrightarrow K^n$ ist; genauso ist $V_\lambda(A) := V_\lambda(L_A)$.

Beispiele:

- $A = \begin{pmatrix} 3 & 0 & -1 \\ 1 & 2 & -1 \\ -1 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{R})$ hat den Eigenvektor $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ zum Eigenwert 2.
- Die Matrix der Drehung des $\mathbb{R}^3$ um die x-Achse um den Winkel φ, $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi) & \cos(\varphi) \end{pmatrix} \in M_3(\mathbb{R})$ hat den Eigenvektor $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ zum Eigenwert 1 (Eigenvektoren zum Eigenwert 1 nennt man auch *Fixvektoren*).
  Allgemeiner hat jede Drehung des $\mathbb{R}^3$ um eine Achse die Vektoren in Richtung der Achse als Fixvektoren.
- Sei $C^\infty(\mathbb{R}) := \mathcal{D}(\mathbb{R})$ der $\mathbb{R}$-Vektorraum der unendlich oft differenzierbaren Funktionen $f : \mathbb{R} \longrightarrow \mathbb{R}$, $D : \mathcal{D}(\mathbb{R}) \longrightarrow \mathcal{D}(\mathbb{R})$ die Ableitungsabbildung $f \longmapsto f'$.
  Dann ist $\lambda \in \mathbb{R}$ Eigenwert von D mit zugehörigem Eigenvektor $f_\lambda(x) = \exp(\lambda x)$.
  Ist W der $\mathbb{C}$-Vektorraum der unendlich oft differenzierbaren Funktionen $f : \mathbb{R} \longrightarrow \mathbb{C}$, die periodisch mit Periode 2π sind, und D wie oben der Ableitungsoperator, so ist für jedes $n \in \mathbb{Z}$ die durch $g_n(x) := e^{inx} = \exp(inx)$ gegebene Funktion g_n ein Eigenvektor von D zum Eigenwert in.
  In beiden Fällen kann man zeigen, dass diese Eigenvektoren (bis auf skalare Vielfache) die einzigen Eigenvektoren von D sind.
- $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M(2 \times 2, \mathbb{R})$ hat keine Eigenwerte in $\mathbb{R}$: Wegen $A^2 = -E_2$ müsste für einen Eigenvektor $\mathbf{x}$ zum Eigenwert λ

gelten:

$$-\mathbf{x} = A^2\mathbf{x} = A(\lambda\mathbf{x}) = \lambda(A\mathbf{x}) = \lambda^2\mathbf{x},$$

also $\lambda^2 = -1$, diese Gleichung ist in $\mathbb{R}$ bekanntlich nicht lösbar. Betrachtet man A aber als Element von $M_2(\mathbb{C})$, so hat $\lambda^2 = -1$ die beiden Lösungen $i,\ -i$, und im nächsten Lemma werden wir sehen, dass diese beiden komplexen Zahlen dann tatsächlich Eigenwerte von A sind.

- Genau dann, wenn $\mathrm{Ker}(f) \neq \{\mathbf{0}\}$ gilt, ist 0 Eigenwert von f; der Kern von f ist dann der Eigenraum V_0.
- Hat die Matrix A den Eigenvektor $\mathbf{x}$ zum Eigenwert λ, so ist für $T \in GL_n(K)$ der Vektor $T^{-1}\mathbf{x}$ ein Eigenvektor der zu A ähnlichen Matrix $T^{-1}AT$ zum Eigenwert λ. Ähnliche Matrizen haben also die gleichen Eigenwerte und ihre Eigenräume zu einem festen Eigenwert sind zueinander isomorph.

Bemerkung: Es ist zweckmäßig, den Nullvektor nicht als Eigenvektor zuzulassen (siehe Definition). Dagegen kann $\lambda = 0$ durchaus als Eigenwert vorkommen (siehe oben).

Lemma 8.2. *Sei V ein endlichdimensionaler K-Vektorraum mit Basis $\mathcal{B} = (v_1, \ldots, v_n)$ und $f : V \longrightarrow V$ linear (ein Endomorphismus von V). Sei $A = M_{\mathcal{B}}(f)$ die Matrix von f bezüglich $\mathcal{B}$.*
Dann ist $v = \sum_{i=1}^n x_i v_i \neq \mathbf{0}$ genau dann Eigenvektor von f zum Eigenwert $\lambda \in K$, wenn $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ Eigenvektor von A zum Eigenwert λ ist. Zu beidem äquivalent ist:
$v \in \mathrm{Ker}(\lambda\ \mathrm{Id}_V - f) \setminus \{\mathbf{0}\}$.
Ferner sind folgende Aussagen äquivalent:

a) *$\lambda \in K$ ist Eigenwert von f,*
b) *$\lambda \in K$ ist Eigenwert von A,*
c) $\det(\lambda E_n - A) = 0$,
d) $\mathrm{Ker}(\lambda\ \mathrm{Id}_V - f) \neq \{\mathbf{0}\}$.

Ist λ Eigenwert von A, so erhält man sämtliche Eigenvektoren von L_A zum Eigenwert λ durch Lösen des linearen Gleichungssystems

$$(\lambda E_n - A)\mathbf{x} = \mathbf{0}$$

Beweis. Alle Aussagen sind klar, wenn man im Fall $V = K^n,\ f = L_A$ (mit der Standardbasis $\mathcal{B}$) ist. Wegen der Korrespondenz zwischen Matrizen und linearen Abbildungen folgt aber hieraus sofort die Aussage für beliebiges V, f. □

Beispiel: Sei wie oben $A = \begin{pmatrix} 3 & 0 & -1 \\ 1 & 2 & -1 \\ -1 & 1 & 1 \end{pmatrix} \in M(3 \times 3, \mathbb{R})$,
$f = L_A : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$.

Die Matrix $A - \lambda E_3$ wird wie folgt umgeformt:

$$\begin{pmatrix} 3-\lambda & 0 & -1 \\ 1 & 2-\lambda & -1 \\ -1 & 1 & 1-\lambda \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2-\lambda & -1 \\ 0 & (\lambda-3)(2-\lambda) & 2-\lambda \\ 0 & 3-\lambda & -\lambda \end{pmatrix}.$$

Ist $\lambda = 2$, so vertauscht man zweite und dritte Zeile und erhält

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix};$$

man findet die Lösung $x_3 = 1$, $x_2 = 2$, $x_1 = 1$, also den Eigenvektor $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} =: v_2$ zum Eigenwert 2.

Ist $\lambda \neq 2$, so dividiere man die zweite Zeile durch $2 - \lambda$ und forme weiter um:

$$\begin{pmatrix} 1 & 2-\lambda & -1 \\ 0 & \lambda-3 & 1 \\ 0 & 3-\lambda & -\lambda \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2-\lambda & -1 \\ 0 & \lambda-3 & 1 \\ 0 & 0 & 1-\lambda \end{pmatrix}.$$

Man sieht, dass diese Matrix für $\lambda = 1$ und für $\lambda = 3$ singulär wird, diese sind also ebenfalls Eigenwerte.

Für $\lambda = 1$ findet man den Eigenvektor $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} =: v_1$ zum Eigenwert 1,

für $\lambda = 3$ den Eigenvektor $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} =: v_3$ zum Eigenwert 3.

Die Vektoren v_1, v_2, v_3 bilden eine Basis des $\mathbb{R}^3$, bezüglich der L_A die Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

in Diagonalgestalt hat.

L_A ist also die lineare Abbildung, die man erhält, indem man in v_1-Richtung keine Änderung vornimmt, in v_2-Richtung um den Faktor 2 und in v_3-Richtung um den Faktor 3 streckt.

Im Weiteren wollen wir $\det(\lambda E_n - A)$ als Polynom in der Variablen λ betrachten. Dafür brauchen wir noch ein paar Vorbereitungen.

Definition und Satz 8.3. *Sei K ein Körper. Auf dem Vektorraum $K[X]$ der Polynome in einer Variablen X über K wird die Gradfunktion* deg *gegeben, indem man für $f = \sum_{i=0}^{m} a_i X^i$ mit $a_m \neq 0$ den Grad von f als $\deg(f) = m$ definiert, mit $\deg 0 = -\infty$.*

Ferner wird auf $K[X]$ eine Multiplikation definiert durch

$$(\sum_{i=0}^{m} a_i X^i) \cdot (\sum_{j=0}^{n} b_j X^j) = \sum_{k=0}^{m+n} (\sum_{i+j=k} a_i b_j) X^k,$$

dabei wird $a_i = 0$ gesetzt, wenn $i > m$ ist, analog für b_j.

Dann gilt:

a) *Die Multiplikation in $K[X]$ ist assoziativ und kommutativ, mit $1 = 1X^0$ als neutralem Element, und für Addition und Multiplikation in $K[X]$ gelten die Distributivgesetze. Der Vektorraum $K[X]$ wird also durch die Multiplikation zu einem kommutativen Ring mit Einselement.*
b) *Sind $f \neq 0 \neq g \in K[X]$, so ist $f \cdot g \neq 0$. Man sagt der Ring $K[X]$ sei* nullteilerfrei.
c) *Das Polynom $X^i = 1X^i$ ist die i-te Potenz des Polynoms $X = 1X^1$.*
d) *Sind $f \neq 0 \neq g \in K[X]$, so ist $\deg(fg) = \deg(f) + \deg(g)$ und $\deg(f+g) \leq \max(\deg(f), \deg(g))$.*

*Sei ferner $R \supseteq K$ ein kommutativer Ring, für den 1_K neutrales Element der Multiplikation ist (eine K-*Algebra*) , für $\lambda \in R$ werde durch $f = \sum_{i=0}^{n} a_i X^i \mapsto i_\lambda(f) := f(\lambda) := \sum_{i=0}^{n} a_i \lambda^i \in R$ die Einsetzungsabbildung $i_\lambda : K[X] \to R$ definiert.*
Dann gilt: Für $f, g \in K[X]$ und $\lambda \in R$ ist

$$\begin{aligned} (f+g)(\lambda) &= f(\lambda) + g(\lambda) \\ (f \cdot g)(\lambda) &= f(\lambda) \cdot g(\lambda), \end{aligned}$$

die Abbildung i_λ ist also ein Homomorphismus von Ringen.

Beweis. Alle Aussagen rechnet man direkt nach. Für die Assoziativität betrachte man etwa f, g wie oben und $h = \sum_{\ell=0}^{p} c_\ell X^\ell$, dann erhält man nach Ausmultiplizieren von $(f \cdot g) \cdot h$ und $f \cdot (g \cdot h)$ die Assoziativität durch Vergleich von $\sum_{i+j+\ell=k} (a_i b_j) c_\ell$ mit $\sum_{i+j+\ell=k} a_i (b_j c_\ell)$, die anderen Aussagen sind genauso leicht zu zeigen. □

Bemerkung. Für die durch $\lambda \mapsto f(\lambda)$ für $f \in K[X]$ definierte Polynomfunktion $K \to K$ schreiben wir auch $\tilde{f}$. Die Tilde über dem f bei der Polynomfunktion wird oft (wenn kein Irrtum entstehen kann) fortgelassen, das wird auch durch die Schreibweise $f(\lambda) = \tilde{f}(\lambda)$ nahegelegt. Dass man damit vorsichtig sein muss, sieht man am folgenden Beispiel:
Ist $K = \mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ der Körper mit zwei Elementen, so ist $\lambda^2 = \lambda$ für alle $\lambda \in K$. Die auf K definierte Polynomfunktion $\lambda \mapsto \lambda^2 - \lambda$ zum Polynom $X^2 - X$ ist also die gleiche wie die zum Nullpolynom $0 = 0 \cdot X^0$, nämlich die Nullfunktion. Insbesondere kann man für die Polynomfunktionen nicht ohne weiteres vom Grad der Funktion sprechen. Beim Übergang zu einem geeigneten größeren Körper, in dem

es Elemente x mit $x^2 \neq x$ gibt, werden die Polynomfunktionen dann allerdings verschieden.
Den abstrakten Polynomring führt man ein, weil es in dieser Situation ausnahmsweise sinnvoll ist, verschiedene Abbildungsvorschriften auch dann zu unterscheiden, wenn die durch sie definierten Abbildungen auf den Elementen von K übereinstimmen; man besitzt dann genug Flexibilität, um bei Bedarf zu geeigneten größeren Körpern oder Ringen überzugehen.

Satz 8.4. (Euklidischer Algorithmus, Division mit Rest) *Sei K ein Körper, $f, g \in K[X]$ mit $g \neq 0$. Dann gibt es $q, r \in K[X]$, so dass*

$$f = qg + r \text{ mit } r = 0 \text{ oder } \deg(r) < \deg(g)$$

gilt. r und q sind eindeutig bestimmt.

Beweis. Wir beweisen diese Aussage durch vollständige Induktion nach $\deg(f)$, beginnend bei $\deg(f) = 0$. Der Induktionsanfang $\deg(f) = 0$ ist trivial. Wir schreiben $f = \sum_{i=0}^{m} a_i X^i, g = \sum_{i=0}^{n} b_i X^i$ mit $a_m \neq 0, b_n \neq 0, m \geq 1$ und nehmen an, die Aussage sei für $\deg(f) < m$ bereits bewiesen.
Ist $\deg(f) < \deg(g)$, so ist die Aussage (mit $q = 0, r = f$) trivial, wir können also $n \leq m$ annehmen. Dann ist der Grad von

$$\begin{aligned} f_1 &:= f - (\frac{a_m}{b_n} X^{m-n})g \\ &= (a_m X^m - (\frac{a_m}{b_n} X^{m-n}) b_n X^n) + \sum_{i=0}^{m-1} c_i X^i \\ &= \sum_{i=0}^{m-1} c_i X^i \end{aligned}$$

(mit gewissen $c_i \in K$, die hier nicht weiter interessieren) offenbar kleiner als m, wir können also nach Induktionsannahme

$$f_1 = q_1 g + r \text{ mit } r = 0 \text{ oder } \deg(r) < \deg(g)$$

schreiben und erhalten

$$f = (q_1 + \frac{a_m}{b_n} X^{m-n})g + r,$$

was mit $q = q_1 + \frac{a_m}{b_n} X^{m-n}$ die gewünschte Zerlegung $f = qg + r$ für f liefert. Die Eindeutigkeit überlege man sich als Übung. □

Beispiel: Durch den üblichen Prozess der Polynomdivision erhält man etwa:

$$(X^4 - 1) = (X^2 + 2X + 1)(X^2 - 2X + 3) + (-4X - 4).$$

Bemerkung. a) Im Beweis benutzt man Division durch den Leitkoeffizienten $b_n \neq 0$ von $g = \sum_{i=0}^{n} b_i X^i$; das Verfahren der Division mit Rest lässt sich daher nicht ohne weiteres auf den Polynomring $R[X]$ über einem Ring R übertragen, da es in einem Ring, der kein Körper ist, vorkommen kann, dass ein $b_n \neq 0$ nicht invertierbar ist.

b) Betrachtet man die Gleichung $f = qg + r$, so sieht man, dass alle gemeinsamen Teiler von f und g auch $r = f - qg$ teilen und damit auch gemeinsame Teiler von g und r sind. Da der Schluss sich umkehren lässt, sind genauer die gemeinsamen Teiler von f und g genau die gemeinsamen Teiler von g und r.

Definition und Korollar 8.5. *Sei K ein Körper.*

a) *Sei $f \in K[X]$, $f \neq 0$, $a \in K$ mit $f(a) = 0$. Dann gibt es ein eindeutig bestimmtes $q \in K[X]$ mit $f = (X - a)q$.*

b) *Sind $\beta_1, \ldots, \beta_r$ verschiedene Nullstellen von $0 \neq f \in K[X]$, so gibt es eindeutig bestimmte $e_i \in \mathbb{N} \setminus \{0\}$, $g \in K[X]$ mit*

$$f = \prod_{i=1}^{r} (X - \beta_i)^{e_i} g \text{ und } g(\beta_i) \neq 0 \text{ für } 1 \leq i \leq r.$$

Der Exponent e_i in dieser Darstellung heißt die Vielfachheit *der Nullstelle β_i des Polynoms f, ist $e_i = 1$, so spricht man von einer* einfachen Nullstelle*, sonst von einer* mehrfachen*.*

c) *Seien $f, g \in K[X]$ mit $n > \max(\deg(f), \deg(g))$, seien $u_1, \ldots, u_n \in K$ paarweise verschieden mit $f(u_i) = g(u_i)$ für $1 \leq i \leq n$.*
Dann ist $f = g$.
Insbesondere gilt: Hat K unendlich viele Elemente, so folgt aus $f(\lambda) = g(\lambda)$ für alle $\lambda \in K$, dass $f = g$ gilt.

d) *Ist $K = \mathbb{C}$ der Körper der komplexen Zahlen (siehe Satz 5.16) und $f \in \mathbb{C}[X]$ mit $\deg(f) = n > 0$ ein nicht konstantes normiertes Polynom (d.h., $f = X^n + \sum_{i=0}^{n-1} c_i X^i$), so gibt es $\beta_1, \ldots, \beta_n \in \mathbb{C}$ (nicht notwendig verschieden) mit $f = \prod_{i=1}^{n} (X - \beta_i)$.*

Beweis. a) Wir teilen f mit Rest durch $X - a$. Wäre der Rest hierbei nicht 0, so hätte er wegen $\deg(X - a) = 1$ Grad 0, wäre also gleich einer Konstanten $c \in K$. Setzen wir in die Polynomgleichung $f = (X - a)q + c$ den Wert $a \in K$ ein, so erhalten wir

$$0 = f(a) = (a - a)q(a) + c,$$

also $c = 0$.

b) Zunächst ist klar, dass man eine Darstellung

$$f = \prod_{i=1}^{r} (X - \beta_i)^{e_i} g \text{ und } g(\beta_i) \neq 0 \text{ für } 1 \leq i \leq r.$$

erhält, indem man a) so oft iteriert, bis der verbleibende Faktor g in keinem der β_i verschwindet.

Hat man zwei derartige Darstellungen

$$f = \prod_{i=1}^{r}(X-\beta_i)^{e_i} g = \prod_{i=1}^{r}(X-\beta_i)^{e'_i} g'$$

und ist etwa $e_1 \geq e'_1$, so hat man nach Ausklammern des Faktors $(X-\beta_1)^{e'_1}$ die Gleichung

$$(X-\beta_1)^{e'_1}\big((X-\beta_1)^{e_1-e'_1}\prod_{i=2}^{r}(X-\beta_i)^{e_i} g - \prod_{i=2}^{r}(X-\beta_i)^{e'_i} g'\big) = 0,$$

und da $K[X]$ nullteilerfrei und $(X-\beta_1)^{e'_1} \neq 0$ ist, ist hierin der in Klammern stehende zweite Faktor 0, und man erhält

$$(X-\beta_1)^{e_1-e'_1}\prod_{i=2}^{r}(X-\beta_i)^{e_i} g = \prod_{i=2}^{r}(X-\beta_i)^{e'_i} g'.$$

Einsetzen von β_1 in diese Gleichung liefert dann $e_1 - e'_1 = 0$, da sonst die linke Seite 0 ergäbe und die rechte nicht. Das iteriert man für die anderen Faktoren $(X-\beta_i)$ und erhält am Ende $g = g'$.
c) In b) sehen wir, dass $f = \prod_{i=1}^{r}(X-\beta_i)^{e_i} g$ Grad $\deg(g) + \sum_{i=1}^{r} e_i$ hat, insbesondere muss $r \leq n$ für die Anzahl r der verschiedenen Nullstellen eines Polynoms $f \neq 0$ vom Grad n gelten. Anders gesagt: Nimmt ein Polynom f in n verschiedenen Stellen $a_1, \ldots, a_n$ den Wert 0 an, so muss $\deg(f) \geq n$ oder $f = 0$ gelten.
Da in der Situation von c) $\deg(f-g) < n$ gilt und $f-g$ in den n verschiedenen Stellen $a_1, \ldots, a_n$ den Wert 0 annimmt, ist $f - g = 0$, also $f = g$.
d) Folgt durch wiederholte Anwendung von a) aus der Tatsache, dass in $\mathbb{C}$ jedes nicht konstante Polynom wenigstens eine Nullstelle hat. □

Bemerkung. Sind $f \in K[X]$ und $a, c \in K$ mit $f(a) = c$ und hat $f - c$ in a eine e-fache Nullstelle, so sagt man auch, f nehme in a den Wert c mit der Vielfachheit e an.

Wir erinnern daran, dass die zunächst für Matrizen mit Einträgen aus einem Körper K entwickelte Determinantentheorie aus Abschnitt 7 auch über einem beliebigen kommutativen Ring gilt. Insbesondere liefert für einen beliebigen kommutativen Ring R die Leibniz'sche Formel

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)}$$

eine Definition der Determinante einer Matrix $A \in M_n(R)$, die folgende Eigenschaften hat:

- $\det(A)$ ist sowohl als Funktion der Zeilen von A als auch als Funktion der Spalten von A eine alternierende n-fache Multilinearform mit Werten in R
- A ist in $M_n(R)$ genau dann invertierbar, wenn $\det(A)$ im Ring R invertierbar ist.

- Für $A, B \in M_n(R)$ gilt $\det(AB) = \det(A)\det(B)$.
- Für die durch $\tilde{a}_{ij} = (-1)^{i+j}\det(A_{ji})$ definierte Komplementärmatrix $\tilde{A} = (\tilde{a}_{ij}) \in M_n(R)$ der Matrix $A \in M_n(R)$ gilt
$$\tilde{A}A = A\tilde{A} = \det(A)E_n.$$

Definition 8.6. *Das charakteristische Polynom der Matrix $A \in M_n(K)$ ist gegeben als*
$$\chi_A := \det(XE_n - A) \in K[X]$$
(dabei ist
$$XE_n - A = \begin{pmatrix} X - a_{11}X^0 & \dots & -a_{1n}X^0 \\ \vdots & & \vdots \\ -a_{n1}X^0 & \dots & X - a_{nn}X^0 \end{pmatrix} \in M_n(K[X])$$
als Matrix mit Koeffizienten im Polynomring $K[X]$ aufzufassen).

Korollar 8.7. *$\lambda \in K$ ist genau dann Eigenwert der Matrix $A \in M_n(K)$, wenn $\chi_A(\lambda) = 0$ gilt.*
Insbesondere kann A nicht mehr als n verschiedene Eigenwerte haben.

Beweis. Die erste Behauptung ist klar nach Lemma 8.2. Da ein von 0 verschiedenes Polynom vom Grad n höchstens n verschiedene Nullstellen hat und $\chi_A \neq 0$ ist, folgt auch die zweite Behauptung. □

Lemma 8.8. *Für $A \in M_n(K)$ ist*
$$\chi_A = \sum_{i=0}^{n} a_i X^i$$
mit $a_n = 1$, man sagt: χ_A ist ein normiertes Polynom vom Grad n.
Ferner gilt: $-a_{n-1} = \mathrm{tr}(A)$ ist die Spur
$$\mathrm{Spur}(A) = \mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$
der Matrix A, und man hat
$$a_0 = (-1)^n \cdot \det(A).$$

Beweis. Die Behauptungen folgen direkt aus der Leibniz'schen Formel für die Determinante. □

Definition und Lemma 8.9. *Sind A und A' aus $M_n(K)$ zueinander ähnliche (konjugierte) Matrizen, so ist $\chi_A = \chi_{A'}$.*
Ist V ein endlichdimensionaler K-Vektorraum, $f \in \mathrm{End}(V)$, A die Matrix von f bezüglich einer (beliebigen) Basis $\mathcal{B}$, so ist
$$\chi_f := \chi_A$$

Beweis. Der Lemma-Anteil hiervon ist wegen $T^{-1}(X \cdot E_n - A)T = X \cdot E_n - T^{-1}AT$ und der Multiplikativität der Determinante klar. □

Definition und Lemma 8.10. *Sei V ein endlichdimensionaler K-Vektorraum. $f \in \mathrm{End}(V)$ heißt* diagonalisierbar*, wenn eine der folgenden äquivalenten Aussagen gilt:*

a) *V hat eine Basis aus Eigenvektoren von f.*
b) *Bezüglich einer geeigneten Basis von V hat die Matrix von f Diagonalgestalt.*
c) *Ist $\mathcal{B}$ Basis von V und $A = M_{\mathcal{B}}(f)$, so gibt es $T \in \mathrm{GL}_n(K)$, so dass $T^{-1}AT$ eine Diagonalmatrix ist.*

Ist die Matrix von f bzüglich einer geeigneten Basis von V eine Dreiecksmatrix, so heißt f trigonalisierbar.

Eine Matrix $A \in M_n(K)$ heißt diagonalisierbar bzw. trigonalisierbar, wenn der Endomorphismus L_A von K^n die jeweilige Eigenschaft hat. Äquivalent dazu ist, dass es $T \in GL_n(K)$ gibt, so dass $T^{-1}AT$ Diagonalgestalt bzw. Dreiecksgestalt hat.

Beweis. Die Äquivalenz der Bedingungen ist klar. □

Beispiel:

a) Die im vorigen Beispiel diskutierte Matrix $A \in M_3(\mathbb{R})$ ist diagonalisierbar, mit $T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix}$ ist $T^{-1}AT = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.
b) Eine Drehung ($\neq$ Id) in $\mathbb{R}^2$ (insbesondere die oben diskutierte Drehung um 90° mit Matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$) hat keinen Eigenvektor in $\mathbb{R}^2$, ist also nicht diagonalisierbar.
Da für eine obere Dreiecksmatrix der erste Standardbasisvektor $\mathbf{e}_1$ und für eine untere Dreiecksmatrix $\mathbf{e}_n$ ein Eigenvektor ist, hat jede trigonalisierbare Matrix Eigenvektoren. Eine nichttriviale Drehung im $\mathbb{R}^2$ ist also auch nicht trigonalisierbar.
Allerdings wird etwa die Matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, als Matrix über $\mathbb{C}$ betrachtet, diagonalisierbar: Man findet die beiden Eigenwerte $i, -i$ mit zugehörigen Eigenvektoren $\begin{pmatrix} 1 \\ i \end{pmatrix}, \begin{pmatrix} 1 \\ -i \end{pmatrix} \in \mathbb{C}^2$ (dabei ist wie üblich i die imaginäre Einheit mit $i^2 = -1$).
c) Die Matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ hat als einzigen Eigenwert 1 und alle Eigenvektoren sind Vielfache von $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (unabhängig davon, über welchem Körper man die Matrix betrachtet).
Die Matrix ist also weder in $M(2 \times 2, \mathbb{R})$ noch in $M(2 \times 2, \mathbb{C})$ diagonalisierbar. Da sie Dreiecksgestalt hat, ist sie natürlich trigonalisierbar.

Lemma 8.11. *Sei $A \in M_n(K)$ und $T \in GL_n(K)$. Genau dann hat $T^{-1}AT$ Diagonalgestalt, wenn die Spalten von T Eigenvektoren von A sind.*
Die Matrix A ist also genau dann diagonalisierbar, wenn es in $GL_n(K)$ eine Matrix gibt, deren Spalten Eigenvektoren von A sind.

Beweis. Ist $D = T^{-1}AT$, so ist $D\mathbf{x} = \lambda\mathbf{x}$ äquivalent zu $AT\mathbf{x} = \lambda T\mathbf{x}$, also ist $\mathbf{x}$ genau dann Eigenvektor von D, wenn $T\mathbf{x}$ Eigenvektor von A zum gleichen Eigenwert ist. Da andererseits D genau dann Diagonalgestalt hat, wenn die Standard-Basisvektoren $\mathbf{e}_j$ Eigenvektoren von D sind, folgt die Behauptung. □

Definition 8.12. *Die Matrix $A \in M_n(\mathbb{R})$ heißt orthogonal, wenn $A^{-1} = {}^tA$ gilt.*

Beispiel:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

ist orthogonal.

Beispiel: Das *Standardskalarprodukt* auf $\mathbb{R}^n$ ist definiert durch

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} x_i y_i = {}^t\mathbf{y}\ \mathbf{x}.$$

Vektoren $v_1, \ldots, v_n \in \mathbb{R}^n$ bilden eine Orthonormalbasis, wenn $\langle v_i, v_j \rangle = \delta_{ij}$ gilt (man überzeugt sich leicht, dass sie dann linear unabhängig sind, also in der Tat eine Basis des $\mathbb{R}^n$ bilden).
Man sieht: $A \in GL_n(\mathbb{R})$ ist genau dann eine orthogonale Matrix, wenn die Spaltenvektoren von A eine Orthonormalbasis bilden.

Satz 8.13. *Ist $A \in M_n(\mathbb{R})$ symmetrisch ($A = {}^tA$), so ist A diagonalisierbar. Genauer gilt: Es gibt eine orthogonale Matrix $S \in \mathrm{GL}_n(\mathbb{R})$, so dass*

$${}^tSAS = S^{-1}AS$$

Diagonalgestalt hat.
Der Raum $\mathbb{R}^n$ hat also für symmetrisches $A \in M_n(\mathbb{R})$ eine Orthonormalbasis aus Eigenvektoren von A.

Beweis. Das werden wir erst im nächsten Abschnitt beweisen. □

Lemma 8.14. *Sei V ein K-Vektorraum, $f \in \mathrm{End}(V)$. Die Vektoren $v_1, \ldots, v_r$ seien Eigenvektoren von f zu den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$. Dann sind $v_1, \ldots, v_r$ linear unabhängig.*

Beweis. Wir beweisen das durch vollständige Induktion nach der Anzahl r der Vektoren.
Für $r = 1$ ist die Behauptung trivial (Induktionsanfang), wir betrachten also $r > 1$ und nehmen an, die Behauptung sei für $r' < r$ Eigenvektoren bewiesen (Induktionsannahme).

Ist dann

$$\sum_{i=1}^{r} a_i v_i = \mathbf{0}$$

mit Skalaren $a_i \in K$, so wenden wir f auf diese Gleichung an und erhalten

$$\mathbf{0} = \sum_{i=1}^{r} a_i f(v_i) = \sum_{i=1}^{r} \lambda_i a_i v_i.$$

Wir multiplizieren die erste dieser beiden Gleichungen mit λ_1 und haben jetzt die beiden folgenden Gleichungen:

$$\begin{aligned} \lambda_1 a_1 v_1 + \lambda_1 a_2 v_2 + \cdots + \lambda_1 a_r v_r &= \mathbf{0} \\ \lambda_1 a_1 v_1 + \lambda_2 a_2 v_2 + \cdots + \lambda_r a_r v_r &= \mathbf{0}. \end{aligned}$$

Wir subtrahieren die erste Gleichung von der zweiten und haben

$$(\lambda_2 - \lambda_1) a_2 v_2 + \cdots + (\lambda_r - \lambda_1) a_r v_r = \mathbf{0}.$$

Da nach Induktionsannahme die Vektoren $v_2, \ldots, v_r$ linear unabhängig sind, ist

$$(\lambda_2 - \lambda_1) a_2 = \ldots = (\lambda_r - \lambda_1) a_r = 0,$$

und da alle $\lambda_j - \lambda_1 \neq 0$ sind, folgt

$$a_2 = \ldots = a_r = 0,$$

wegen $\sum_{i=1}^{r} a_i v_i = \mathbf{0}$ also auch $a_1 v_1 = \mathbf{0}$ und damit wegen $v_1 \neq \mathbf{0}$ auch $a_1 = 0$.
Wir haben also gezeigt, dass aus $\sum_{i=1}^{r} a_i v_i = \mathbf{0}$ folgt, dass $a_1 = \ldots = a_r = 0$ gilt, die Vektoren $v_1, \ldots, v_r$ sind also wie behauptet linear unabhängig. □

Satz 8.15. *Sei V ein K-Vektorraum, $f \in \mathrm{End}(V)$, $\lambda_1, \ldots, \lambda_r \in K$ seien paarweise verschiedene Eigenwerte von f, $U_i := V_{\lambda_i}$ die jeweiligen Eigenräume ($1 \leq i \leq n$).*
Dann bilden die U_i eine direkte Summe*, jeder Vektor u aus $U_1 + \cdots + U_r = \{u_1 + \ldots + u_r \mid u_i \in U_i$ für $1 \leq i \leq r\}$ lässt sich also nur auf eine Weise als*

$$u = u_1 + \cdots + u_r \text{ mit } u_i \in U_i \text{ für } 1 \leq i \leq n$$

darstellen.

Beweis. Ist $u_1 + \cdots + u_r = u_1' + \cdots + u_r'$ mit $u_i, u_i' \in U_i$, so ist $u_1 - u_1' + \cdots + u_r - u_r' = \mathbf{0}$, und da Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind, muss $u_1 = u_1', \ldots, u_r = u_r'$ gelten. □

Satz 8.16. *Sei V ein K-Vektorraum der Dimension n, $f \in \mathrm{End}(V)$ so, dass*

$$\chi_f = \prod_{i=1}^{n} (X - \beta_i)$$

mit paarweise verschiedenen $\beta_1, \ldots, \beta_n \in K$ gilt (das charakteristische Polynom χ_f von f zerfällt über K vollständig in verschiedene Linearfaktoren*).*
Dann ist f diagonalisierbar.

Beweis. Ist

$$\chi_f = \prod_{i=1}^{n} (X - \beta_i),$$

so sind die β_i Eigenwerte von f. Sind $v_1, \ldots, v_n$ Eigenvektoren zu diesen Eigenwerten, so sind diese Vektoren nach Lemma 8.14 linear unabhängig, da die β_i als paarweise verschieden vorausgsetzt wurden. Sie bilden also wegen $\dim(V) = n$ eine Basis von V, die aus Eigenvektoren von f besteht, d.h., f ist diagonalisierbar. □

Falls χ_f zwar in Linearfaktoren zerfällt, diese aber nicht paarweise verschieden sind (wenn es also Linearfaktoren gibt, die zu einer höheren Potenz in χ_f aufgehen), so wird es schwieriger zu entscheiden, ob f diagonalisierbar ist. Dies sieht man zum Beispiel durch Betrachten der Matrizen $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ und $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$: Während die zweite offenbar diagonal (und damit diagonalisierbar) ist, ist die erste nicht diagonalisierbar, beide Matrizen haben aber das gleiche charakteristische Polynom $(X-1)^2$. Immerhin können wir noch den folgenden Satz zeigen:

Satz 8.17. *Sei V ein K-Vektorraum der Dimension n, $f \in \mathrm{End}(V)$ so, dass*

$$\chi_f = \prod_{i=1}^{n} (X - \beta_i)$$

mit (nicht notwendig verschiedenen) $\beta_1, \ldots, \beta_n \in K$ gilt (χ_f zerfällt über K vollständig in Linearfaktoren*).*
Dann ist f trigonalisierbar.
Umgekehrt gilt: Ist f trigonalisierbar, so zerfällt das charakteristische Polynom χ_f von f als

$$\chi_f = \prod_{i=1}^{n} (X - \beta_i)$$

mit $\beta_1, \ldots, \beta_n \in K$ (die aber nicht notwendig paarweise verschieden sind).
Insbesondere gilt also: Über $\mathbb{C}$ ist jeder Endomorphismus trigonalisierbar.

Beweis. Ist f trigonalisierbar, so hat f bezüglich einer geeigneten Basis $v_1, \ldots, v_n$ von V die (o. E. obere) Dreiecksmatrix

$$\begin{pmatrix} \beta_1 & * & * \\ & \ddots & * \\ & & \beta_n \end{pmatrix},$$

für das charakteristische Polynom χ_f von f gilt also

$$\chi_f = \prod_{i=1}^{n}(X - \beta_i).$$

Die andere Richtung der Behauptung zeigen wir durch Induktion nach $n = \dim(V)$:
Der Induktionsanfang $n = 1$ ist trivial. Sei also $n > 1$ und die Behauptung für Räume W mit $\dim(W) < n$ gezeigt (Induktionsannahme).
Da das charakteristische Polynom von f wie angegeben zerfällt, ist jedenfalls β_1 ein Eigenwert, es gibt also einen Eigenvektor v_1 zum Eigenwert β_1. Wir ergänzen ihn zu einer Basis $\mathcal{B}'$ von V. Bezüglich dieser hat f die Matrix

$$A' = \begin{pmatrix} \beta_1 & \dots \\ 0 & \\ \vdots & B \\ 0 & \end{pmatrix}$$

mit einer $(n-1) \times (n-1)$-Matrix B. Wegen der Formel für die Determinante einer Blockmatrix aus Korollar 7.19 ist

$$\prod_{i=1}^{n}(X - \beta_i) = \chi_f = (X - \beta_1)\chi_B.$$

Wie im Beweis von Definition und Korollar 8.5 können wir hier den Faktor $(X - \beta_1)$ kürzen und sehen, dass

$$\chi_B = \prod_{i=2}^{n}(X - \beta_i)$$

gilt.
Nach Induktionsannahme ist B trigonalisierbar, es gibt also $T' \in GL_{n-1}(K)$, so dass $(T')^{-1}BT$ obere Dreiecksgestalt hat. Setzt man

$$T = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & T' & \\ 0 & & & \end{pmatrix} \in GL_n(K),$$

so hat $T^{-1}A'T$ Dreiecksgestalt, A' und damit f ist also trigonalisierbar. □

Korollar 8.18. *Sei $A \in M_n(K)$ so, dass χ_A über K vollständig in Linearfaktoren zerfällt. Dann ist A trigonalisierbar, d. h., es gibt $T \in \mathrm{GL}_n(K)$, so dass $T^{-1}AT$ Dreiecksgestalt hat (mit den Eigenwerten $\beta_1, \dots, \beta_n$ als Diagonalelementen).*
Sind die Linearfaktoren paarweise verschieden, so ist A sogar diagonalisierbar.

Definition 8.19. *Sei $A \in M_n(K)$ mit $\chi_A = (X-\beta)^e g$ mit einem Polynom $g \in K[X]$ mit $g(\beta) \neq 0$. Dann heißt e die* algebraische Vielfachheit *des Eigenwerts β von A.*
Ist s die Dimension des Eigenraums zum Eigenwert β von A, so heißt s die geometrische Vielfachheit *des Eigenwerts β.*
Entsprechend sind algebraische und geometrische Vielfachheit der Eigenwerte eines Endomorphismus f eines endlichdimensionalen K-Vektorraums definiert.

Bemerkung. Nach Definition und Lemma 8.10 ist die Matrix A bzw. der Endomorphismus f von V (mit Matrix A bzgl. einer geeigneten Basis) genau dann diagonalisierbar, wenn V eine Basis aus Eigenvektoren hat.
Da die Summe der geometrischen Vielfachheiten der Eigenwerte von A die Maximalzahl linear unabhängiger Eigenvektoren ist, ist die Matrix $A \in M_n(K)$ bzw. der Endomorphismus f also genau dann diagonalisierbar, wenn die Summe der geometrischen Vielfachheiten der Eigenwerte gleich der Dimension n des zu Grunde liegenden Vektorraums ist.

Lemma 8.20. *Ist $A \in M_n(K)$ (bzw. $f \in \mathrm{End}(V)$, V endlichdimensionaler K-Vektorraum), β ein Eigenwert von A (von f) mit algebraischer Vielfachheit e und geometrischer Vielfachheit s, so ist $e \geq s$.*

Beweis. Ist V_β der Eigenraum von f zum Eigenwert β und $(v_1, \ldots, v_n)$ eine Basis von V, die durch Ergänzung einer Basis von V_β entsteht, so hat f bezüglich dieser Basis eine Blockmatrix

$$\begin{pmatrix} \beta E_s & B \\ 0 & C \end{pmatrix}$$

und daher charakteristisches Polynom $\chi_f = (X-\beta)^s \chi_C$. Also ist die algebraische Vielfachheit e von β wenigstens so groß wie s. □

Satz 8.21. *Sei V ein n-dimensionaler K-Vektorraum, $f \in \mathrm{End}(V)$, $\chi_f = \prod_{i=1}^r (X-\beta_i)^{e_i}$ mit paarweise verschiedenen β_i, $e_i \in \mathbb{N} \setminus \{0\}$, $U_i = V_{\beta_i}$ der Eigenraum zum Eigenwert β_i. Dann gilt:*
f ist genau dann diagonalisierbar, wenn $\dim(U_i) = e_i$ für $1 \leq i \leq r$ gilt (wenn also die algebraischen Vielfachheiten gleich den geometrischen Vielfachheiten sind).

Beweis. Ist f diagonalisierbar, so liest man die Gleichheit von algebraischen und geometrischen Vielfachheiten direkt an der Matrix von f bezüglich einer Basis ab, die aus Eigenvektoren besteht.
Sind umgekehrt die algebraischen Vielfachheiten e_i der Eigenwerte β_i gleich ihren geometrischen Vielfachheiten s_i, so ist $\sum_{i=1}^r s_i = \dim(V)$. Da die Eigenräume $U_i = V_{\beta_i}$ zu den Eigenwerten β_i nach Satz 8.15 eine

direkte Summe bilden, ist V die direkte Summe der Eigenräume von f und f daher diagonalisierbar.

$\square$

Beispiel: Für die Matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in M(2 \times 2, K)$ (K beliebiger Körper) ist 1 der einzige Eigenwert; die algebraische Vielfachheit ist 2, die geometrische Vielfachheit ist 1. Die Matrix ist trigonalisierbar, aber nicht diagonalisierbar.

Bemerkung. Im Grunde genommen sagt dieser Satz, dass man die Frage, ob die Matrix A diagonalisierbar ist oder nicht, dadurch entscheidet, dass man versucht, A zu diagonalisieren:
Man bestimmt zunächst die Eigenwerte, indem man das charakteristische Polynom und dessen Nullstellen berechnet und stellt dann für jeden Eigenwert β durch Bestimmen der Dimension des Lösungsraumes des linearen Gleichungssystems $(A - \beta E_n)\mathbf{x} = \mathbf{0}$ fest, ob er die maximal mögliche geometrische Vielfachheit hat.
Ist dies für einen Eigenwert nicht der Fall, so ist die Matrix nicht diagonalisierbar, andernfalls ist es praktisch kein zusätzlicher Aufwand, für die Lösungsräume der Gleichungssysteme $(A - \beta E_n)\mathbf{x} = \mathbf{0}$ neben der Dimensionsbestimmung auch gleich noch Basen zu bestimmen. Diese sind dann Basen der jeweiligen Eigenräume und ergeben zusammengenommen eine Basis von $V = K^n$, die aus Eigenvektoren von A besteht, bezüglich der die Matrix des Endomorphismus L_A also eine Diagonalmatrix D ist. Ist $S \in M_n(K)$ die Matrix, deren Spalten diese Eigenvektoren sind, so ist $S^{-1}AS = D$ eine Diagonalmatrix.
Sind zwei Matrizen diagonalisierbar, so sieht man leicht, dass sie genau dann zueinander ähnlich (konjugiert) sind, wenn sie die gleichen Eigenwerte mit den gleichen Vielfachheiten haben; ferner ist klar, dass eine nicht diagonalisierbare Matrix niemals zu einer diagonalisierbaren ähnlich sein kann. Offen bleibt im Moment die Frage, wie man von zwei nicht diagonalisierbaren Matrizen entscheidet, ob sie zueinander ähnlich sind. Wir werden diese Frage bei der Behandlung der Jordan'schen Normalform weiter untersuchen.

Eine geometrische Anwendung der Diagonalisierungstheorie ist:

Definition und Korollar 8.22 (Satz über die Hauptachsentransformation)**.**
Sei $A \in M_n(\mathbb{R})$ symmetrisch, Q_A das durch

$$Q_A(\mathbf{x}) := Q_A(x_1, \ldots, x_n) := {}^t\mathbf{x}A\mathbf{x} = \sum_{i,j=1}^{n} a_{ij}x_ix_j$$

definierte quadratische Polynom in den Koordinaten x_i von $\mathbf{x} \in \mathbb{R}^n$ (bzgl. der Standardbasis).

A und Q_A heißen positiv definit, wenn Q_A auf $\mathbb{R}^n \setminus \{\mathbf{0}\}$ nur positive Werte (d.h Werte > 0) annimmt, indefinit, wenn es dort positive und negative (d.h. < 0) Werte annimmt.

a) *A ist genau dann positiv definit, wenn alle Eigenwerte positiv sind, genau dann indefinit, wenn A positive und negative Eigenwerte hat.*

b) *Es gibt eine Orthonormalbasis $\mathbf{t}_1, \ldots, \mathbf{t}_n$ von $\mathbb{R}^n$, so dass für $c \in \mathbb{R}$ gilt:*

$$\{\mathbf{x} \in \mathbb{R}^n \mid Q_A(\mathbf{x}) = c\} = \{\sum_{i=1}^n y_i \mathbf{t}_i \mid \sum_{i=1}^n \lambda_i y_i^2 = c\},$$

wobei die λ_i die (evtl. mehrfach vorkommenden) Eigenwerte von A sind.

c) *Sind zuätzlich $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$,*

$$\begin{aligned} \mathcal{Q} &:= \{\mathbf{x} \in \mathbb{R}^n \mid {}^t\mathbf{x}A\mathbf{x} + {}^t\mathbf{x}\mathbf{b} + \mathbf{c} = 0\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \mid \textstyle\sum_{i,j=1}^n a_{ij}x_ix_j + \sum_{j=1}^n b_jx_j + c = 0\} \end{aligned}$$

die durch $A, \mathbf{b}, c$ gegebene Quadrik.

Dann gibt es $\mathbf{a} \in \mathbb{R}^n$ und $T \in O_n(\mathbb{R})$ (mit Spalten $\mathbf{t}_1, \ldots, \mathbf{t}_n$), so dass $\mathcal{Q}$ bezüglich des (kartesischen) Koordinatensystems mit Ursprung in $\mathbf{a}$ und Achsen in Richtung der $\mathbf{t}_i$ gegeben ist als

$$\mathcal{Q} = \{\mathbf{a} + \sum_{i=1}^n x_i' \mathbf{t}_i \mid \sum_{i=1}^r \lambda_i x_i'^2 + c' = 0\}$$

oder als

$$\mathcal{Q} = \{\mathbf{a} + \sum_{i=1}^n x_i' \mathbf{u}_i \mid \sum_{i=1}^r \lambda_i x_i'^2 + \mu x_n' = 0\};$$

dabei ist $r = \mathrm{rg}(A)$, $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ alle von 0 verschieden.

Ein Orthonormalsystem $\mathbf{t}_1, \ldots, \mathbf{t}_n$ von Vektoren mit dieser Eigenschaft heißt ein Hauptachsensystem der Quadrik $\mathcal{Q}$. Die λ_i sind dabei gegeben durch

$$\chi_A = X^{n-r} \prod_{i=1}^r (X - \lambda_i),$$

die $\mathbf{t}_i$ sind Eigenvektoren von A zu den Eigenwerten λ_i.

Beweis. (Im Fall $\mathbf{b} = \mathbf{0}$:) Als $\mathbf{t}_i$ nehme man die Spalten der Matrix T aus dem vorigen Satz. Der allgemeine Fall erfordert noch eine geeignete Koordinatentransformation, die den Ursprung verschiebt. Einzelheiten findet man z.B. im Buch von Lorenz. □

Beispiel:

a) Ist $A = E_n$ die Einheitsmatrix, so ist $\{\mathbf{x} \in \mathbb{R}^n \mid Q_A(\mathbf{x}) = r^2\}$ die Oberfläche der Kugel vom Radius r um den Ursprung als Mittelpunkt im $\mathbb{R}^n$, für $n = 2$ also der Rand des Kreises vom Radius r um den Ursprung.
b) Ist etwa $n = 2, A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\mathbf{b} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $c = 0$, so ist $\{\mathbf{x} \in \mathbb{R}^n \mid {}^t\mathbf{x}A\mathbf{x} + {}^t\mathbf{x}\mathbf{b} + \mathbf{c} = 0\}$ die Normalparabel im $\mathbb{R}^2$.
c) Ist $n = 2, A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, $c = 1$, so ist $\{\mathbf{x} \in \mathbb{R}^2 \mid Q_A(\mathbf{x}) = 1\} = \{\mathbf{x} \mid x_1x_2 = 1\}$ eine zweiästige Hyperbel. Für $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ erhält man $\{\mathbf{x} \mid x_1^2 - x_2^2 = 0\}$, das ist die gleiche Figur, nur um einen Winkel von 45° gedreht.
d) Für positiv definites A erhält man im Fall $n = 2, \mathbf{b} = \mathbf{0}, c > 0$ als $\{\mathbf{x} \in \mathbb{R}^n \mid Q_A(\mathbf{x}) = c\}$ eine Ellipse mit Halbachsen $\lambda_1^{-1}, \lambda_2^{-1}$, im Fall $n = 3, c > 0$ ein Ellipsoid mit Halbachsen $\lambda_1^{-1}, \lambda_2^{-1}, \lambda_3^{-1}$. Für indefinites A erhält man eine Hyperbel bzw. ein Hyperboloid. Explizite Beispiele und Bilder finden Sie im MAPLE-Worksheet zum Thema Hauptachsen.

Bemerkung. Im Fall $\mathbf{b} \neq \mathbf{0}$ gilt:

a) Ist $r = n$, so erreicht man stets die erste der angegebenen Normalformen für die Quadrik.
b) Ist $r = n$ und sind die λ_i paarweise verschieden, so sind die $\mathbf{u}_i$ bis auf einen Faktor ± 1 eindeutig bestimmt.
c) Im Fall $n = 2, r = 1$ erhält man hier (falls $\mu \neq 0$) z. B. eine Parabel.

Zusammenfassung:

Eigenvektoren eines Endomorphismus f des Vektorraums V sind Vektoren $v \neq \mathbf{0}$, so dass $f(v) = \lambda v$ für ein $\lambda \in K$ ist; dieses λ heißt der zugehörige Eigenwert. Eigenvektoren zu verschiedenen Eigenwerten sind linear unabhängig, Eigenräume zu verschiedenen Eigenwerten bilden eine direkte Summe.

Ist V endlichdimensional, so sind die Eigenwerte genau die Nullstellen des charakteristischen Polynoms $\chi_f = \det(X\mathrm{Id}_\mathrm{V} - f)$. Ist f diagonalisierbar(ähnlich, konjugiert zu einer Diagonalmatrix), so zerfällt χ_f in Linearfaktoren, das gleiche gilt sogar, wenn f nur trigonalisierbar (ähnlich, konjugiert zu einer (oberen oder unteren) Dreiecksmatrix) ist.
Zerfällt umgekehrt χ_f in Linearfaktoren, so ist f trigonalisierbar, sind diese Linearfaktoren paarweise verschieden, so ist f sogar diagonalisierbar. Kommen Linearfaktoren mehrfach vor, so können wir im Moment noch nicht entscheiden, ob f diagonalisierbar ist oder nicht, es gibt Beispiele für beide Möglichkeiten.

Reelle symmetrische Matrizen sind stets diagonalisierbar, ihre Diagonalisierung erlaubt es z.B., Ellipsen in der Ebene bzw. Ellipsoide im Raum in Hauptachsengestalt zu bringen.

9. Bilinearformen, hermitesche Formen und Skalarprodukte

Wir haben bereits in Abschnitt 7 Multilinearformen betrachtet. Ein Spezialfall hiervon ist:

Definition und Lemma 9.1. *Sei V ein K-Vektorraum. Eine Bilinearform auf V ist eine 2-fache Multilinearform, also eine Abbildung $\beta : V \times V \longrightarrow K$, für die gilt:*

$$\begin{aligned} \beta(v_1 + v_2, w) &= \beta(v_1, w) + \beta(v_2, w) \\ \beta(v, w_1 + w_2) &= \beta(v, w_1) + \beta(v, w_2) \\ \beta(\lambda v, w) &= \lambda\beta(v, w) = \beta(v, \lambda w) \end{aligned}$$

(für alle $v, v_1, v_2, w, w_1, w_2 \in V$, $\lambda \in K$).
(β ist in jedem Argument linear.)
Gilt $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$, so heißt β symmetrisch. Gilt $\beta(v, v) = 0$ für alle $v \in V$, so heißt β alternierend, in diesem Fall gilt $\beta(v, w) = -\beta(w, v)$ für alle $v, w \in V$.

Beispiel:

a) Das *Standardskalarprodukt* auf $\mathbb{R}^n$ ist definiert durch

$$\langle \mathbf{x}, \mathbf{y} \rangle := {}^t\mathbf{x}\mathbf{y} = {}^t\mathbf{y}\mathbf{x} = \sum_{i=1}^{n} x_i y_i.$$

Es gilt für $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$:

$$\begin{aligned} \langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} \rangle &= \langle \mathbf{x}_1, \mathbf{y} \rangle + \langle \mathbf{x}_2, \mathbf{y} \rangle \\ \langle \mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2 \rangle &= \langle \mathbf{x}, \mathbf{y}_1 \rangle + \langle \mathbf{x}, \mathbf{y}_2 \rangle \\ \langle \lambda\mathbf{x}, \mathbf{y} \rangle &= \lambda\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \lambda\mathbf{y} \rangle \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \mathbf{y}, \mathbf{x} \rangle \end{aligned}$$

Das *Standardskalarprodukt* ist also eine symmetrische Bilinearform auf $\mathbb{R}^n$.

Die genauso definierte symmetrische Bilinearform auf K^n für einen beliebigen Grundkörper K heißt die Einheitsform.

Ist $U \subseteq \mathbb{R}^n$ ein Unterraum mit Basis $\mathcal{B} = (v_1, \ldots, v_r)$, so ist die Einschränkung des Standardskalarprodukts auf U natürlich ebenfalls eine symmetrische Bilinearform.

b) Ist $A \in M_n(K)$, so wird durch

$$(\mathbf{x}, \mathbf{y}) \mapsto \beta_A(\mathbf{x}, \mathbf{y}) := {}^t\mathbf{y}A\mathbf{x} = \sum_{i,j=1}^{n} a_{ij} y_i x_j$$

eine Bilinearform β_A auf K^n definiert. Diese ist symmetrisch, wenn die Matrix A symmetrisch ist. Ist $2 \neq 0$ in K (die Charakteristik $\mathrm{char}(K)$ von K ist nicht 2) und A schiefsymmetrisch, so ist β_A alternierend.

Umgekehrt sei β eine Bilinearform auf K^n und $a_{ij} := \beta(\mathbf{e}_i, \mathbf{e}_j)$ für die Standardbasis $(\mathbf{e}_i)_i$ von K^n, sei $A = (a_{ij}) \in M_n(K)$ (A heißt die *Gram'sche Matrix* von β).
Dann rechnet man sofort nach, dass $\beta = \beta_A$ ist.

Speziell für den Fall des Grundkörpers $\mathbb{C}$ betrachten wir noch eine Variante des Begriffs Bilinearform.
Dafür brauchen wir noch ein paar Eigenschaften der komplexen Zahlen:

Definition und Lemma 9.2. *Für $z = a + bi$ sei die* komplex konjugierte *Zahl als $\overline{z} = a - bi$ definiert. Dann gilt:*

a) *Die Abbildung $z \longmapsto \overline{z}$ (*komplexe Konjugation*) ist ein Automorphismus des Körpers $\mathbb{C}$, d.h., es gilt $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$, $\overline{z_1 \cdot z_2} = \overline{z}_1 \cdot \overline{z}_2$ für alle $z_1, z_2 \in \mathbb{C}$.*
b) *Für $z = a + bi$ ($a, b \in \mathbb{R}$) ist*
$$\mathrm{Re}(z) := a = \frac{z + \overline{z}}{2}, \quad \mathrm{Im}(z) := b = \frac{z - \overline{z}}{2i}.$$
Es gilt $\mathbb{R} = \{z \in \mathbb{C} \mid z = \overline{z}\}$,
$$i\mathbb{R} = \{bi \mid b \in \mathbb{R}\} = \{z \in \mathbb{C} \mid z = -\overline{z}\},$$
$|\mathrm{Re}(z)| \leq |z|$ *für alle $z \in \mathbb{C}$.*
c) *Für $z = a + bi \in \mathbb{C}$ sei $|z| = \sqrt{a^2 + b^2} \in \mathbb{R}$ der Betrag von z. Dann gilt*
$$|z|^2 = |\overline{z}|^2 = z \cdot \overline{z},\ |z_1 z_2| = |z_1|\ |z_2|, |\mathrm{Re}(z)| \leq |z|$$
für alle $z, z_1, z_2 \in \mathbb{C}$, und man hat $|z| \geq 0$ mit $|z| = 0$ nur für $z = 0$.
d) *Für den komplexen Betrag gilt die* Dreicksungleichung*:*
$|z_1 + z_2| \leq |z_1| + |z_2|$ *für $z_1, z_2 \in \mathbb{C}$.*

Beweis. a) bis c) rechnet man leicht nach (Übung), nur die Dreicksungleichung ist nicht offensichtlich. Für sie quadriere man beide Seiten. Man hat dann $(|z_1| + |z_2|)^2 = |z_1|^2 + |z_2|^2 + 2|z_1||z_2|$ und und $|z_1 + z_2|^2 = |z_1|^2 + |z_2|^2 + 2\mathrm{Re}(z_1\overline{z_2})$. Wegen $|\mathrm{Re}(z_1\overline{z_2})| \leq |z_1|\ |\overline{z_2}| = |z_1||z_2|$ folgt die Behauptung. □

Definition 9.3. *Sei V ein Vektorraum über dem Körper $\mathbb{C}$ der komplexen Zahlen. Eine Abbildung $\beta : V \times V \longrightarrow \mathbb{C}$ heißt eine* hermitesche Form*, wenn für alle $u_1, u_2, u, v, v_1, v_2 \in V, \lambda \in \mathbb{C}$ gilt:*

$$\begin{array}{lcl}
\beta(u_1 + u_2, v) & = & \beta(u_1, v) + \beta(u_2, v) \\
\beta(\lambda u, v) & = & \lambda\beta(u, v) \\
\beta(u, v_1 + v_2) & = & \beta(u, v_1) + \beta(u, v_2) \\
\beta(u, \lambda v) & = & \overline{\lambda}\beta(u, v) \\
\beta(v_2, v_1) & = & \overline{\beta(v_1, v_2)}
\end{array}$$

Eine Matrix $A \in M_n(\mathbb{C})$ heißt hermitesch, wenn $A^* := {}^t\overline{A} = A$ gilt, schiefhermitesch, wenn $A^* := {}^t\overline{A} = -A$ gilt, dabei entsteht ${}^t\overline{A} = (\overline{a_{ij}})$ aus ${}^tA$ durch komplexe Konjugation aller Einträge.
Eine Matrix $U \in M_n(\mathbb{C})$ heißt unitär, wenn $U^* := {}^t\overline{U} = U^{-1}$ gilt.

Beispiel: Sei $A \in M_n(\mathbb{C})$ eine hermitesche Matrix.
Dann wird durch

$$\beta_A(\mathbf{x}, \mathbf{y}) := {}^t\overline{\mathbf{y}} A\mathbf{x} = \sum_{i,j=1}^{n} a_{ij}\overline{y_i}x_j$$

eine hermitesche Form β_A auf $\mathbb{C}^n$ definiert.
Ist $A = E_n$, so erhält man das Standard-Skalarprodukt $(\mathbf{x}, \mathbf{y}) \mapsto \sum_{i=1}^n x_i\overline{y_i}$. Genau wie oben kann man umgekehrt auch für beliebiges hermitesches β eine (notwendig hermitesche) Gram'sche Matrix A mit $\beta = \beta_A$ definieren.

Lemma 9.4. *Ist $\beta : V \times V \longrightarrow \mathbb{C}$ eine hermitesche Form, so ist $\beta(v, v) \in \mathbb{R}$ für alle $v \in V$.*

Beweis. Setzt man in der Definition $v_1 = v_2 = v$ ein, so erhält man $\beta(v, v) = \overline{\beta(v, v)}$. □

Definition 9.5. *Sei V ein K-Vektorraum mit $K = \mathbb{R}$ oder $K = \mathbb{C}$.*
$\beta : V \times V \longrightarrow \mathbb{C}$ sei eine symmetrische Bilinearform, falls $K = \mathbb{R}$ gilt, eine hermitesche Form im Falle $K = \mathbb{C}$.
β heißt positiv definit, wenn $\beta(v, v) > 0$ für alle $v \in V$, $v \neq \mathbf{0}$ gilt. Eine positiv definite symmetrische Bilinearform bzw. hermitesche Form β auf V heißt auch ein Skalarprodukt.
Ein endlichdimensionaler $\mathbb{R}$-Vektorraum mit einem Skalarprodukt heißt auch ein euklidischer Raum, *ein endlichdimensionaler $\mathbb{C}$-Vektorraum mit einem Skalarprodukt heißt ein* unitärer Raum.
Lässt man hier die Einschränkung auf endlichdimensionale Räume fort, so spricht man in beiden Fällen auch von einem Prä-Hilbert-Raum.

Beispiel:

a) Das Standardskalarprodukt auf $\mathbb{R}^n$ ist offenbar positiv definit, es ist also ein Skalarprodukt im Sinne der obigen Definition.
Ebenso ist die Einschränkung des Standardskalarprodukts auf einen beliebigen Unterraum $U \subseteq \mathbb{R}^n$ ein Skalarprodukt auf U.

b) Die hermitesche Standardform auf $\mathbb{C}^n$ ist

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{j=1}^{n} x_j\overline{y}_j.$$

Sie ist positiv definit, da $\sum_{i=1}^n |x_i|^2 \leq 0$ äquivalent zu $\mathbf{x} = \mathbf{0}$ ist.
Ist $U \subseteq \mathbb{C}^n$ ein Unterraum mit Basis $\mathcal{B} = (v_1, \ldots, v_r)$, so ist die Einschränkung des Standardskalarprodukts auf U ebenfalls eine hermitesche Form.

c) Dagegen ist für $V = \mathbb{C}^2$ die Form
$$\beta(\mathbf{x}, \mathbf{y}) = x_1\overline{y}_1 - x_2\overline{y}_2$$
zwar hermitesch, aber nicht positiv definit.

d) Sei V der (unendlichdimensionale) $\mathbb{C}$-Vektorraum der stetigen Fuktionen $f : [0, 1] \longrightarrow \mathbb{C}$.
Auf V wird dann ein Skalarprodukt durch
$$\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}dx$$
definiert, seine Einschränkung auf den in V enthaltenen $\mathbb{R}$-Vektorraum der stetigen Fuktionen $f : [0, 1] \longrightarrow \mathbb{R}$ ist natürlich ebenfalls ein Skalarprodukt.

Lemma 9.6. *Sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer K-Vektorraum (mit $K = \mathbb{R}$ oder $K = \mathbb{C}$) mit Skalarprodukt $\beta(v, w) = \langle v, w\rangle$. Dann gilt für die durch*
$$\|v\| := \sqrt{\langle v, v\rangle}$$
gegebene Norm *von v*

a) $\|v\| \geq 0$ *mit* $\|v\| = 0$ *nur für* $v = \mathbf{0}$.
b) $\|\lambda v\| = |\lambda|\ \|v\|$ *für alle* $\lambda \in \mathbb{C}$.
c) $\|v + w\| \leq \|v\| + \|w\|$ *(Dreiecksungleichung)*
d) *(*Cauchy-Schwarz'sche Ungleichung*)*
$$|\langle v, w\rangle| \leq \|v\| \cdot \|w\|.$$
In dieser Ungleichung steht genau dann das Gleichheitszeichen, wenn v und w linear abhängig sind.
e) *(*Parallelogrammgleichung*)*
$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

Beweis. a) und b) rechnet man leicht nach (Übung).
Die Cauchy-Schwarz'sche Ungleichung ist richtig, wenn $\langle v, w\rangle = 0$ gilt oder v und w linear abhängig sind, wir nehmen also an, dass keines von beiden der Fall ist. Ferner ändern sich beide Seiten der Ungleichung nicht, wenn man v durch $\frac{|\langle v,w\rangle|}{\langle v,w\rangle}v$ ersetzt. Da $\langle\frac{|\langle v,w\rangle|}{\langle v,w\rangle}v, w\rangle = |\langle v, w\rangle|$ gilt, können wir also im Weiteren zusätzlich annehmen, dass $\langle v, w\rangle = |\langle v, w\rangle|$ ist.
Dann gilt für $\lambda \in \mathbb{R}$
$$0 < \langle v + \lambda w, v + \lambda w\rangle = \|v\|^2 + 2\lambda|\langle v, w\rangle| + \lambda^2\|w\|^2.$$
Die rechte Seite ist eine quadratische Gleichung in der reellen Variablen λ, die keine reellen Nullstellen hat, also ist die Diskriminante $|\langle v, w\rangle|^2 - \|v\|^2\|w\|^2$ negativ, und das ist gerade die Behauptung.
Die Dreiecksungleichung folgt dann wegen
$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\mathrm{Re}(\langle v, w\rangle) \leq \|v\|^2 + \|w\|^2 + 2|\langle v, w\rangle|$$
aus der Cauchy-Schwarz'schen Ungleichung. $\square$

Bemerkung. In Analogie zur Situation in $\mathbb{R}^2$ und $\mathbb{R}^3$ schreibt man im euklidischen Fall

$$-1 \leq \cos(\alpha) = \frac{\langle v, w\rangle}{\|v\| \; \|w\|} \leq 1$$

und nennt α den Winkel zwischen den Geraden in Richtung von v und w.

Definition 9.7. *Sei $K = \mathbb{R}$ oder $K = \mathbb{C}$, sei V ein K-Vektorraum. Eine Abbildung $v \mapsto \|v\|$ heißt eine* Norm *auf V, wenn gilt:*

$$\begin{array}{rcl} \|v\| & \geq & 0 \text{ mit } \|v\| = 0 \text{ nur für } v = \mathbf{0} \\ \|\lambda v\| & = & |\lambda| \; \|v\| \text{ für } \lambda \in K \\ \|v + w\| & \leq & \|v\| + \|w\| \text{ (Dreiecksungleichung).} \end{array}$$

Das Paar $(V, \| \quad \|)$ heißt dann ein normierter Raum.

Korollar 9.8. *Ein euklidischer oder unitärer Vektorraum $(V, \langle \quad , \quad \rangle)$ wird durch*

$$\|v\| := \sqrt{\langle v, v\rangle}$$

zu einem normierten Raum.

Definition 9.9. *Sei V ein K-Vektorraum mit symmetrischer Bilinearform oder (im Fall $K = \mathbb{C}$) hermitescher Form $\beta : V \times V \longrightarrow K$.*

a) *Vektoren $v, w \in V$ heißen* orthogonal *oder* senkrecht *zueinander (bezüglich β), wenn $\beta(v, w) = 0$ gilt.*
b) *Unterräume U_1, U_2 von V heißen orthogonal (oder senkrecht) zueinander, wenn $\beta(u_1, u_2) = 0$ für alle $u_1 \in U_1, u_2 \in U_2$ gilt.*
   Sind $U_1, \ldots, U_r$ zueinander paarweise orthogonale Teilräume mit $U_1 \oplus \ldots \oplus U_r = V$, so schreibt man auch
   $$V = U_1 \perp \ldots \perp U_r$$
   und sagt, V sei die orthogonale direkte Summe der U_j.
   Ist dann $v = u_1 + \cdots + u_r \in V$ mit $u_i \in U_i$ $(1 \leq i \leq r)$, so heißen die u_i die orthogonalen Projektionen von v auf die U_i.
c) *Eine Basis $(v_1, \ldots, v_n)$ von V heißt* Orthogonalbasis *von V* bezüglich *β, falls die v_i paarweise orthogonal sind. Die Basis $(v_1, \ldots, v_n)$ heißt* Orthonormalbasis, *wenn*
   $$\beta(v_i, v_j) = \delta_{ij} \text{ für } 1 \leq i, j \leq n$$
   gilt.

Offensichtlich bilden die Standardbasisvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$ eine Orthonormalbasis des $\mathbb{R}^n$ bzw. des $\mathbb{C}^n$ bezüglich des Standardskalarprodukts.

Lemma 9.10. *Sei V ein K-Vektorraum mit symmetrischer Bilinearform oder (im Fall $K = \mathbb{C}$) hermitescher Form β, seien $v_1, \ldots, v_n \in V$ paarweise orthogonale Vektoren mit $\beta(v_i, v_i) \neq 0$ für $1 \leq i \leq n$.*
Dann sind $v_1, \ldots, v_n$ linear unabhängig.

Beweis. Sind $a_1, \ldots, a_n \in K$ mit $\sum_{i=1}^n a_i v_i = \mathbf{0}$, so ist

$$0 = \beta(\sum_{i=1}^n a_i v_i, v_j) = a_j \beta(v_j, v_j) \ (1 \leq j \leq n),$$

mit $\beta(v_j, v_j) \neq 0$, also $a_j = 0$ für $1 \leq j \leq n$. □

Definition und Lemma 9.11. *Sei V ein euklidischer oder unitärer Vektorraum über $K = \mathbb{R}$ bzw. $K = \mathbb{C}$, sei $\mathbf{0} \neq u \in V$.*

a) *Für $\mathbf{0} \neq v \in V$ ist $v_u := \frac{\langle v,u \rangle}{\langle u,u \rangle} u$ die orthogonale Projektion von v auf die von u aufgespannte Gerade $\mathrm{Lin}(u) = K \cdot u$ und $v - v_u$ die orthogonale Projektion auf $\mathrm{Lin}(u)^\perp$.*
b) *$s_u(v) := v - 2v_u$ heißt die Spiegelung von v an der zu u orthogonalen Hyperebene $\mathrm{Lin}(u)^\perp$.*

Beispiel: Ist $V = \mathbb{R}^n$, $u = \lambda \mathbf{e}_1$ mit $\lambda \neq 0$ und $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, so ist $\mathbf{x} - \mathbf{x}_u = \sum_{i=2}^n x_i \mathbf{e}_i$ und $s_u(\mathbf{x}) = -x_1 \mathbf{e}_1 + \sum_{i=2}^n x_i \mathbf{e}_i$.

Der folgende Satz zeigt, dass sich ein beliebiges Skalarprodukt durch einen geeigneten Basiswechsel auf das Standardskalarprodukt zurückführen lässt und liefert auch gleich einen Algorithmus für die Bestimmung der Matrix des Basiswechsels.

Satz 9.12 (Gram-Schmidt Orthogonalisierung)**.** *Sei $K = \mathbb{R}$ oder $K = \mathbb{C}$, sei V ein euklidischer bzw. unitärer K-Vektorraum mit Skalarprodukt $\beta(\ \ ,\ \ ) = \langle\ \ ,\ \ \rangle$, sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V.*

Dann gibt es eine Basis $\mathcal{B}' = (w_1, \ldots, w_n)$ von V, die bezüglich $\langle\ \ ,\ \ \rangle$ eine Orthonormalbasis ist und für die

$$\mathrm{Lin}(v_1, \ldots, v_j) = \mathrm{Lin}(w_1, \ldots, w_j) \quad \textit{für } 1 \leq j \leq n$$

gilt (für die also die Matrix des Basiswechsels von $\mathcal{B}$ zu $\mathcal{B}'$ eine obere Dreiecksmatrix ist).

Beweis. Es reicht offenbar, eine Basis $(w_1, \ldots, w_n)$ zu finden, für die $\langle w_i, w_i \rangle$ nicht 1 sein muss, die aber die anderen im Satz angegebenen Bedingungen erfüllt, da man dann durch Übergang zu $\sqrt{\langle w_i, w_i \rangle}^{-1} w_i$ auch die noch fehlende Normierungsbedingung erfüllen kann.
Wir führen den Beweis durch Induktion nach n, der Fall $n = 1$ ist trivial. Sei also $n \geq 2$ und die Behauptung gelte für Vektorräume der Dimension $n - 1$.
Sei $w_1 = v_1$, für $2 \leq i \leq n$ sei

$$v_i' = v_i - \frac{\langle v_i, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1$$

die Projektion von v_i auf den zu $v_1 = w_1$ orthogonalen Unterraum.
Dann ist $\langle v_i', w_1 \rangle = 0$ für $2 \leq i \leq n$ und $\mathrm{Lin}(w_1, v_2', \ldots v_j') = \mathrm{Lin}(v_1, v_2, \ldots v_j)$ für $2 \leq j \leq n$. Nach Induktionsannahme gibt es $w_2, \ldots, w_n \in V$, die

eine Orthogonalbasis von $\mathrm{Lin}(v_2', \dots v_n') = \mathrm{Lin}(w_1)^\perp$ bilden und daher auch alle orthogonal zu w_1 sind, und für die

$$\mathrm{Lin}(v_2, \dots, v_j) = \mathrm{Lin}(w_2, \dots, w_j) \quad \text{für } 1 \le j \le n$$

gilt. Die Vektoren $w_1, \dots, w_n$ sind dann zusammen wie gewünscht. □

Lemma 9.13. *Sei V ein euklidischer bzw. unitärer K-Vektorraum ($K = \mathbb{R}$ oder $K = \mathbb{C}$) mit Skalarprodukt $\beta(\ ,\ ) = \langle\ ,\ \rangle$, sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V und $(w_1, \dots w_n)$ eine weitere Basis mit $w_j = \sum_{i=1}^n t_{ij} v_i$.*
Seien $A = (\langle v_i, v_j\rangle)_{i,j}$ und $B = (\langle w_i, w_j\rangle)_{i,j}$ die Gram-Matrizen von $\langle\ ,\ \rangle$ bezüglich der beiden Basen.
Dann ist

$$B = {}^tT A \overline{T}.$$

Beweis. Nachrechnen! □

Korollar 9.14. *Ist A eine positiv definite hermitesche Matrix, so gibt es $T \in \mathrm{GL}_n(\mathbb{C})$ mit $A = {}^tT \cdot \overline{T}$. Die Matrix T kann als obere (untere) Dreiecksmatrix gewählt werden; ist A reell, so kann auch T reell gewählt werden.*

Beweis. Ist $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ eine Orthonormalbasis des $\mathbb{C}^n$ bezüglich des Skalarprodukts β_A und $U \in GL_n(\mathbb{C})$ die Matrix mit den Spalten $\mathbf{u}_1, \dots, \mathbf{u}_n$, so ist der i,j-Eintrag von $\overline{{}^tU} AU$ gleich ${}^t\overline{\mathbf{u}_i} A \mathbf{u}_j = \delta_{ij}$, d.h., es gilt $\overline{{}^tU} AU = E_n$. Mit $T := \overline{U^{-1}}$ folgt die Behauptung. Ist A reell, so kann hier durchweg $\mathbb{C}$ durch $\mathbb{R}$ ersetzt werden. □

Bemerkung.

a) Das Gram-Schmidt-Verfahren liefert den Übergang von A zu ${}^tT^{-1} A \overline{T}^{-1} = E_n$ als eine Abfolge von simultanen Zeilen- und Spaltenumformungen: In jedem Schritt des Verfahrens wird ein Vektor v_j durch einen Vektor $\tilde{v}_j = v_j + \sum_{k=1}^{j-1} \lambda_{kj} v_k$ ersetzt. Die Gram-Matrix wird in diesem Schritt geändert, indem gleichzeitig zur j-ten Zeile für $1 \le k \le j-1$ die mit λ_{kj} multiplizierte k-te Zeile addiert und zur j-ten Spalte für $1 \le k \le j-1$ die mit $\overline{\lambda_{kj}}$ multiplizierte k-te Spalte addiert wird. Abschließend werden die v_j normiert, also jede Zeile/Spalte mit $\frac{1}{\|v_j\|}$ multipliziert.
b) Das Gram-Schmidt-Verfahren ist nichts anderes als eine Verallgemeinerung des Verfahrens der quadratischen Ergänzung.
c) Der Beweis des Satzes von Gram und Schmidt liefert sogar einen leicht implementierbaren Algorithmus
d) Das Gram-Schmidt-Verfahren kann auch angewendet werden, um in unendlichdimensionalen Vektorräumen mit Skalarprodukt Orthogonal- bzw. Orthonormalsysteme zu finden. Insbesondere in Vektorräumen von Funktionen mit einem über das Integral definierten Skalarprodukt ist das eine der häufigsten Anwendungen

des Verfahrens; dies ist auch die Situation, in der das Verfahren von Gram und Schmidt eingeführt wurde.

Beispiel. Die Matrix $\begin{pmatrix} a & b \\ \overline{b} & c \end{pmatrix}$ wird durch die Transformation:

$$\begin{array}{ll} \text{2. Zeile} & \longmapsto \text{2. Zeile} - \frac{\overline{b}}{a} \cdot \text{erste Zeile} \\ \text{2. Spalte} & \longmapsto \text{2. Spalte} - \frac{b}{a} \cdot \text{erste Spalte} \end{array}$$

in Diagonalgestalt überführt, diese liefert die Matrix $\begin{pmatrix} a & 0 \\ 0 & c - \frac{|b|^2}{a} \end{pmatrix}$, die zu der Orthogonalbasis aus $v_1, v_2' = v_2 - \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = v_2 - \frac{\overline{b}}{a} v_1$ gehört. Dementsprechend geht für $v = x_1 v_1 + x_2 v_2$ der Wert $\langle v, v \rangle = a|x_1|^2 + cg|x_2^2| + 2\mathrm{Re}(bx_1\overline{x}_2)$ (mit $v = x_1' v_1' + x_2' v_2'$, $v_1' = v_1$, $v_2' = v_2 - \frac{\overline{b}}{a} v_1$) über in $\langle v, v \rangle = a|x_1'|^2 + (c - \frac{|b|^2}{a})|x_2'|^2$ mit $x_1' = x_1 + \frac{\overline{b}}{a} x_2$, $x_2' = x_2$ (insbesondere für reelle a, b, c, x_1, x_2 ist das genau die Formel der quadratischen Ergänzung).

Satz 9.15. *Sei V ein (endlichdimensionaler) euklidischer oder unitärer Raum über K ($K = \mathbb{R}$ oder $K = \mathbb{C}$) mit Skalarprodukt $\langle\ ,\ \rangle$, sei $U \subseteq V$ ein Untervektorraum. Dann gilt:*

a) *Das orthogonale Komplement $U^\perp = \{v \in V \mid \langle v, u \rangle = 0$ für alle $u \in U\}$ ist ein Unterraum von V.*
b) *Es ist $V = U \oplus U^\perp$, insbesondere ist*

$$\dim(U^\perp) = \dim(V) - \dim(U).$$

Man schreibt in dieser Situation auch: $V = U \perp U^\perp$.

Beweis. Sei $(v_1, \ldots, v_n)$ eine Basis von V, deren erste r Vektoren eine Basis von U bilden. Wir wenden auf diese Basis das Gram-Schmidt-Verfahren an und erhalten eine Orthonormalbasis $(w_1, \ldots, w_n)$, deren erste r Vektoren eine Basis von U bilden. Dann ist offenbar $\mathrm{Lin}(w_{r+1}, \ldots, w_n)$ eine Basis von $U^\perp$, und das zeigt alle Behauptungen. □

Bemerkung. a) gilt auch für beliebige hermitesche oder schiefhermitesche Formen (bzw. symmetrische oder alternierende Bilinearformen). Auch die Aussage von Teil b) lässt sich verallgemeinern; das werden wir in Teil II der Vorlesung weiter untersuchen.

Definition und Lemma 9.16. *Sei V ein unitärer Vektorraum über $\mathbb{C}$ mit Skalarprodukt $\langle\ ,\ \rangle$. Ein Endomorphismus $f \in \mathrm{End}(V)$ heißt* selbstadjungiert, *wenn*

$$\langle f(v), w \rangle = \langle v, f(w) \rangle$$

für alle $v, w \in V$ gilt.

a) *f ist genau dann selbstadjungiert, wenn bezüglich jeder Orthonormalbasis $\mathcal{B} = (v_1, \ldots, v_n)$ von V die Matrix $A = M_{\mathcal{B}}(f)$ hermitesch ist.*

b) *Ist f selbstadjungiert, so sind alle Eigenwerte von f reell.*

Beweis. a) folgt mit $A = (a_{ij})$ aus $\langle f(v_j), v_i\rangle = \langle \sum_k a_{kj} v_k, v_i\rangle = a_{ij}$ und $\langle v_j, f(v_i)\rangle = \overline{a_{ji}}$.
Für b) sei $v \neq \mathbf{0}$ mit $f(v) = \lambda v$. Dann ist $\lambda\langle v, v\rangle = \langle f(v), v\rangle = \langle v, f(v)\rangle = \bar{\lambda}\langle v, v\rangle$, also $\lambda = \bar{\lambda}$, d. h., $\lambda \in \mathbb{R}$. □

Satz 9.17. *Sei $A \in M_n(K)$ mit $K = \mathbb{R}$ oder $K = \mathbb{C}$ eine hermitesche Matrix. Dann gibt es eine orthogonale ($K = \mathbb{R}$) bzw unitäre Matrix U, für die ${}^t\overline{U}AU$ eine reelle Diagonalmatrix ist.*

Beweis. Wir betrachten den selbstadjungierten Endomorphismus $f = L_A$ von $V = \mathbb{C}^n$ (mit dem Standardskalarprodukt) und zeigen durch Induktion nach n, dass V eine Orthonormalbasis aus Eigenvektoren von f hat; die gesuchte Matrix U erhält man dann, indem man diese Vektoren als Spalten von U einträgt.
Das charakteristische Polynom $\chi_A = \chi_f$ hat in $\mathbb{C}$ wenigstens eine Nullstelle λ, diese ist nach dem vorigen Lemma reell. Sei $\mathbf{v}$ ein Eigenvektor von f zum Eigenwert λ, und $\mathbf{v}$ kann als Lösung des reellen linearen Gleichungssystems $(A - \lambda E_n)\mathbf{v} = \mathbf{0}$ in $\mathbb{R}^n$ gewählt werden.
Sei U das orthogonale Komplement von $\mathrm{Lin}(v)$, wir wissen, dass $V = \mathrm{Lin}(v) \perp U$ gilt, insbesondere ist $\dim(U) = n - 1$. Für $u \in U$ ist $\langle v, f(u)\rangle = \langle f(v), u\rangle = \lambda\langle v, u\rangle = 0$, also auch $f(u) \in \mathrm{Lin}(v)^\perp = U$. Wir können also $f|U : U \to U$ als Endomorphismus des euklidischen Raums U (mit dem auf U eingeschränkten Skalarpodukt) auffassen und finden nach Induktionsannahme eine Orthonormalbasis aus Eigenvektoren von $f|_U$ von U. Diese bildet zusammen mit v die gesuchte Orthonormalbasis von $V = \mathbb{C}^n$. Man prüft leicht nach, dass man die Orthonormalbasis tatsächlich in $\mathbb{R}^n$ wählen kann, wenn A reell ist, so dass man in diesem Fall als U eine reelle orthogonale Matrix erhält. □

Definition und Satz 9.18. *Sei $A \in M_n(\mathbb{R})$ eine symmetrische Matrix, sei n_+ die Anzahl der (strikt) positiven und n_- die Anzahl der negativen Eigenwerte von A.*
Ein Teilraum $U \subseteq \mathbb{R}^n$ heißt positiv (bzw. negativ) definit für β_A, wenn die Einschränkung von β_A auf $U \times U$ positiv (bzw. negativ) definit ist.
Dann gilt (Trägheitssatz von Sylvester)*: Jeder maximale für β_A positiv definite Teilraum von $\mathbb{R}^n$ hat Dimension n_+.*
Das Paar (n_+, n_-) (bzw. oft auch die Differenz $n_+ - n_-$) heißt die Signatur von β_A.

Beweis. Sei $(v_1, \ldots, v_n)$ eine Orthonormalbasis des $\mathbb{R}^n$ aus Eigenvektoren von A zu den Eigenwerten $\lambda_1, \ldots, \lambda_n$, die so angeordnet ist, dass die ersten n_+ Eigenwerte > 0 sind, sei $U_1 = \mathrm{Lin}(v_1, \ldots, v_{n_+})$ und $U_2 = U_1^\perp = \mathrm{Lin}(v_{n_+ + 1}, \ldots, v_n)$. Dann ist wegen der Dimensionsformel für Unterräume $U \cap U_2 \neq \{\mathbf{0}\}$ für jeden Unterraum $U \subseteq \mathbb{R}^n$ mit $\dim(U) > n_+$, also gibt es in jedem solchen U einen Vektor $u \neq \mathbf{0}$ mit $\langle u, u\rangle \leq 0$, der Unterraum U kann also nicht positiv definit sein.

Andererseits sei U ein beliebiger Teilraum, so dass die Einschränkung von β_A auf $U \times U$ positiv definit ist.
Sei $(u_1, \ldots, u_r)$ eine Orthonormalbasis bezüglich β_A von U, die durch $v_{r+1}, \ldots, v_n$ zu einer Basis von V ergänzt wird. Setzt man wie im Beweis des Gram-Schmidt-Satzes

$$u_j = v_j - \sum_{i=1}^{r} \frac{\beta_A(v_j, u_i)}{\beta_A(u_i, u_i)} u_i$$

für $r+1 \leq j \leq n$, so spannen $u_{r+1}, \ldots, u_n$ einen $n-r$-dimensionalen Teilraum U' auf, der bezüglich β_A auf U senkrecht steht ($\beta_A(u, u') = 0$ für alle $u \in U, u' \in U'$). Ist $r < n_+$, so ist $U' \cap U_1 \neq \{\mathbf{0}\}$, und mit $\mathbf{0} \neq u' \in U' \cap U_1$ ist $U \oplus \mathbb{R}u'$ ein bezüglich β_A positiv definiter Unterraum, der U echt enthält, ein solches U ist also nicht maximal positiv definit bezüglich β_A. □

Wir wollen jetzt noch ein paar geometrische Anwendungen des Skalarprodukts betrachten.

Lemma 9.19. *Sei $V = \mathbb{R}^n$, $\langle\ ,\ \rangle$ das Standardskalarprodukt. Sei $H_0 \subseteq \mathbb{R}^n$ ein $(n-1)$-dimensionaler Teilraum (eine Hyperebene durch $\mathbf{0}$), $\mathbf{v} \in V$ und $H = \mathbf{v} + H_0$ (die zu H_0 parallele affine Hyperebene durch $\mathbf{v}$). Sei $\mathbf{y} \in V$.*
Dann gibt es genau ein $\mathbf{u}_0 \in H$, so dass $\mathbf{y} - \mathbf{u}_0 \in H_0^\perp$ gilt. Für dieses $\mathbf{u}_0$ gilt:

$$\|\mathbf{y} - \mathbf{u}_0\| = \min\{\|\mathbf{y} - \mathbf{x}\| \mid \mathbf{x} \in H\} =: d(\mathbf{y}, H).$$

Beweis. Wir haben $V = H_0 \perp H_0^\perp$, bekommen also eine eindeutige Zerlegung $\mathbf{y} - \mathbf{v} = \mathbf{u} + \mathbf{w}$ mit $\mathbf{u} \in H_0, \mathbf{w} \in H_0^\perp$. Dann ist $\mathbf{u}_0 = \mathbf{u} + \mathbf{v}$ der eindeutige Vektor $\mathbf{u}_0 \in H$ mit $\mathbf{y} - \mathbf{u}_0 \in H_0^\perp$.
Ist $\mathbf{x} \in H$ beliebig, so ist $\mathbf{u}' := \mathbf{u}_0 - \mathbf{x} \in H_0$, und man hat

$$\begin{aligned} \|\mathbf{y} - \mathbf{x}\| &= \|(\mathbf{y} - \mathbf{u}_0) + \mathbf{u}'\| \\ &= \|\mathbf{y} - \mathbf{u}_0\| + \|\mathbf{u}'\| \quad \text{wegen } \mathbf{y} - \mathbf{u}_0 \in H_0^\perp,\ \mathbf{u}' \in H_0 \\ &\geq \|\mathbf{y} - \mathbf{u}_0\|. \end{aligned}$$

□

Korollar 9.20. (Hesse'sche Normalform) *Sei $H = \mathbf{v} + H_0$ wie im vorigen Lemma. Dann gilt für $\mathbf{0} \neq \mathbf{a} \in H_0^\perp$:*

a) $H = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{x} - \mathbf{v} \rangle = 0\}$
b) *Mit* $\langle \mathbf{a}, \mathbf{v} \rangle = \sum_{i=1}^n a_i v_i =: b$ *ist*
$H = \{\mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^n a_i x_i = b\} = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{x} \rangle = b\}$
c) *Für* $\mathbf{y} \in \mathbb{R}^n$ *ist*

$$d(\mathbf{y}, H) = \frac{|\sum_{i=1}^n a_i y_i - b|}{\|\mathbf{a}\|};$$

insbesondere gilt für $\|\mathbf{a}\| = 1$ *(*$\mathbf{a}$ *ein Einheitsnormalenvektor):*

$$d(\mathbf{y}, H) = |\sum_{i=1}^{n} a_i y_i - b| \quad \textit{falls } \|\mathbf{a}\| = 1.$$

Beweis. a): Für $\mathbf{0} \neq \mathbf{a} \in H_0^\perp$ ist $H_0 = \mathrm{Lin}(\mathbf{a})^\perp$, also ist $\mathbf{x} \in H = \mathbf{v} + H_0$ genau dann, wenn $\mathbf{x} - \mathbf{v} \in \mathrm{Lin}(\mathbf{a})^\perp$ gilt, also genau dann, wenn $\langle \mathbf{a}, \mathbf{x} - \mathbf{v} \rangle = 0$ gilt.
b) folgt direkt aus a).
c) Sei $\mathbf{u}_0$ wie im vorigen Lemma. Da $H_0^\perp$ Dimension 1 hat, ist $\mathbf{y} - \mathbf{u}_0 = \lambda \mathbf{a}$ mit $\lambda \in \mathbb{R}$, und wir haben $\langle \mathbf{y} - \mathbf{u}_0, \mathbf{a} \rangle = \lambda \|\mathbf{a}\|^2$. Damit wird

$$\begin{aligned} d(\mathbf{y}, H) &= \|\mathbf{y} - \mathbf{u}_0\| \\ &= |\lambda| \; \|\mathbf{a}\| \\ &= \frac{|\langle \mathbf{y} - \mathbf{u}_0, \mathbf{a} \rangle|}{\|\mathbf{a}\|} \\ &= \frac{|\langle \mathbf{y}, \mathbf{a} \rangle - b|}{\|\mathbf{a}\|}. \end{aligned}$$

□

10. Dimensionsformel und Quotientenraum

Bei der Behandlung linearer Abbildungen ist es oft hilfreich, im Urbildraum V der linearen Abbildung $f : V \longrightarrow W$ die Äquivalenzklassen unter der Relation

$$v_1 \sim v_2 \Leftrightarrow f(v_1) = f(v_2)$$

zu betrachten. Offenbar gilt für diese Relation:

$$\begin{aligned} \{v \in V \mid v \sim v_1\} &= \{v \in V \mid v - v_1 \in \mathrm{Ker}(f)\} \\ &= \{v = v_1 + u \mid u \in \mathrm{Ker}(f)\} \\ &=: v_1 + \mathrm{Ker}(f). \end{aligned}$$

Wir verallgemeinern das und definieren:

Definition 10.1. *Sei G eine Gruppe, $U \subseteq G$ eine Untergruppe. Für $x \in G$ heißt $xU := \{xu \mid u \in U\}$ die Linksnebenklasse von x nach U und $Ux := \{ux \mid u \in U\}$ die Rechtsnebenklasse von x nach U. Ist $Ux = xU$ für alle $x \in G$, so heißt U ein Normalteiler in G.*

Beispiel: Ist $f : G \to H$ ein Homomorphismus von Gruppen, so ist $x\mathrm{Ker}(f) = \mathrm{Ker}(f)x = \{g \in G \mid f(g) = f(x)\}$, der Kern von f ist also ein Normalteiler. Speziell können wir hier $G = V$ und $H = W$ als Vektorräume über dem Körper K wählen und als f eine lineare Abbildung von V nach W betrachten; wir erhalten dann wie oben als Nebenklassen die Mengen $x + \mathrm{Ker}(f)$. Ist hier zum Beispiel $V = \mathbb{R}^3$, $W = \mathbb{R}^2$ und f durch $f(x_1, x_2, x_3) = (x_1, x_2)$ gegeben, so ist $\mathbf{x} + \mathrm{Ker}(f) = \{(x, y, x_3) \mid x, y \in \mathbb{R}\}$ die zur x_1, x_2-Ebene parallele Ebene in der Höhe x_3.
Generell gilt offenbar: Ist die Gruppe G abelsch (kommutativ), so ist jede Untergruppe ein Normalteiler.

Lemma 10.2. *Sei G eine Gruppe, $U \subseteq G$ eine Untergruppe, $x, y \in G$. Dann ist $y \in xU$ äquivalent zu $x^{-1}y \in U$ und $y \in Ux$ äquivalent zu $yx^{-1} \in U$.*
Durch $x \sim_\ell y \leftrightarrow x^{-1}y \in U$ bzw. $x \sim_r y \leftrightarrow yx^{-1} \in U$ werden Äquivalenzrelationen gegeben, deren Äquivalenzklassen die Links- bzw. die Rechtsnebenklassen von U in G sind.

Beweis. Klar, die Eigenschaften einer Äquivalenzrelation (Reflexivität, Symmetrie, Transitivität) rechnet man leicht nach. □

Satz 10.3. a) *Sei G eine Gruppe, $H \subseteq G$ ein Normalteiler (also $xH = Hx$ für alle $x \in G$). Dann wird auf der Nebenklassenmenge G/H durch*

$$(xH) \circ (yH) := (xy)H$$

eine wohldefinierte Verknüpfung eingeführt, bezüglich der G/H eine Gruppe ist. G/H mit dieser Verknüpfung heißt die Faktorgruppe *oder* Quotientengruppe *von G nach H.*

b) *Sei V ein K-Vektorraum, $U \subseteq V$ ein Unterraum. Dann wird die Faktorgruppe $V/U = \{v + U \mid v \in V\}$ durch*

$$\lambda(v + U) = \lambda v + U \quad (\lambda \in K)$$

ein K-Vektorraum (der Faktorraum *oder* Quotientenraum *von V nach U).*

Beweis. a): Wir müssen zunächst zeigen, dass die Verknüpfung wohldefiniert ist, d.h., wir müssen zeigen, dass das Ergebnis der Verknüpfung der Nebenklasse $N_1 = xH = x'H$ mit der Nebenklasse $N_2 = yH = y'H$ nicht davon abhängt, welche mögliche Darstellung der Nebenklassen man ausgewählt hat. Genauer ist zu zeigen:
Sind $x, x', y, y' \in G$ mit $xH = x'H, yH = y'H$, so gilt $xyH = x'y'H$.
Wir nutzen dafür aus, dass H ein Normalteiler ist und benutzen, dass man beim Rechnen mit Nebenklassen Klammern versetzen darf, dass also

$$(xy)H = \{(xy)h \mid h \in H\} = \{x(yh) \mid h \in H\} = \{xz \mid z \in yH\} = x(yH)$$

gilt und erhalten:

$$\begin{aligned}
x'y'H &= x'(y'H) \text{ wegen der Klammerregel} \\
&= x'(yH) \text{ weil } yH = y'H \text{ gilt} \\
&= x'(Hy) \text{ weil } H \text{ Normalteiler ist} \\
&= (x'H)y \text{ wegen der Klammerregel} \\
&= (xH)y \text{ weil } xH = x'H \text{ gilt} \\
&= x(Hy) \text{ wegen der Klammerregel} \\
&= x(yH) \text{ weil } H \text{ Normalteiler ist} \\
&= (xy)H \text{ wegen der Klammerregel.}
\end{aligned}$$

(Wer mag, kann auch stattdessen nachrechnen, dass aus $x' = xh_1, y' = yh_2$ folgt, dass es ein $h_3 \in H$ mit $x'y' = xyh_3$ gibt. Man muss dabei ausnutzen, dass man wegen der Normalteilereigenschaft von H ein $h_1' \in H$ mit $h_1 y = yh_1'$ finden kann.)
Dass für die so definierte Verknüpfung das Assoziativgesetz gilt, folgt dann sofort aus dem Assoziativgesetz für G. Auch dass die Nebenklasse $H = eH$ neutrales Element bezüglich dieser Verknüpfung ist und dass die Nebenklasse $a^{-1}H$ invers zur Nebenklasse aH ist, sieht man sofort.
b): Da die additive Gruppe von V kommutativ ist, brauchen wir uns um die Normalteilerbedingung keine Sorgen zu machen: Jeder Unterraum U von V ist auch Normalteiler in $(V, +)$, wir können also die Faktorgruppe V/U bilden und müssen zeigen, dass wir auf die angegebene Weise für diese eine Multiplikation mit Skalaren $\lambda \in K$ definieren können.
Seien also $v_1, v_2 \in V$ mit $v_1 + U = v_2 + U$, d.h. $v_1 - v_2 \in U$. Dann ist, weil U ein Unterraum ist, $\lambda(v_1 - v_2) \in U$, also $\lambda v_1 + U = \lambda v_2 + U$.

Die Verknüpfung ist also wohldefiniert, und die Gültigkeit von **V1** bis **V4** aus Definition 3.1 folgt wie oben direkt aus deren Gültigkeit für V. $\square$

Definition 10.4. *Die Elemente $v + U$ des Faktorraums V/U heißen* affine Unterräume *der Dimension* $\dim(U)$.
Ist $\dim(U) = 1$, *so spricht man von* affinen Geraden *(Geraden, die nicht notwendig durch den Ursprung gehen), ist* $\dim(U) = 2$, *so spricht man von* affinen Ebenen *(Ebenen, die nicht notwendig durch den Ursprung gehen).*

Bemerkung. Ist speziell $V = \mathbb{R}^3$, so gibt es für den Unterraum U die Möglichkeiten:

- $U = \{\mathbf{0}\}$. Die Nebenklasse $v + U$ besteht nur aus dem Vektor v, V/U ist isomorph zu V.
- $\dim(U) = 1$, d. h., U ist eine Gerade g durch den Ursprung. Die Nebenklasse $v + U$ ist als Punktmenge die Parallele zu g durch den Punkt P_v mit Ortsvektor v (affine Gerade durch den Punkt P_v parallel zu g). Der Faktorraum V/U ist zweidimensional, ein vollständiges Repräsentantensystem für seine Elemente (Nebenklassen) findet man in jeder Ursprungsebene, die die Gerade g nicht enthält (und daher ein zu U komplementärer Unterraum ist).
- $\dim(U) = 2$, d. h., U ist eine Ebene E durch den Ursprung. Die Nebenklasse $v + U$ ist als Punktmenge die zu E parallele Ebene durch den Punkt P_v mit Ortsvektor v. Der Faktorraum V/U ist 1-dimensional, ein vollständiges Repräsentantensystem für seine Elemente (Nebenklassen) findet man in jeder Ursprungsgeraden, die die Ebene E nicht enthält (und daher ein zu U komplementärer Unterraum ist).
- $U = V$, der Faktorraum V/U besteht nur aus der Nullklasse: $V/U = \{\mathbf{0}\}$.

Satz 10.5. *Sei V ein endlichdimensionaler K-Vektorraum, $U \subseteq V$ ein Unterraum, $p_U : V \longrightarrow V/U$ die Projektion $v \mapsto v + U$, so ist p_U linear mit* $\operatorname{Ker}(p_U) = U$, *und für jeden zu U komplementären Unterraum U' von V ist die Einschränkung*

$$p_U|_{U'} : U' \longrightarrow V/U$$

von p_U auf U' ein Isomorphismus.
Insbesondere hat man

$$\dim_K(V/U) = \dim_K V - \dim_K U,$$

und für jede Basis $(u'_1, \ldots, u'_r)$ von U' bilden die Nebenklassen $(u'_1 + U, \ldots, u'_r + U)$ eine Basis von V/U.

Beweis. Man rechnet sofort nach, dass p_U linear und surjektiv ist und Kern U hat. Die Dimensionsformel liefert dann die Behauptung. Alternativ kann man auch, ähnlich wie im Beweis der Dimensionsformel, aus Basen für U und für V/U eine Basis für V konstruieren (Übung). $\square$

Bemerkung. a) Im Fall möglicherweise unendlicher Dimension ist diese Gleichung sinngemäß verstanden auch richtig. (Ist $\dim(V)$ unendlich und $\dim U$ endlich, so ist $\dim(V/U)$ unendlich.)

b) Der Satz zeigt, dass der Quotientenraum (oder Faktorraum) es in gewisser Weise ermöglicht, mit allen zu U komplementären Unterräumen gleichzeitig zu arbeiten, ohne einen von ihnen wirklich anzugeben. Das macht manche Schlüsse eleganter, ohne sie aber eigentlich inhaltlich zu verändern. Wem komplementäre Unterräume sympathischer sind, der kann in der Linearen Algebra immer statt des Quotientenraums mit komplementären Unterräumen arbeiten. Anders sieht es bei den Gruppen aus: Ist G eine Gruppe und $H \subseteq G$ ein Normalteiler, so wird man im allgemeinen keine Untergruppe in G finden, die die Rolle des komplementären Unterraums in obigem Satz spielen könnte; das wird in der Vorlesung EAZ eingehender behandelt werden. Auch wenn man den Begriff des Vektorraums über einem Körper zu dem eines Moduls über einem Ring (etwa über dem Ring $\mathbb{Z}$ der ganzen Zahlen) verallgemeinert, hat man keinen Ersatz für den komplementären Unterraum.

Beispiel:

a) $V = \mathbb{R}^3$, $U = \left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \mid x \in \mathbb{R} \right\}$. V/U ist zweidimensional; zwei Vektoren in V sind genau dann in der gleichen Klasse modulo U, wenn sie sich höchstens in der x-Koordinate unterscheiden: Durch Übergang zu V/U vernachlässigt man Unterschiede, die in U liegen, man vergisst quasi die x-Koordinate des Vektors.

b) Ist $f : V \longrightarrow W$ eine lineare Abbildung von K-Vektorräumen und $U = \mathrm{Ker}(f)$, so sind zwei Vektoren $v, v' \in V$ genau dann in der gleichen Klasse modulo U, wenn $f(v) = f(v')$ gilt. Beim Übergang von V zu V/U vernachlässigt man alle Information über den Vektor v, die sich nicht aus seinem Bild $f(v)$ unter f gewinnen lässt.

Der folgende Satz liefert mit Hilfe des Begriffs Faktorraum eine weitere Version der Dimensionsformel für Kern und Bild einer linearen Abbildung:

Satz 10.6. *(*Homomorphiesatz*):*
Sei $f : V \longrightarrow W$ eine lineare Abbildung von K-Vektorräumen, $U \subseteq \mathrm{Ker}(f)$ ein Unterraum von V.

Dann gibt es genau eine lineare Abbildung $\tilde{f}_U : V/U \longrightarrow W$*, so dass* $f = \tilde{f}_U \circ p_U$ *gilt; dabei ist* p_U *die durch*

$$p_U(v) := v + U$$

definierte Projektion von V *auf* V/U*.*
Man sagt auch: Das Diagramm

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow p_U & \nearrow \tilde{f}_U & \\ V/U & & \end{array}$$

ist kommutativ *(oder kommutiert).*
Wenn man ein solches kommutatives Diagramm hat, so sagt man ferner, die Abbildung f faktorisiere *über* V/U*.*
Die Abbildung $\tilde{f}_U$ *ist genau dann injektiv, wenn* $U = \mathrm{Ker}(f)$ *ist; sie definiert dann einen Isomorphismus von* V/U *auf das Bild* $\mathrm{Im}(f)$ *von* f*, man hat also*

$$V/\mathrm{Ker}(f) \cong \mathrm{Im}(f). \tag{10.1}$$

Beweis. Man könnte das aus Satz 4.29 und Satz 10.5 durch die Isomorphie zwischen V/U und einem beliebigen zu U komplementären Unterraum folgern, für spätere Verallgemeinerungen ist aber der folgende Beweis ausbaufähiger, der auf die Benutzung des komplementären Unterraums völlig verzichtet und im Grunde genommen auch einfacher ist:
Das Bild eines Vektors $v \in V$ unter f hängt nur von seiner Klasse modulo $\mathrm{Ker}(f)$ ab, da ja für $u \in \mathrm{Ker}(f)$ offenbar $f(v+u) = f(v) + f(u) = f(v)$ gilt. Anders gesagt: Für alle Vektoren $v' \in v + \mathrm{Ker}(f)$ gilt $f(v') = f(v)$.
Da $U \subseteq \mathrm{Ker}(f)$ vorausgesetzt wurde, gilt die gleiche Aussage erst recht, wenn wir die Klasse von v modulo $\mathrm{Ker}(f)$ durch die (kleinere) Klasse von v modulo U ersetzen.
Wir können also $\tilde{f}_U$ durch

$$\tilde{f}_U(v + U) := f(v)$$

definieren, da wir uns soeben überzeugt haben, dass diese Definition nicht von der Auswahl des Repräsentanten der Nebenklasse abhängt.
Dass dieses $\tilde{f}_U$ linear ist, rechnet man schnell nach:

$$\begin{aligned} \tilde{f}_U((v_1 + U) + \lambda(v_2 + U)) &= \tilde{f}_U((v_1 + \lambda v_2) + U) \\ &= f(v_1 + \lambda v_2) \\ &= f(v_1) + \lambda f(v_2) \\ &= \tilde{f}_U(v_1 + U) + \lambda \tilde{f}_U(v_2 + U). \end{aligned}$$

Umgekehrt sieht man sofort, dass die Anforderung $f = \tilde{f}_U \circ p_U$ die Abbildung $\tilde{f}_U$ eindeutig festlegt:

Hat man eine Abbildung $g : V/U \longrightarrow W$ mit $f = g \circ p_U$, so gilt zwangsläufig

$$g(v+U) = g(p_U(v)) = f(v) = \tilde{f}_U(v+U)$$

für alle $v + U \in V/U$.
Schließlich ist $\tilde{f}_U$ genau dann injektiv, wenn der Kern dieser Abbildung gleich dem Nullvektor $\{\mathbf{0} + U\}$ des Vektorraums V/U ist. Nach Definition von $\tilde{f}_U$ ist

$$\mathrm{Ker}(\tilde{f}_U) = \{v + U \mid f(v) = \mathbf{0}\} = \{v + U \mid v \in \mathrm{Ker}(f)\},$$

das ist genau dann gleich $\{\mathbf{0} + U\}$, wenn $\mathrm{Ker}(f) \subseteq U$ gilt, was wegen der Voraussetzung $U \subseteq \mathrm{Ker}(f)$ äquivalent zu $U = \mathrm{Ker}(f)$ ist.
Da das Bild von $\tilde{f}_U$ offenbar gleich $\mathrm{Im}(f)$ ist, folgt der Rest der Behauptung. □

Bemerkung. a) Sind A, B Mengen, $f : A \longrightarrow B$ eine Abbildung, so heißt für $b \in B$ das Urbild

$$f^{-1}(b) := \{a \in A \mid f(a) = b\}$$

von b auch die Faser von b unter f (oder die Faser über b); man stellt sich quasi alle Elemente mit dem gleichen Bild b an einer Schnur aufgereiht vor, die in b befestigt ist. Ist $f : V \longrightarrow W$ lineare Abbildung von K-Vektorräumen, so ist $V/\mathrm{Ker}(f)$ die Menge der Fasern von f.

b) Der Homomorphiesatz wird häufig angewendet, wenn es bequem ist, die in f enthaltene Information in einen trivialen Anteil (Projektion auf $V/\mathrm{Ker}(f)$) und einen nichttrivialen Anteil ($\tilde{f}_U$ mit $U = \mathrm{Ker}(f)$) aufzuspalten.
$\tilde{f}_U$ heißt auch die von f *induzierte Abbildung* von V/U nach W.

c) Die Version des Homomorphiesatzes für Gruppen lautet:
Sei $f : G \longrightarrow H$ ein Homomorphismus von Gruppen, $U \subseteq \mathrm{Ker}(f)$ ein Normalteiler von G.
Dann gibt es genau einen Gruppenhomomorphismus $\tilde{f}_U : G/U \longrightarrow H$, so dass $f = \tilde{f}_U \circ p_U$ gilt; dabei ist p_U die durch

$$p_U(g) := gU$$

definierte Projektion von G auf G/U.
Man sagt auch: Das Diagramm

$$\begin{array}{ccc} G & \xrightarrow{f} & H \\ \downarrow p_U & \nearrow \tilde{f}_U & \\ G/U & & \end{array}$$

ist *kommutativ* (oder kommutiert).

Die Abbildung $\tilde{f}_U$ ist genau dann injektiv, wenn $\mathrm{Ker}(f) = U$ gilt, sie definiert dann einen Isomorphismus von G/U auf das Bild $\mathrm{Im}(f)$ von f, man hat also

$$G/\mathrm{Ker}(f) \cong \mathrm{Im}(f). \tag{10.2}$$

Der Beweis geht ganz genauso wie der oben gegebene für Vektorräume.

Das folgende Korollar ist die Version des Homomorphiesatzes für Quotientenraumvermeider.

Korollar 10.7. *Sei $f : V \longrightarrow W$ eine lineare Abbildung von K-Vektorräumen, $U \subseteq \mathrm{Ker}(f)$ ein Unterraum, $U' \subseteq V$ ein Unterraum von V mit $U \oplus U' = V$.*
Sei $p : V \longrightarrow U'$ die durch $p(u + u') = u'$ für $u \in U$, $u' \in U'$ definierte lineare Abbildung. Dann ist

$$f = f|_{U'} \circ p.$$

Beweis. Übung. Sie können diese Aussage entweder direkt beweisen oder unter Benutzung des Homomorphiesatzes und der Isomorphie zwischen V/U und U'. □

11. Bilinearformen, Dualraum und adjungierte Abbildung

Auch in diesem Abschnitt ist V immer ein Vektorraum über dem Körper K. In Abschnitt 9 haben wir Bilinearformen und hermitesche Formen definiert (Definition und Lemma 9.1, Definition 9.3) und deren Untersuchung begonnen. Wir setzen diese Untersuchung jetzt fort und verknüpfen sie mit der Theorie des Dualraums (Abschnitt 5, Definition und Korollar 5.9), insbesondere den dort eingeführten Begriffen des Annullators (Definition und Satz 5.10) und der transponierten Abbildung (Definition und Satz 5.11).
Insbesondere sei daran erinnert, dass für endlich-dimensionales V zu jeder Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V der Dualraum die dazu duale Basis $\mathcal{B}^* = (v_1^*, \ldots, v_n^*)$ mit $v_i^*(v_j) = \delta_{ij}$ hat, so dass man insbesondere nach Wahl einer Basis durch lineare Fortsetzung von $v_i \mapsto v_i^*$ einen Isomorphismus $V \to V^*$ bekommt. Dieser ist aber natürlich von der Wahl der Basis abhängig, also nicht kanonisch (kanonisch nennt man Abbildungen, die nicht von Auswahlen abhängen).
Dagegen hat man:

Satz 11.1. *Sei V ein K-Vektorraum, V^* sein Dualraum, $(V^*)^* =: V^{**}$ sein* Bidualraum *(der Dualraum von V^*).*
Für $v \in V$ sei $v^{**} : V^* \to K$ gegeben durch

$$v^{**}(f) := f(v) \text{ für } f \in V^*.$$

Dann ist $v^{**} \in V^{**}$ und die durch $\iota(v) := v^{**} \in V^{**}$ gegebene Abbildung $\iota : V \longrightarrow V^{**}$ eine injektive lineare Abbildung.
Ist V endlichdimensional, so ist ι ein (kanonischer) Isomorphismus von V auf V^{**}.

Beweis. Zunächst muss man nachrechnen, dass $\iota(v)$ linear ist:
Sind $\varphi_1, \varphi_2 \in V^*, \lambda \in K$, so ist

$$\iota(v)(\varphi_1+\lambda\varphi_2) = (\varphi_1+\lambda\varphi_2)(v) = \varphi_1(v)+\lambda\varphi_2(v) = \iota(v)(\varphi_1)+\lambda\iota(v)(\varphi_2).$$

Dann muss man nachrechnen, dass die Abbildung ι linear ist:
Sind $v_1, v_2 \in V, \lambda \in K$, so ist für alle $\varphi \in V^*$

$$\begin{aligned} \iota(v_1 + \lambda v_2)(\varphi) &= \varphi(v_1 + \lambda v_2) \\ &= \varphi(v_1) + \lambda\varphi(v_2) \text{ (weil } \varphi \text{ linear ist)} \\ &= \iota(v_1)(\varphi) + \lambda\iota(v_2)(\varphi). \end{aligned}$$

Da es zu jedem $v \neq \mathbf{0}$ aus V ein $\varphi \in V^*$ gibt mit $\varphi(v) \neq 0$ (weil v linear unabhängig ist), ist $\mathrm{Ker}(\iota) = \{\mathbf{0}\}$, also ist ι injektiv. Im endlich dimensionalen Fall ist $\dim(V) = \dim(V^*) = \dim((V^*)^*)$, also ist in diesem Fall ι sogar bijektiv.

□

Bilinarformen auf endlichdimensionalen Vektorräumen können mit Hilfe von Matrizen beschrieben werden:

Definition und Lemma 11.2. a) *Sei V ein K-Vektorraum mit Basis $\mathcal{B} = (v_1, \ldots, v_n)$, $\beta : V \times V \longrightarrow K$ eine Bilinearform.*
Sei $A = (a_{ij}) \in M_n(K) = M(n \times n, K)$ gegeben durch $a_{ij} := \beta(v_i, v_j)$.
Dann heißt A die Gram-Matrix *von β bezüglich $\mathcal{B}$, man schreibt $A = M_{\mathcal{B}}(\beta)$.*
Es gilt

$$\beta(\sum_{i=1}^{n} x_i v_i, \sum_{i=1}^{n} y_i v_i) = \beta_A(\mathbf{x}, \mathbf{y}) = \sum_{i,j=1}^{n} a_{ij} x_i y_j,$$

β ist genau dann symmetrisch, wenn A symmetrisch ist

b) *Durch $A \mapsto \beta_A$ (mit $\beta_A(\mathbf{x}, \mathbf{y}) = {}^t\mathbf{x}A\mathbf{y}$) und $\beta \mapsto M_{\mathcal{B}}(\beta)$ werden zueinander inverse Bijektionen zwischen $M_n(K)$ und der Menge* $\mathrm{Bil}(K^n)$ *der Menge der Bilinearformen auf K^n gegeben. Diese sind Isomorphismen von K-Vektorräumen, wenn man* $\mathrm{Bil}(K^n)$ *durch $(\beta+\beta')(\mathbf{x}, \mathbf{y}) := \beta(\mathbf{x}, \mathbf{y}) + \beta'(\mathbf{x}, \mathbf{y})$, $(\lambda\beta)(\mathbf{x}, \mathbf{y}) := \lambda \cdot \beta(\mathbf{x}, \mathbf{y})$ zu einem K-Vektorraum macht.*

Beweis. Nachrechnen! □

Beispiel: Das *Standardskalarprodukt* auf $\mathbb{R}^n$ ist definiert durch

$$\langle \mathbf{x}, \mathbf{y} \rangle := {}^t\mathbf{x}\mathbf{y} = {}^t\mathbf{y}\mathbf{x} = \sum_{i=1}^{n} x_i y_i.$$

Es gilt für $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$:

$$\begin{aligned} \langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} \rangle &= \langle \mathbf{x}_1, \mathbf{y} \rangle + \langle \mathbf{x}_2, \mathbf{y} \rangle \\ \langle \mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2 \rangle &= \langle \mathbf{x}, \mathbf{y}_1 \rangle + \langle \mathbf{x}, \mathbf{y}_2 \rangle \\ \langle \lambda\mathbf{x}, \mathbf{y} \rangle &= \lambda\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \lambda\mathbf{y} \rangle \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \mathbf{y}, \mathbf{x} \rangle \end{aligned}$$

Das *Standardskalarprodukt* ist also eine symmetrische Bilinearform auf $\mathbb{R}^n$.
Seine Gram-Matrix bezüglich der Standardbasis ist die Einheitsmatrix E_n.
Die genauso definierte symmetrische Bilinearform auf K^n für einen beliebigen Grundkörper K heißt die Einheitsform.
Ist $U \subseteq \mathbb{R}^n$ ein Unterraum mit Basis $\mathcal{B} = (v_1, \ldots, v_r)$, so ist die Einschränkung des Standardskalarprodukts auf U natürlich ebenfalls eine symmetrische Bilinearform.
Ihre Gram-Matrix A bezüglich $\mathcal{B}$ ist $A = (\langle v_i, v_j \rangle)$. Ist $T \in M(n \times r, \mathbb{R})$ die Matrix mit den Spalten $v_1, \ldots, v_r$, so ist $A = {}^tTT$.

Lemma 11.3. *Sei V ein K-Vektorraum mit Basen $\mathcal{B} = (v_1, \ldots, v_n), \mathcal{B}' = (v'_1, \ldots, v'_n)$, $\beta : V \times V \longrightarrow K$ eine Bilinearform.*
Seien $A, A' \in M_n(K)$ die Gram-Matrizen von β bezüglich der Basen $\mathcal{B}, \mathcal{B}'$, sei $T \in GL_n(K)$ die Übergangsmatrix von der Basis $\mathcal{B}$ zur Basis $\mathcal{B}'$, also $v'_j = \sum_{i=1}^{n} t_{ij} v_i$ für $1 \leq j \leq n$.

Dann gilt

$$A' = {}^tTAT.$$

Ist $\mathcal{B} = (v_1, \ldots, v_n)$ *eine Basis von* K^n *und* $T \in GL_n(K)$ *die Matrix mit den Spalten* $v_1, \ldots, v_n$*, so hat die Einheitsform bezüglich* $\mathcal{B}$ *die Matrix* ${}^tTT$*.*

Beweis. Nachrechnen! □

Definition und Lemma 11.4. *Sei* V *ein endlichdimensionaler* K*-Vektorraum mit Basis* $\mathcal{B} = (v_1, \ldots, v_n)$ *und* $\beta : V \times V \longrightarrow K$ *eine Bilinearform.*

a) *Durch*

$$\tilde{\beta}_1(v)(w) := \beta(w, v), \quad \tilde{\beta}_2(v)(w) := \beta(v, w)$$

werden lineare Abbildungen $\tilde{\beta}_1, \tilde{\beta}_2 : V \longrightarrow V^*$ *definiert.*

b) *Ist* $f : V \to V^*$ *eine lineare Abbildung, so wird durch* $(v, w) \mapsto (f(v))(w) \in K$ *eine Bilinearform* α *auf* V *definiert, für die* $\tilde{\alpha}_2 = f$ *gilt.*

c) *Hat* β *bezüglich der Basis* $\mathcal{B}$ *die Gram-Matrix* A*, so ist*

$$M^{\mathcal{B}}_{\mathcal{B}^*}(\tilde{\beta}_1) = A, \quad M^{\mathcal{B}}_{\mathcal{B}^*}(\tilde{\beta}_2) = {}^tA.$$

d) *Das Diagramm*

$$\begin{tikzcd}
\mathrm{Bil}(V) \arrow[rr, "\beta \mapsto \tilde{\beta}_1"] \arrow[dr, "\beta \mapsto M_{\mathcal{B}}(\beta)"'] & & \mathrm{Hom}(V, V^*) \arrow[dl, "f \mapsto M^{\mathcal{B}}_{\mathcal{B}^*}(f)"] \\
& M_n(K) &
\end{tikzcd}$$

ist kommutativ und alle Abbildungen in diesem Diagramm sind Isomorphismen.

Beweis. Nachrechnen. □

Bemerkung. Ist β symmetrisch und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V bezüglich β (also $\beta(v_i, v_j) = \beta(v_j, v_i) = \delta_{ij}$) und $\mathcal{B}^* = (v_1^*, \ldots, v_n^*)$ die dazu duale Basis des Dualraums V^*, so gilt für die Abbildung $\tilde{\beta} = \tilde{\beta}_1 = \tilde{\beta}_2 : V \longrightarrow V^*$:

$$\tilde{\beta}(v_j) = v_j^* \quad (1 \leq j \leq n).$$

Lemma 11.5. *Sei* V *ein endlichdimensionaler* K*-Vektorraum und* $\beta : V \times V \longrightarrow K$ *eine Bilinearform.*
Die folgenden Aussagen sind äquivalent:

a) *Ist* $v \in V$ *mit* $\beta(w, v) = 0$ *für alle* $w \in V$*, so ist* $v = \mathbf{0}$*.*

b) $\tilde{\beta}_1$ *ist injektiv.*

c) $\tilde{\beta}_1$ *ist surjektiv.*

d) *Zu jedem* $\varphi \in V^*$ *gibt es* $v \in V$ *mit* $\beta(w, v) = \varphi(w)$ *für alle* $w \in V$*.*

e) *Ist* $v \in V$ *mit* $\beta(v,w) = 0$ *für alle* $w \in V$*, so ist* $v = \mathbf{0}$.
f) $\tilde{\beta}_2$ *ist injektiv.*
g) $\tilde{\beta}_2$ *ist surjektiv.*
h) *Zu jedem* $\varphi \in V^*$ *gibt es* $v \in V$ *mit* $\beta(v,w) = \varphi(w)$ *für alle* $w \in V$.
i) *Die Gram-Matrix von* β *bezüglich einer beliebigen Basis* $\mathcal{B}$ *von* V *ist invertierbar.*

Beweis. Dass a) und b) sowie c) und d) äquivalent sind, ist nach Definition von $\tilde{\beta}_1$ klar, und b) und c) sind äquivalent, weil $\dim(V) = \dim(V^*)$ gilt. Ebenso sind e), f), g), h) zueinander äquivalent.
Gilt a) (und damit b), c) ,d)) und ist $v \in V, v \neq \mathbf{0}$, so gibt es ein $\varphi \in V^*$ mit $\varphi(v) \neq 0$, wegen c) gibt es also ein $w \in V$ mit $\tilde{\beta}_1(w) = \varphi$, also $\beta(v,w) = \varphi(v) \neq 0$, also gilt auch e) und damit f), g) und h). Genauso folgen a),b),c),d) aus e).
Die Äquivalenz von i) mit b), c) folgt aus dem vorigen Lemma, Teil c). □

Wir wollen die Bilinearformen, die den Bedingungen des vorigen Lemmas genügen, näher untersuchen:

Definition 11.6. *Sei* V *ein* K*-Vektorraum. Eine Bilinearform* $\beta : V \times V \longrightarrow K$ *heißt* nichtausgeartet *oder* regulär*, wenn gilt:*
Ist $v \in V$ *und* $\beta(v,w) = 0$ *für alle* $w \in V$*, so ist* $v = \mathbf{0}$.

Bemerkung. Nach dem vorigen Lemma kann man in obiger Definition auch die Rollen des ersten und des zweiten Arguments in β vertauschen, die Definition ist also nicht so unsymmetrisch, wie sie auf den ersten Blick aussieht.

Beispiel:

a) Die Einheitsform β_0 auf K^n mit $\beta_0(\mathbf{x},\mathbf{y}) = {}^t\mathbf{x}\mathbf{y}$ ist nichtausgeartet, da $\beta_0(\mathbf{x}, \mathrm{e}_j) = x_j$ für $1 \leq j \leq n$ gilt.
b) Die symmetrische Bilinearform
$$\beta(\mathbf{x},\mathbf{y}) = x_1y_1 - x_2y_2 \text{ auf } K^2$$
ist nichtausgeartet.
c) Die symmetrische Bilinearform $\beta(\mathbf{x},\mathbf{y}) = x_1y_3 - x_2y_3 + x_3y_1 - x_3y_2$ auf $\mathbb{R}^3$ ist ausgeartet: Man hat $\beta(\left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right), \mathbf{y}) = 0$ für alle $\mathbf{y} \in \mathbb{R}^3$.

Im Weiteren werden wir uns meist auf die Untersuchung symmetrischer oder alternierender Bilinearformen beschränken, weil diese in Anwendungen am häufigsten vorkommen. Zudem überlegt man sich leicht, dass man für $\mathrm{char}(K) \neq 2$ jede Bilinearform eindeutig als Summe einer symmetrischen und einer alternierenden Bilinearform schreiben kann (Übung).

Definition 11.7. *Sei β eine symmetrische oder alternierende Bilinearform auf dem K-Vektorraum V.*
Für einen Unterraum $U \subseteq V$ ist das orthogonale Komplement $U^\perp$ *von U (bezüglich β) durch*

$$U^\perp := \{v \in V \mid \beta(u,v) = 0 \text{ für alle } u \in U\}$$

definiert.
Das Radikal $\operatorname{rad}(V,\beta) = \operatorname{rad}_\beta(V)$ *von (V,β) ist definiert durch*

$$\begin{aligned} \operatorname{rad}(V,\beta) &= \{v \in V \mid \beta(v,w) = 0 \text{ für alle } w \in V\} \\ &= \{v \in V \mid \beta(w,v) = 0 \text{ für alle } w \in V\} \\ &= V^\perp. \end{aligned}$$

Beispiel: Die symmetrische Bilinearform $\beta(\mathbf{x},\mathbf{y}) = x_1y_3 - x_2y_3 + x_3y_1 - x_3y_2$ auf $\mathbb{R}^3$ hat als Radikal den vom Vektor $\left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right)$ erzeugten Unterraum des $\mathbb{R}^3$.

Wie bei der Betrachtung des Zusammenhangs zwischen Matrizen und linearen Abbildungen stellt sich auch bei der Beschreibung von Bilinearformen durch Matrizen die Frage, ob und wie man zu einer gegebenen Bilinearform eine Basis des zu Grunde liegenden Vektorraums findet, bezüglich der die Matrix der Bilinearform eine besonders einfache Gestalt hat.
Wir haben zunächst:

Lemma 11.8. *Sei β eine symmetrische oder alternierende Bilinearform auf dem K-Vektorraum V.*
Ist U ein zu $\operatorname{rad}(V,\beta)$ *komplementärer Unterraum von V, so ist*

$$\beta|_{U\times U} : U \times U \longrightarrow K$$

eine nichtausgeartete Bilinearform.
Sind $(v_1,\ldots,v_r)$ eine Basis von $\operatorname{rad}(V,\beta)$ *und $(v_{r+1},\ldots,v_n)$ eine Basis von U, so ist $(v_1,\ldots,v_n)$ eine Basis von V, bezüglich der β eine Gram-Matrix der Gestalt*

$$\begin{pmatrix} 0_r & 0_{r,n-r} \\ 0_{n-r,r} & A \end{pmatrix}$$

mit einer invertierbaren Matrix $A \in M_{n-r}(K)$ hat.

Beweis. Klar. □

Wir können uns also im Weiteren auf die Untersuchung nichtausgearteter symmetrischer bzw. alternierender Bilinearformen beschränken.

Lemma 11.9. *Sei V ein endlich-dimensionaler K-Vektorraum mit nichtausgearteter symmetrischer oder alternierender Bilinearform β und $U \subseteq V$ ein Unterraum.*
Dann ist $\dim(U) + \dim(U^\perp) = \dim(V)$.

Beweis. Durch den Isomorphismus $\tilde{\beta}_1$ wird $U^\perp$ bijektiv auf den Annullator $\mathrm{Ann}(U) = \{\varphi \in V^* \mid \varphi|_U \equiv 0\} \subseteq V^*$ abgebildet. Nach Definition und Satz 5.10 folgt die Behauptung.
Alternativ können wir auch $\tilde{\beta}_1^{(U)} : V \to U^*$ durch $\tilde{\beta}_1^{(U)}(v) = \beta_1(v)|_U$ definieren und beobachten, dass $\tilde{\beta}_1^{(U)}$ surjektiv mit Kern $U^\perp$ ist, die Behauptung folgt dann aus Satz 4.29. □

Korollar 11.10. *In der Situation des vorigen Lemmas ist* $(U^\perp)^\perp = U$.

Beweis. Offenbar ist $(U^\perp)^\perp \supseteq U$. Aus Dimensionsgründen folgt die Behauptung. □

Lemma 11.11. *Sei V ein K-Vektorraum, $\beta : V \times V \longrightarrow K$ eine symmetrische oder alternierende Bilinearform, $U \subseteq V$ ein endlichdimensionaler Unterraum, für den $\beta|_{U\times U}$ nichtausgeartet ist.*

Dann ist $V = U \oplus U^\perp$.

Beweis. Da $\beta|_{U\times U}$ nichtausgeartet ist, ist $U \cap U^\perp = \{\mathbf{0}\}$, die Räume $U, U^\perp$ bilden also eine direkte Summe.
Ist $v \in V$, so wird durch $u \mapsto \beta(v, u)$ eine Linearform $\varphi \in U^*$ definiert, und weil $\beta|_{U\times U}$ nichtausgeartet ist, gibt es $u' \in U$ mit $\beta(u', u) = \beta(v, u)$ für alle $u \in U$. Also ist $u'' := v - u' \in U^\perp$ und daher $v = u' + u'' \in U + U^\perp$, es gilt also auch $V = U + U^\perp$. □

Bemerkung. Besitzt der Unterraum U im vorigen Lemma speziell eine Orthogonalbasis $(u_1, \ldots, u_r)$ aus Vektoren u_j mit $\beta(u_j, u_j) \neq 0$ $(1 \leq j \leq r)$, so lassen sich die Projektionen $u' \in U$, $u'' \in U^\perp$ von $v \in V$ auf $U, U^\perp$ als

$$\begin{aligned} u' &= \textstyle\sum_{j=1}^r \frac{\beta(v,u_j)}{\beta(u_j,u_j)} u_j, \\ u'' &= v - v_1 \end{aligned}$$

berechnen (Übung).

Beispiel: Ist $V = \mathbb{F}_2^n$, so heißen Unterräume $C \subseteq \mathbb{F}_2^n$ auch lineare *Codes*, da sie benutzt werden, um Daten für die Speicherung oder für die Übermittlung von Nachrichten zu codieren. Man übermittelt Nachrichten, indem man die einzelnen Zeichen oder Wörter der Nachricht in Bitfolgen der Länge n (also Elemente von $\mathbb{F}_2^n$) verwandelt (codiert) und dafür nur Folgen (Codeworte) in C zulässt. Für $\mathbf{x} \in V$ nennt man $w(\mathbf{x}) = \#\{1 \leq j \leq n \mid x_j \neq 0\}$ das *(Hamming-) Gewicht* von $\mathbf{x}$.
Ist das *Minimalgewicht* $w(C) = \min\{w(\mathbf{x}) \mid 0 \neq \mathbf{x} \in C\}$ gleich d, so unterscheiden sich je zwei Elemente von C in wenigstens d Stellen, d heißt deshalb auch der *Minimalabstand* von C.
Hat C den Minimalabstand $2t + 1$, so kann der Code t bei der Nachrichtenübermittlung entstandene Fehler korrigieren.
Empfängt man $\mathbf{y} \in \mathbb{F}_2^n$, so decodiert man es als dasjenige (eindeutig bestimmte) $\mathbf{x} \in C$ mit $w(\mathbf{x} - \mathbf{y}) \leq t$ (sofern es ein solches gibt).
Sind bei der Übermittlung nicht mehr als t Fehler aufgetreten, so erhält

man auf diese Weise korrekt den gesendeten Vektor aus C zurück.
Man definiert hier den zu C dualen Code als das Orthokomplement $C^\perp$ bezüglich der Standardbilinearform $\beta(\mathbf{x},\mathbf{y}) = \sum_{j=1}^n x_j y_j$. Die Eigenschaften von C und $C^\perp$ hängen zusammen; besonders interessiert man sich für selbstduale Codes, also C mit $C = C^\perp$, also gerade solche Teilräume C, für die $\beta|_C$ identisch verschwindet (und die maximal mit dieser Eigenschaft sind).

Die schwächere Eigenschaft $\beta(\mathbf{x},\mathbf{x}) = 0$ für alle $\mathbf{x} \in C$ erreicht man, idem man zum erweiterten Code $\tilde{C} \subseteq \mathbb{F}_2^{n+1}$ übergeht:

$$\tilde{C} := \{\mathbf{x} \in \mathbb{F}_2^{n+1} \mid \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in C,\ x_{n+1} = \sum_{j=1}^n x_j\}.$$

Es ist in der Codierungstheorie üblich, die Transponierten der Vektoren einer Basis in die Zeilen einer Matrix, der sogenannten *Erzeugermatrix* G_C (*generator matrix*) einzutragen, die Matrix $G_{C^\perp}$ heißt dann *Kontrollmatrix* und hat die Eigenschaft

$$C = \{\mathbf{x} \in \mathbb{F}_2^n \mid G_{C^\perp} \cdot \mathbf{x} = \mathbf{0}\},$$

für selbstduale Codes stimmen also Erzeugermatrix und Kontrollmatrix überein.

Ein Beispiel ist der *Hammingcode* $C_H \subseteq \mathbb{F}_2^7$ mit Kontrollmatrix

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

und Erzeugermatrix

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

Der Code hat Minimalgewicht 3. Der erweiterte Code ist selbstdual und hat Minimalgewicht 4.

Satz 11.12. (Gram-Schmidt, verallgemeinert) *Sei K ein Körper mit* $\mathrm{char}(K) \neq 2$, *V ein endlichdimensionaler K-Vektorraum, $\beta : V \times V \longrightarrow K$ eine symmetrische Bilinearform auf V.*
Dann hat V eine Orthogonalbasis bezüglich β, also eine Basis $(v_1, \ldots, v_n)$ mit $\beta(v_i, v_j) = 0$ für $i \neq j$.
Ist β nicht ausgeartet, so ist dabei $\beta(v_i, v_i) \neq 0$ für alle i.

Beweis. Wir beweisen das durch Induktion nach $n = \dim(V)$, der Anfang $\dim(V) = 1$ ist trivial. Wir betrachten also $n > 1$ und nehmen an, die Behauptung sei für Räume kleinerer Dimension als n bewiesen. Iat

β identisch 0, so ist die Aussage trivial, andernfalls gibt es $v \in V$ mit $\beta(v,v) \neq 0$, denn sonst wäre wegen der Polarisierungsformel

$$2\beta(v,w) = \beta(v+w, v+w) - \beta(v,v) - \beta(w,w)$$

die Bilinearform β identisch gleich 0 (an dieser Stelle benötigen wir die Voraussetzung $\mathrm{char}(K) \neq 2$).
Der von v erzeugte Unterraum $U \subseteq V$ liefert wegen Lemma 11.11 eine Zerlegung $V = U \oplus U^\perp$ in eine orthogonale direkte Summe. Da $U^\perp$ Dimension $n-1$ hat, besitzt U nach Induktionsannahme eine Orthogonalbasis; ergänzt man diese um den Vektor v, so erhält man die gesuchte Orthogonalbasis von V. $\square$

Bemerkung. Man beachte, dass dieser eventuell etwas unkonstruktiv wirkende Beweis in Wahrheit nichts anderes als das Gram-Schmidt'sche Orthogonalisierungsverfahren aus Abschnitt 9 ist - jedenfalls dann, wenn man beim Versuch, es auf eine vorgegebene Basis von V anzuwenden, nie einen Vektor u mit $\beta(u,u) = 0$ erhält. Wenn das passiert, muss man das Verfahren unterbrechen und den Vektor zunächst durch einen besser geeigneten ersetzen. Auch dies lässt sich leicht algorithmisch formulieren, man büßt aber die Dreiecksgestalt der Matrix der Basistransformation ein.

Da man für den Grundkörper $K = \mathbb{R}$ durch Übergang von v_i zu $\sqrt{|\beta(v_i,v_i)|}^{-1} v_i$ stets $\beta(v_i,v_i) = \pm 1$ erreichen kann, haben wir einen neuen Beweis des Trägheitssatzes von Sylvester (Definition und Satz 9.18):

Satz 11.13. (Trägheitssatz von Sylvester) *Sei V ein endlichdimensionaler $\mathbb{R}$-Vektorraum der Dimension n, $\beta : V \times V \longrightarrow \mathbb{R}$ eine nichtausgeartete symmetrische Bilinearform.*
Dann gibt es eine Basis von V, bezüglich der die Matrix von β Diagonalgestalt mit Einträgen ± 1 hat. Dabei sind die Anzahlen p der Einträge $+1$ und q der Einträge -1 von der Auswahl der Basis unabhängig. Das Paar (p,q) (oder gelegentlich auch die Zahl $p-q$) heißt die Signatur *von β.*

Bemerkung. Alles bisherige kann mit leichten Modifikationen auch für hermitesche Formen auf einem komplexen Vektorraum durchgeführt werden. Allerdings ist die Abbildung $\tilde{\beta}_1 : V \to V^*$ dann nicht linear, sondern *semilinear*, d.h., für $v \in V$ und $a \in \mathbb{C}$ gilt $\tilde{\beta}_1(av) = \bar{a}\tilde{\beta}_1(v)$, und $\tilde{\beta}_2$ bekommt Werte im Raum der semilinearen Abbildungen von V nach $\mathbb{C}$ (ist dafür aber linear). Man kann sich aber überzeugen (Übung), dass das nichts an der Richtigkeit der (entsprechend modifizierten) oben benutzten Argumente ändert. Insbesondere gilt der neue Beweis des Trägheitssatzes von Sylvester ebenfalls für hermitesche Formen.

Korollar 11.14. *Sei $A \in GL_n(\mathbb{C})$ eine reguläre hermitesche Matrix.*

Dann gibt es $T \in GL_n(\mathbb{C})$, so dass ${}^tTA\bar{T}$ eine Diagonalmatrix mit Einträgen ± 1 ist. Ist A reell symmetrisch, so kann auch $T \in GL_n(\mathbb{R})$ gewählt werden.

Satz 11.15. (Determinantenkriterium von Jacobi) *Sei $A \in M_n(\mathbb{R})$ symmetrisch, für $1 \leq k \leq n$ sei $A_k \in M_k(\mathbb{R})$ die aus den ersten k Zeilen und Spalten von A gebildete $(k \times k)$-Matrix, $d_k := \det(A_k)$.*
Dann gilt: A ist genau dann positiv definit, wenn $d_k > 0$ für $1 \leq k \leq n$ gilt.

Beweis. Die Behauptung können wir auch wie folgt ausdrücken: Die durch A gegebene Bilinearform $\beta = \beta_A$ ist genau dann positiv definit, wenn für $1 \leq k \leq n$ ihre Einschränkung auf den von den ersten k Vektoren der Standardbasis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ erzeugten Unterraum U_k bezüglich dieser Basisvektoren eine Gram-Matrix A_k mit positiver Determinante d_k hat.
Ist β positiv definit, so ist auch die Einschränkung von β auf U_k positiv definit, und nach dem Satz von Gram-Schmidt kann man die Matrix A_k für jedes k als $A_k = {}^tT_k \cdot T_k$ mit $T \in GL_k(\mathbb{R})$ schreiben, hat also $d_k = (\det(T_k))^2 > 0$ für $1 \leq k \leq n$.
Sind umgekehrt alle d_k positiv, so können wir eine leicht umformulierte Version des Gram-Schmidt-Verfahrens anwenden, um aus der Standardbasis eine Orthogonalbasis $(v_1, \ldots, v_n)$ von $\mathbb{R}^n$ bezüglich β zu konstruieren, für die $v_1, \ldots, v_k$ für $1 \leq k \leq n$ den Raum U_k erzeugen und für die $\beta(v_j, v_j) > 0$ für $1 \leq j \leq n$ gilt; das impliziert offenbar, dass β und damit A positiv definit ist.
Zunächst ist $v_1 = \mathbf{e}_1$ mit $\beta(\mathbf{e}_1, \mathbf{e}_1) = d_1 > 0$. Hat man für $k > 1$ bereits paarweise orthogonale Vektoren $v_1, \ldots, v_{k-1}$ mit den gewünschten Eigenschaften konstruiert, so hat U_k nach Lemma 11.11 eine Zerlegung $U_k = U_{k-1} \oplus U_{k-1}^{\perp}$ mit $\dim(U_{k-1}^{\perp}) = 1$. Wir wählen dann v_k als einen Vektor, der $U_{k-1}^{\perp}$ erzeugt. Damit sind die v_j offenbar paarweise orthogonal und so, dass $v_1, \ldots, v_k$ für $1 \leq k \leq n$ den Raum U_k erzeugen.
Die Determinante $d'_k = \prod_{j=1}^{k} \beta(v_j, v_j)$ der Gram-Matrix von $\beta|_{U_k \times U_k}$ bezüglich $(v_1, \ldots, v_k)$ unterscheidet sich von d_k nur um ein von 0 verschiedenes Quadrat, ist also für alle k ebenfalls positiv. Also sind alle $\beta(v_j, v_j)$ in der Tat positiv, und die Behauptung ist bewiesen. $\square$

Bemerkung. Die Aussage des Korollars gilt auch für komplexe hermitesche Matrizen (man beachte, dass dann alle d_k reell sind).

Für alternierende Bilinearformen ist die Situation sogar noch einfacher.

Satz 11.16. *Sei V endlichdimensional und $\beta : V \times V \longrightarrow K$ eine nichtausgeartete alternierende Bilinearform (also $\beta(v, v) = 0$ für alle $v \in V$, das impliziert, dass β schiefsymmetrisch ist und ist für*

$\mathrm{char}(K) \neq 2$ *äquivalent dazu), V endlichdimensional. Dann hat V eine Basis, bezüglich der β die Matrix*

$$\begin{pmatrix} J & & & 0 \\ & J & & \\ & & \ddots & \\ 0 & & & J \end{pmatrix} \text{ mit } J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

hat.
Insbesondere ist $\dim(V) = 2m$ *gerade.*

In Matrixformulierung: Sei $A \in M(n \times n, K)$ eine alternierende Matrix (also ${}^t\mathbf{x}A\mathbf{x} = 0$ für alle $\mathbf{x} \in K^n$) mit $\det(A) \neq 0$.
Dann gibt es $T \in \mathrm{GL}_n(K)$ mit

$${}^tTAT = \begin{pmatrix} J & & 0 \\ & \ddots & \\ 0 & & J \end{pmatrix}.$$

Beweis. Sei $v \neq \mathbf{0}$ ein beliebiger Vektor in V. Weil β als nichtausgeartet vorausgesetzt ist, gibt es $w \in V$ mit $\beta(v, w) = 1$. Ist $U = \mathrm{Lin}(v, w)$, so hat $\beta|_{U \times U}$ bezüglich der Basis (v, w) die Matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Nach Lemma 11.11 können wir U orthogonal abspalten (also $V = U \oplus U^\perp$ schreiben) und sehen, dass die Behauptung durch vollständige Induktion nach $\dim(V)$ folgt. □

Bemerkung. a) Der Beweis des Satzes kann so geführt werden, dass man einen Algorithmus erhält, vermittels dessen A durch simultane Zeilen- und Spaltenumformungen in die Gestalt

$$\begin{pmatrix} J & & 0 \\ & \ddots & \\ 0 & & J \end{pmatrix}$$

gebracht wird.

b) Ist A alternierend und T wie oben, so ist $\det(A) = (\det(T))^2$, insbesondere ist $\det(A)$ ein Quadrat in K.

c) $\det(A)$ ist (nach der Formel von Leibniz) ein Polynom F_1 in den Einträgen der Matrix A, und zwar ein homogenes Polynom vom Grad $n = 2m$ (ein Polynom
$F(X_1, \ldots, X_r) = \sum_{j_1=0}^{d} \sum_{j_2=0}^{d} \cdots \sum_{j_r=0}^{d} a_{j_1,\ldots,j_n} X_1^{j_1} \cdots X_r^{j_r}$
heißt *homogen* vom Grad n, wenn nur Ausdrücke $X^{j_1} \cdots X^{j_r}$ mit $j_1 + \cdots + j_r = n$ darin vorkommen).

Will man die Determinante nur auf alternierende Matrizen anwenden, so kann man die Variablen X_{ii} gleich 0 setzen und für X_{ij} mit $i > j$ die Einsetzung $X_{ij} = -X_{ji}$ vornehmen, man erhält ein Polynom F in den Variablen $X_{12}, \ldots, X_{1n}, X_{23}, \ldots, X_{n-1,n}$ mit $F(a_{12}, \ldots, a_{1n}, a_{23}, \ldots, a_{n-1,n}) = \det(A)$ für jede alternierende Matrix $A \in M_n(K)$ (und für beliebigen Körper K).

Man kann dann zeigen: Es gibt ein Polynom P in den Koeffizienten a_{ij} mit $i < j$ von A mit $P(X_{12}, \ldots, X_{1n}, X_{23}, \ldots, X_{n-1,n})^2 = F(X_{12}, \ldots, X_{n-1,n})$; das Polynom P ist homogen vom Grad n, es heißt die *Pfaff'sche Form.*

Wir haben in Abschnitt 9 in der Situation eines euklidischen oder unitären Vektorraumes zu einem Endomorphismus des Vektorraums die adjungierte Abbildung betrachtet. Jetzt wollen wir zusammenstellen, was man zu diesem Begriff in der allgemeinen Situation eines Vektorraums mit symmetrischer Bilinearform aussagen kann und wie der Zusammenhang dieses Begriffs mit dem Dualraum ist.
Wir erinnern zunächst an Definition und Satz 5.11: Sind V, W Vektorräume über dem Körper K, $f : V \longrightarrow W$ eine lineare Abbildung, so wird die transponierte Abbildung ${}^t f =: f^* : W^* \longrightarrow V^*$ durch

$$ {}^t f(\psi) := \psi \circ f \quad (\psi \in W^*) $$

eine lineare Abbildung definiert. Sind $\mathcal{B}_1, \mathcal{B}_2$ von V bzw. Wund $A = M_{\mathcal{B}_2}^{\mathcal{B}_1}(f)$ die Matrix von f bezüglich der Basen $\mathcal{B}_1$ von V, $\mathcal{B}_2$ von W, so ist ${}^t A = M_{\mathcal{B}_1^*}^{\mathcal{B}_2^*}(f^t)$, wo $\mathcal{B}_1^*$ und $\mathcal{B}_2^*$ die zu $\mathcal{B}_1$ bzw. $\mathcal{B}_2$ dualen Basen von V^* bzw. W^* sind.

Satz 11.17. *Sei V ein endlichdimensionaler K-Vektorraum, $\beta : V \times V \longrightarrow K$ eine (beliebige) nichtausgeartete Bilinearform. Sei $f \in \mathrm{End}(V)$.*

Dann gibt es genau einen Endomorphismus $f^{\mathrm{ad}} \in \mathrm{End}(V)$ mit

$$ \beta(f(v), w) = \beta(v, f^{\mathrm{ad}}(w)) \text{ für alle } v, w \in V. $$

f^{ad} ist die (eindeutig bestimmte) lineare Abbildung, die das Diagramm

$$ \begin{array}{ccc} V & \xrightarrow{f^{\mathrm{ad}}} & V \\ \tilde{\beta}_1 \downarrow & & \downarrow \tilde{\beta}_1 \\ V^* & \xrightarrow{{}^t f} & V^* \end{array} $$

kommutativ macht, f^{ad} heißt die bezüglich β zu f adjungierte Abbildung, man bezeichnet sie häufig auch mit f^*.

Beweis. Wir haben

$$ \beta(f(v), w) = \tilde{\beta}_1(w)(f(v)) = (({}^t f \circ \tilde{\beta}_1)(w))(v) $$

und

$$ \beta(v, f^{\mathrm{ad}}(w)) = \tilde{\beta}_1(f^{\mathrm{ad}}(w))(v) = ((\tilde{\beta}_1 \circ f^{\mathrm{ad}})(w))(v) $$

für alle $v, w \in V$. Die Bedingung

$$ \beta(f(v), w) = \beta(v, f^{\mathrm{ad}}(w)) \text{ für alle } v, w \in V. $$

ist also äquivalent zu $\tilde{\beta}_1 \circ f^{\mathrm{ad}} = {}^t f \circ \tilde{\beta}_1$, also zu
$f^{\mathrm{ad}} = (\tilde{\beta}_1)^{-1} \circ {}^t f \circ \tilde{\beta}_1$. Man definiert nun f^{ad} durch diese Gleichung und hat die Behauptung gezeigt. $\square$

Korollar 11.18. *Sei V ein endlichdimensionaler K-Vektorraum, $\beta : V \times V \longrightarrow K$ eine nichtausgeartete symmetrische Bilinearform und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V bezüglich β. Sei $f \in \mathrm{End}(V)$ mit Matrix A bezüglich $\mathcal{B}$.*
Dann hat die bezüglich β zu f adjungierte Abbildung f^{ad} bezüglich der Basis $\mathcal{B}$ die Matrix ${}^tA$.

Beweis. Ohne Einschränkung ist $V = K^n$, $\mathcal{B}$ die Standardbasis und β die Einheitsform, also $\beta(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i$ für $\mathbf{x}, \mathbf{y} \in K^n$, und $f = L_A$. Wir haben dann für alle $\mathbf{x}, \mathbf{y} \in K^n$

$$\beta(A\mathbf{x}, \mathbf{y}) = {}^t(A\mathbf{x})\mathbf{y} = \mathbf{x}({}^tA\mathbf{y}),$$

mit $g := L_{{}^tA}$ ist also $\beta(f(\mathbf{x}), \mathbf{y}) = \beta(\mathbf{x}, g(\mathbf{y}))$ für alle $\mathbf{x}, \mathbf{y} \in K^n$, also $g = f^{\mathrm{ad}}$.
Alternativ können wir das auch in Matrizen ausschreiben: Wir haben für $1 \leq j, k \leq n$:

$$\begin{aligned}
\beta(A\mathbf{e}_j, \mathbf{e}_k) &= \sum_{i=1}^n a_{ij}\beta(\mathbf{e}_i, \mathbf{e}_k) \\
&= \sum_{i=1}^n a_{ij}\delta_{ik} \\
&= a_{kj} \\
&= \sum_{i=1}^n a_{ki}\delta_{ij} \\
&= \sum_{i=1}^n a_{ki}\beta(\mathbf{e}_j, \mathbf{e}_i) \\
&= \beta(\mathbf{e}_j, {}^tA\mathbf{e}_k) \\
&= \beta(\mathbf{e}_j, g(\mathbf{e}_k)).
\end{aligned}$$

Wegen der Bilinearität von β folgt daraus, dass $\beta(L_A(\mathbf{x}), \mathbf{y}) = \beta(\mathbf{x}, L_{{}^tA}\mathbf{y})$ für alle $\mathbf{x}, \mathbf{y} \in K^n$ gilt, dass also wie behauptet $L_{{}^tA}$ die zu L_A adjungierte Abbildung ist. □

Zum Abschluss dieses Paragraphen stellen wir noch einige (in der Vorlesung nicht behandelte) Dinge zusammen, die den Zusammenhang zwischen Dualraum und Bilinearformen weiter untersuchen und ausnutzen. Spezialfälle hiervon für den Fall reeller oder komplexer Vektorräume mit Skalarprodukt werden wir später noch gesondert (und vereinfacht) betrachten.

Definition und Lemma 11.19. *Seien U, V Vektorräume über K.*

a) *Ist* $\beta : U \times V \longrightarrow K$ *eine Bilinearform,* $M \subseteq U$, $N \subseteq V$, *so sind die orthogonalen Komplemente von* M, N *bezüglich* β *definiert durch:*

$$\begin{aligned} M^{\perp} &= \{v \in V \mid \beta(u,v) = 0 \text{ für alle } u \in M\} \\ {}^{\perp}N &= \{u \in U \mid \beta(u,v) = 0 \text{ für alle } v \in N\}. \end{aligned}$$

b) *Ist* $M \subseteq V$, $F \subseteq V^*$, *so ist der* Annullator *von* M *bzw.* F *definiert durch:*

$$\begin{aligned} \mathrm{Ann}(M) = M^0 &:= \{f \in V^* \mid f(v) = 0 \text{ für alle } v \in M\} \\ \mathrm{Ann}(F) = F^0 &:= \{v \in V \mid f(v) = 0 \text{ für alle } f \in F\}. \end{aligned}$$

Es gilt: Die Mengen $M^{\perp}, {}^{\perp}N, M^0, F^0$ *sind Unterräume des jeweiligen Vektorraums.*

Lemma 11.20. *Seien* U, V, β *wie bisher. Es gilt für* $M \subseteq U$*:*

a) $M_1 \subseteq M_2 \Rightarrow M_1^{\perp} \supseteq M_2^{\perp}$, $M_1^0 \supseteq M_2^0$.
b) $M^{\perp} = \mathrm{Lin}(M)^{\perp}$, $M^0 = \mathrm{Lin}(M)^0$
c) ${}^{\perp}(M^{\perp}) \supseteq M$, $(M^0)^0 \supseteq M$.
d) $({}^{\perp}(M^{\perp}))^{\perp} = M^{\perp}$, $M^{000} = M^0$.

Für Teilräume gilt ferner

e) $(M_1 + M_2)^{\perp} = M_1^{\perp} \cap M_2^{\perp}$, $(M_1 + M_2)^0 = M_1^0 \cap M_2^0$,
$M_1^{\perp} + M_2^{\perp} \subseteq (M_1 \cap M_2)^{\perp}$, $M_1^0 + M_2^0 \subseteq (M_1 \cap M_2)^0$.

Analoge Aussagen gelten für ${}^{\perp}N$ *(mit* $N \subset V$*) und* F^0 *(*$F \subseteq U^*$ *oder* $F \subseteq V^*$*).*

Im Weiteren sei stets $U = V$ und β symmetrisch oder schiefsymmetrisch.
Der Unterschied zwischen $M^{\perp}$ und ${}^{\perp}M$ entfällt dann, und wir schreiben häufig $\tilde{\beta} := \beta_1$.

Satz 11.21. *Sei* V *ein* K*-Vektorraum,* $U \subseteq V$ *ein Unterraum. Dann gilt:*

a) *Es ist* $V^*/U^0 \cong U^*$, *ein Isomorphismus wird durch*

$$f + U^0 \longmapsto f|_U$$

gegeben.
Insbesondere kann jedes $g \in U^*$ *zu* $\tilde{g} \in V^*$ *fortgesetzt werden (d.h.,* $\tilde{g}|_U = g$*).*

b) *Es gilt* $(V/U)^* \cong U^0$, *ein Isomorphismus wird durch*

$$\overline{f} \longrightarrow \overline{f} \circ \pi_U$$

gegeben, wo $\pi_U : V \longrightarrow V/U$ *die Projektion ist (also* $\pi_U(v) = v + U$*).*

c) *Ist* V *endlichdimensional, so ist*

$$\dim(U) + \dim(U^0) = \dim(V)$$

und analog

$$\dim(F) + \dim(F^0) = \dim(V)$$

für einen Teilraum $F \subseteq V^*$.

Beispiel. Sei $A = (a_{ij}) \in M(p \times n, K)$ eine Matrix mit Zeilen ${}^t\mathbf{z}_1, \ldots, {}^t\mathbf{z}_p$, ${}^t\mathbf{z}_i = (a_{i1}, \ldots, a_{in}) \in K^n = V$.
Für $1 \leq i \leq p$ sei $f_i \in V^*$ gegeben durch

$$f_i(\mathbf{x}) = \sum_{j=1}^{n} a_{ij} x_j.$$

Der Isomorphismus φ mit $\varphi(\mathbf{e}_j) = \mathbf{e}_j^*$ von $V = K^n$ nach V^* bildet $\mathbf{z}_i$ auf f_i ab für $1 \leq i \leq p$, also hat $\mathrm{Lin}(f_1, \ldots, f_p)$ die gleiche Dimension wie $\mathrm{Lin}(\mathbf{z}_1, \ldots, \mathbf{z}_p)$, nämlich $\mathrm{rg}(A)$.

Mit $U := \{\mathbf{x} \in \mathbb{K}^n \mid A\mathbf{x} = \mathbf{0}\}$ gilt

$$U = (\mathrm{Lin}(f_1, \ldots, f_p))^0,$$

also

$$\dim(U) = n - \dim(\mathrm{Lin}(f_1, \ldots, f_p)) = n - \mathrm{rg}(A)$$

nach Satz 11.21 c).

Der Satz enthält also die bekannte Formel für die Dimension des Lösungsraums eines linearen Gleichungssystems.

12. Hauptachsentransformation, Spektralsatz und euklidische Bewegungen

Definition und Lemma 12.1. *Sei $K = \mathbb{R}$ oder $K = \mathbb{C}$, V ein euklidischer oder unitärer Vektorraum über K mit Skalarprodukt $\langle\ ,\ \rangle$, sei $f \in \mathrm{End}(V)$. Dann gibt es genau eine Abbildung $f^* \in \mathrm{End}(V)$ mit*

$$\langle fv, w\rangle = \langle v, f^*w\rangle \quad \textit{für alle } v, w \in V.$$

f^* heißt die zu f adjungierte Abbildung. *Ist $f = f^*$, so heißt f* selbstadjungiert.

Ist $\mathcal{B}$ eine Orthonormalbasis von V und $A = M_{\mathcal{B}}(f)$ die Matrix von f bezüglich $\mathcal{B}$, so hat f^* bezüglich $\mathcal{B}$ die Matrix $A^* := {}^t\overline{A}$; diese heißt die zu A adjungierte Matrix.

A heißt selbstadjungiert (oder hermitesch), wenn $A = A^*$ gilt.

Beweis. Im euklidischen Fall ist das bereits in Korollar 11.18 gezeigt worden; im unitären Fall wählt man zunächst eine Orthonormalbasis $\mathcal{B}$, setzt $A = M_{\mathcal{B}}(f)$ und beweist dann wie in Korollar 11.18, dass die lineare Abbildung, deren Matrix bezüglich $\mathcal{B}$ die adjungierte Matrix A^* ist, zu f adjungiert ist. Alternativ modifiziert man den Beweis von Korollar 11.18 für den hermiteschen Fall. □

Bemerkung. Der Zusammenhang zwischen der Matrix von f und der Matrix von f^* wird komplizierter, wenn die Basis, bezüglich der die Matrizen betrachtet werden, keine Orthonormalbasis ist.

Lemma 12.2. *Seien $K, V, \langle\ ,\ \rangle$ wie oben, $f, g \in \mathrm{End}(V)$, $\lambda \in K$. Dann gilt:*

a) $(f + g)^* = f^* + g^*$
b) $(\lambda f)^* = \overline{\lambda} f^*$
c) $(f \circ g)^* = g^* \circ f^*$
d) $(f^*)^* = f$

für alle $f, g \in \mathrm{End}(V)$, $\lambda \in K$.

Beispiel. Skalarprodukte und adjungierte Abbildungen lassen sich auch für unendlichdimensionale Vektorräume definieren; allerdings ist dann die Existenz der adjungierten Abbildung nicht mehr garantiert.

Betrachte $V = \mathbb{C}[X]$ mit dem Skalarprodukt

$$\langle f, g\rangle := \int_0^1 f(t)\overline{g(t)}dt.$$

Für $h \in \mathbb{C}[X]$ hat man den Endomorphismus M_h von V, der durch

$$M_h(f) := hf$$

gegeben ist.
Man sieht: $M_h^* = M_{\overline{h}}$; in diesem Fall existiert also die adjungierte Abbildung.

Sei andererseits D der durch $D(f) = f'$ gegebene Ableitungsoperator. Falls es hierzu eine adjungierte Abbildung D^* gibt, so gilt für alle $f, g \in V$

$$\langle f, D^*g\rangle + \langle f, Dg\rangle = f(1)\overline{g}(1) - f(0)\overline{g}(0)$$

(partielle Integration). Ist also $g(1) = 1$, $g(0) = 0$ und $h = D^*g + Dg$, so ist

$$\int_0^1 f(t)\ \overline{h}(t)dt = f(1) \quad \text{für alle } f \in \mathbb{C}[X].$$

Speziell für $f = (X-1)\cdot h$ erhält man

$$\int_0^1 (t-1)^2\ |h(t)|^2 dt = 0, \ \text{ also } h = 0,$$

das ist ein Widerspruch zu

$$\int_0^1 f(t)\ \overline{h}(t)dt = f(1) \quad \text{für alle } f \in \mathbb{C}[X].$$

Eine adjungierte Abbildung zu D existiert also nicht.
Wir haben bereits in Abschnitt 9 gesehen, dass alle Eigenwerte selbstadjungierter Abbildungen reell sind (Definition und Satz 9.16) und gezeigt, dass es für jeden selbstadjungierten Endomorphismus eines euklidischen oder unitären Raumes eine Orthonormalbasis aus Eigenvektoren gibt (Satz 9.17, auch Satz von der Hauptachsentransformation genannt).
Geometrische Folgerungen, die auch den Namen Hauptachsentransformation rechtfertigen, haben wir am Ende von Abschnitt 9 betrachtet. Jetzt sollen noch einige weitere Folgerungen gezogen werden.

Korollar 12.3. *Sei $A \in M_n(\mathbb{C})$ hermitesch, $T \in \mathrm{GL}_n(\mathbb{C})$ so, dass ${}^tTA\overline{T} = B$ Diagonalgestalt mit p Einträgen $+1$, q Einträgen -1, $r = n-p-q$ Einträgen 0 auf der Diagonale hat.*
Dann ist p die Anzahl der positiven, q die Anzahl der negativen Eigenwerte von A und $r = n - \mathrm{rg}(A)$ die Vielfachheit von 0 als Eigenwert von A.

Beweis. Sei $U \in U_n(\mathbb{C})$ eine unitäre Matrix, so dass

$${}^t\bar{U}AU = U^{-1}AU =: D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

eine Diagonalmatrix ist; die Einträge $\lambda_1, \ldots, \lambda_n$ von D sind dann die Eigenwerte von A.
Sind $\mathbf{u}_1, \ldots, \mathbf{u}_n$ die Spalten von U, so sei (für $1 \le i \le n$) $\mathbf{s}_i = \frac{\bar{\mathbf{u}}_i}{\sqrt{|\lambda_i|}}$ falls $\lambda_i \neq 0$ und $\mathbf{s}_i = \bar{\mathbf{u}}_i$ für $\lambda_i = 0$, ferner sei $S \in GL_n(\mathbb{C})$ die Matrix mit Spalten $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Dann ist ${}^tSA\bar{S}$ eine Diagonalmatrix mit p' Einträgen $+1$, q' Einträgen -1 und r' Einträgen 0, wo p' die Anzahl der positiven Eigenwerte von A, q' die Anzahl der negativen Eigenwerte

von A (jeweils mit Vielfachheiten) und r' die Vielfachheit von 0 als Eigenwert von A ist.
Nach dem Trägheitssatz von Sylvester ist dann aber $p = p', q = q', r = r'$. □

Satz 12.4. (Spektralsatz, zweite Fassung) *Sei V ein euklidischer oder unitärer Raum, $f \in \mathrm{End}(V)$ selbstadjungiert, $\lambda_1, \ldots, \lambda_r$ die verschiedenen Eigenwerte von f.*

Für $\lambda \in \mathrm{Spec}(f) := \{\lambda_1, \ldots, \lambda_r\} = \{\lambda \in \mathbb{R} \mid \lambda$ ist Eigenwert von $f\}$ sei V_λ der Eigenraum von f zu λ. (Die Menge Spec(f) *der Eigenwerte von f heißt auch das* Spektrum *von f)*

Dann gilt:

a) *$V = \bigoplus_{j=1}^r V_{\lambda_j}$, und die V_{λ_j} sind paarweise orthogonal zueinander.*
b) *(*Spektralzerlegung *von f) Ist p_{λ_j} die orthogonale Projektion auf V_{λ_i} bezüglich der Zerlegung aus a), so ist $f = \sum_{j=1}^r \lambda_j p_{\lambda_j}$, und alle p_{λ_j} sind selbstadjungiert.*

Korollar 12.5. *Sei $A \in M_n^{\mathrm{sym}}(\mathbb{R})$ positiv semidefinit symmetrisch (also ${}^t\mathbf{x}A\mathbf{x} \geq 0$ für alle $\mathbf{x} \in \mathbb{R}^n$).*

Dann gibt es genau eine positiv semidefinite *symmetrische Matrix B, so dass $B^2 = A$ gilt. B heißt die positiv* semidefinite Wurzel *von A. Ist A positiv definit, so auch B.*

Beweis. Sei $T \in O_n(\mathbb{R})$ so, dass

$$ {}^tTAT = T^{-1}AT =: D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} $$

eine Diagonalmatrix ist. Man setze

$$ D_1 := \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} $$

und $B := {}^tT^{-1}D_1T^{-1} = TD_1T^{-1}$. Dann ist $B^2 = A$ und B ist symmetrisch positiv semidefinit, definit genau dann, wenn A definit ist.
Um die Eindeutigkeit zu sehen betrachten wir eine beliebige symmetrische positiv semidefinite Matrix B_2 mit $B_2^2 = A$. Dann ist B_2 diagonalisierbar und die Eigenräume von B_2 zu den Eigenwerten μ_i sind die Eigenräume von $A = B_2^2$ zu den Eigenwerten μ_i^2. Also ist $\mu_i = \sqrt{\lambda_i}$ und B_2 hat die gleichen Eigenwerte und die gleichen Eigenräume wie B, also ist $B_2 = B$. □

Korollar 12.6. (Polarzerlegung) *Sei $T \in \mathrm{GL}_n(\mathbb{R})$. Dann hat T eine eindeutige Zerlegung*

$$ T = P \cdot U $$

mit P positiv definit symmetrisch, $U \in O_n(\mathbb{R})$. Diese Zerlegung heißt die Polarzerlegung von T.

Korollar 12.7. *Sei V ein euklidischer oder unitärer Raum über K mit $K = \mathbb{R}$ oder $K = \mathbb{C}$, sei $\alpha : V \times V \to K$ ein Skalarprodukt auf V (also eine positiv definite symmetrische Bilinearform bzw. hermitesche Form). Sei*

$$\beta : V \times V \longrightarrow \begin{cases} \mathbb{C} \\ \mathbb{R} \end{cases}$$

eine weitere (beliebige) symmetrische Bilinearform bzw. hermitesche Form.
Dann hat V eine Orthonormalbasis $\mathcal{B}$ bezüglich des Skalarprodukts α, die gleichzeitig Orthogonalbasis bezüglich β ist.

In Matrixschreibweise: Ist $A \in M_n(\mathbb{C})$ eine positiv definite hermitesche Matrix und $B \in M_n(\mathbb{C})$ eine weitere (beliebige) hermitesche Matrix, so gibt es $T \in GL_n(\mathbb{C})$ mit ${}^tTA\bar{T} = E_n$ und ${}^tTB\bar{T}$ diagonal.
Sind A, B reell, so kann hier auch T reell gewählt werden.

Beweis. Ist $A = E_n$, so ist die Matrixform der Aussage gerade die Aussage von Satz 9.17. Ist A eine beliebige positiv definite Matrix, so finden wir nach Korollar 9.14 zunächst S mit ${}^tSA\bar{S} = E_n$. Mit $B' := {}^tSB\bar{S}$ finden wir dann nach dem ersten Schritt ein U mit ${}^tUE_n\bar{U} = E_n$ (also U unitär) und ${}^tUB'\bar{U}$ diagonal. Mit $T = SU$ haben wir dann die gesuchte Matrix T gefunden.
Die erste Form der Behauptung folgt daraus (nach Wahl einer beliebigen Basis und Übergang zu den Gram-Matrizen bezüglich dieser Basis) unmittelbar. □

Bemerkung. Man beachte, dass es für symmetrische bzw. hermitesche Matrizen zwei grundsätzlich verschiedene Methoden gibt, sie in Diagonalgestalt zu überführen:
Über einem beliebigen Körper K liefert das (verallgemeinerte) Gram-Schmidt-Verfahren eine Methode, zu einer symmetrischen Matrix A ein $T \in GL_n(K)$ zu finden, für das ${}^tTAT$ Diagonalgestalt hat. Die Spalten der Matrix T bilden in diesem Fall eine Orthogonalbasis von K^n bezüglich der Bilinearform β_A.
Ist die Matrix A zusätzlich diagonalisierbar im Sinne von Definition und Lemma 8.10, so findet man ein $S \in GL_n(K)$ mit $S^{-1}AS$ diagonal; die Spalten der Matrix S bilden dann eine Basis des K^n aus Eigenvektoren der Matrix A bzw. der linearen Abbildung $L_A \in \mathrm{End}(K^n)$.
Ist $K = \mathbb{R}$ oder $K = \mathbb{C}$ und A symmetrisch bzw. hermitesch, so haben wir in Satz 9.17 gezeigt, dass A diagonalisierbar im letzteren Sinne ist und dass die Matrix S in diesem Fall unitär (im reellen Fall orthogonal) gewählt werden kann, also mit $S^{-1} = {}^t\bar{S} = S^*$. Im reellen Fall haben wir also $S^{-1}AS = {}^tSAS$ für ein solches (orthogonales) S, und die beiden Anforderungen können *in diesem Fall* mit dem gleichen S

erfüllt werden. Deren Spalten sind dann *zugleich* eine Basis aus Eigenvektoren der linearen Abbildung L_A und eine Orthogonalbasis der symmetrischen Bilinearform β_A.

Satz 12.8. (Singulärwertzerlegung, Cartan-Zerlegung) *Sei $A \in M(m \times n, \mathbb{R})$, $m \leq n$. Dann gibt es Matrizen $U_1 \in O_m(\mathbb{R})$, $U_2 \in O_n(\mathbb{R})$ und $\mu_1, \ldots, \mu_m \in \mathbb{R}_{\geq 0}$, so dass*

$$U_1 A U_2 = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & & 0 \\ 0 & & \mu_m \end{pmatrix}$$

gilt.

Die μ_i sind eindeutig bestimmt, $\mu_1^2, \ldots, \mu_m^2$ sind die Eigenwerte von $A \cdot {}^tA$.

Für $n \leq m$ erhält man entsprechend

$$U_1 A U_2 = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \\ & 0 & \end{pmatrix},$$

wo die μ_j^2 die Eigenwerte von ${}^tAA$ sind.

Beweis. Wir beschränken uns beim Beweis auf den Fall $n \leq m$, der andere Fall geht daraus durch Transponieren hervor.
Sei $V = \mathbb{R}^n, W = \mathbb{R}^n, f := L_A : V \longrightarrow W$ die durch A bezüglich der Standardbasen von V und W gegebene lineare Abbildung. Auf V und auf W haben wir das Standardskalarprodukt $\langle , \rangle$.
Auf V betrachten wir nun die offenbar positiv semidefinite symmetrische Bilinearform β, die durch

$$\beta(\mathbf{x}, \mathbf{y}) := \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = {}^t\mathbf{x}\,({}^tAA)\mathbf{y}$$

gegeben ist, ihre Gram-Matrix bezüglich der Standardbasis von V ist ${}^tAA$.
Sei $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V bezüglich des Standardskalarprodukts, die bezüglich β eine Orthogonalbasis ist, für die also $\beta(v_i, v_j) = \lambda_j \delta_{ij}$ gilt; eine solche Basis von V gibt es nach Korollar 12.7. Die λ_j sind nichtnegativ, sie seien so angeordnet, dass $\lambda_j > 0$ für $1 \leq j \leq p$ und $\lambda_{p+1} = \cdots = \lambda_n = 0$ gilt. Dabei ist $p = \mathrm{rg}({}^tAA) \leq \mathrm{rg}(A) \leq n$, und für $j > p$ gilt

$$0 = \beta(v_j, v_j) = \langle f(v_j), f(v_j) \rangle,$$

also $f(v_j) = 0$ (und daher $\mathrm{rg}(A) = p$).

Für $1 \leq j \leq n$ setzen wir dann $\mu_j = \sqrt{\lambda_j}$ und

$$w_j := \frac{f(v_j)}{\mu_j} \quad \text{falls } j \leq p,$$

die Vektoren $w_1, \ldots, w_p$ bilden dann wegen $\langle f(v_j), f(v_k) \rangle = \mu_j^2 \delta_{jk}$ ein Orthonormalsystem im euklidischen Raum W. Wir ergänzen dieses Orthonormalsystem durch Vektoren $w_{p+1}, \ldots, w_m$ zu einer Orthonormalbasis von W und haben $f(v_j) = \mu_j w_j$ für $1 \leq j \leq m$ sowie $f(v_j) = \mathbf{0}$ für $j > m$. Die Matrix von f bezüglich der Orthonormalbasen $(v_1, \ldots, v_n)$ von V und $(w_1, \ldots, w_m)$ von W hat daher die in der Behauptung angegebene Gestalt

$$\begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \\ 0 & & 0 \end{pmatrix}.$$

Ist $U_1^{-1} \in O_m(\mathbb{R})$ die Matrix mit den Spalten $w_1, \ldots, w_m$ und $U_2 \in O_n(\mathbb{R})$ die Matrix mit den Spalten $v_1, \ldots, v_n$, so ist wie behauptet

$$U_1 A U_2 = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \\ 0 & & 0 \end{pmatrix},$$

wobei $\mu_j = \sqrt{\lambda_j}$ gilt und die λ_j die Eigenwerte von ${}^tAA$ sind. Da man leicht zeigt, dass ${}^tAA$ und $A\,{}^tA$ die gleichen von 0 verschiedenen Eigenwerte haben (mit Vielfachheiten) (Übung), sind die μ_j^2 auch die Eigenwerte von $A\,{}^tA$.
Um die Eindeutigkeit der μ_j zu zeigen, betrachten wir eine Zerlegung

$$U_1 A U_2 = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \\ 0 & & 0 \end{pmatrix}$$

wie in der Behauptung mit zunächst beliebigen $\mu_j \in \mathbb{R}$ und $U_1 \in O_m(\mathbb{R}), U_2 \in O_n(\mathbb{R})$. Dann ist

$$\begin{aligned} \begin{pmatrix} \mu_1^2 & & 0 \\ & \ddots & \\ 0 & & \mu_n^2 \end{pmatrix} &= {}^t(U_1 A U_2)\,(U_1 A U_2) \\ &= U_2^{-1}({}^tAA)U_2, \end{aligned}$$

$\mu_1^2, \ldots, \mu_n^2$ sind also (mit den gleichen Vielfachheiten) genau die Eigenwerte von $({}^tAA)$. $\square$

Bemerkung. Die Singulärwertzerlegung ist ein wichtiges Werkzeug bei der numerischen Behandlung von Matrizen. Im Fall $m = n$, $A \in$

$\mathrm{GL}_n(\mathbb{R})$ erhält man eine Zerlegung, die in der Theorie der Lie-Gruppen eine große Rolle spielt und dort als Cartan-Zerlegung bekannt ist.
Im Rest dieses Abschnitts wollen wir ähnliche Normalformen, wie wir sie für selbstadjungierte Endomorphismen bzw. deren Matrizen gesehen haben, auch für unitäre und orthogonale (und allgemeiner normale) Transformationen herleiten. Wir werden zeigen, dass auch für unitäre Transformationen Orthonormalbasen aus Eigenvektoren existieren; für orthogonale Transformationen ist die Lage geringfügig komplizierter, da sie keine reellen Eigenwerte haben müssen.

Definition und Lemma 12.9. *Sei V ein unitärer Raum über $\mathbb{C}$, $f \in \mathrm{End}(V)$. f heißt normal, wenn $ff^* = f^*f$ gilt. Eine Matrix $A \in M_n(\mathbb{C})$ heißt normal, wenn*

$$A \cdot {}^t\overline{A} = {}^t\overline{A} \cdot A$$

Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V, so ist ein Endomorphismus $f \in \mathrm{End}(V)$ genau dann normal, wenn seine Matrix bezüglich der Basis $\mathcal{B}$ normal ist.
Insbesondere gilt: Ist die Matrix von f bezüglich einer *Orthonormalbasis von V eine normale Matrix, so ist die Matrix von f bezüglich jeder beliebigen Orthonormalbasis von V eine normale Matrix.*

Beweis. Klar. □

Bemerkung. Nach den bisherigen Ergebnissen zum Spektralsatz gibt es zu einer reellen Matrix A genau dann eine Orthonormalbasis des $\mathbb{R}^n$ aus Eigenvektoren von A, wenn A symmetrisch ist (dass aus $A = {}^tT \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} T$ die Symmetrie von A folgt, ist trivial.) Genauso gibt es für $A \in M_n(\mathbb{C})$ genau dann eine Orthonormalbasis des $\mathbb{C}^n$ (bezüglich des Standardskalarprodukts) aus Eigenvektoren von A zu reellen Eigenwerten, wenn A hermitesch ist.
Der Begriff „normal“ dient dazu, hier auch den Fall nicht reeller Eigenwerte zu behandeln.

Lemma 12.10. *Sei K ein beliebiger Körper, V ein K-Vektorraum, $f, g \in \mathrm{End}(V)$ mit $f \circ g = g \circ f$.*

a) *Ist λ ein Eigenwert von f, $V_\lambda := V_\lambda(f)$ der zugehörige Eigenraum von f, so ist $g(V_\lambda) \subseteq V_\lambda$.*
b) *Ist $K = \mathbb{R}$ oder $K = \mathbb{C}$ und V euklidisch bzw. unitär, so ist $g^*(V_\lambda^\perp) \subseteq V_\lambda^\perp$.*

Satz 12.11. *Sei V ein unitärer Raum über $\mathbb{C}$. Dann gibt es zu $f \in \mathrm{End}(V)$ genau dann eine Orthonormalbasis (bezüglich des Standardskalarprodukts) von V aus Eigenvektoren von f, wenn f normal ist.*

Allgemeiner gilt: Ist $M \subseteq \mathrm{End}(V)$ eine Unteralgebra, die kommutativ

und unter Adjungiertenbildung abgeschlossen ist, so gibt es eine Orthonormalbasis von V (bezüglich des Standardskalarprodukts), die aus simultanen Eigenvektoren der Elemente von M besteht.

Beweis. Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V, die aus Eigenvektoren des Endomorphismus f besteht, so ist die Matrix A von f bezüglich dieser Basis eine Diagonalmatrix. Man sieht dann sofort, dass $A^*A = AA^*$ gilt, also ist A und damit f normal nach Definition/Lemma 12.9.
Um umgekehrt die Existenz einer Orthonormalbasis aus Eigenvektoren für ein normales f zu zeigen, stellen wir zunächst fest, dass diese aus der allgemeineren Aussage über kommutative und unter Adjunktion abgeschlossene Algebren von Endomorphismen folgt. Ist nämlich $f \in \mathrm{End}(V)$ normal, so ist die von f und f^* erzeugte Unteralgebra

$$\mathbb{C}[f, f^*] := \{\sum_{i,j=0}^{m} a_{ij} f^i (f^*)^j \mid m \in \mathbb{N}_0, a_{ij} \in \mathbb{C}\}$$

eine kommutative und unter Adjunktion abgeschlossene Teilalgebra von $\mathrm{End}(V)$, eine Orthonormalbasis aus gemeinsamen Eigenvektoren aller Elemente von $\mathbb{C}[f, f^*]$ besteht dann natürlich insbesondere aus Eigenvektoren von $f \in \mathbb{C}[f, f^*]$.
Sei also jetzt M eine kommutative und unter Adjunktion abgeschlossene Teilalgebra von $\mathrm{End}(V)$. Wir zeigen die Behauptung durch Induktion nach $n = \dim(V)$, der Induktionsanfang $n = 1$ ist wieder einmal trivial. Ist $n > 1$ und die Behauptung für unitäre Räume gezeigt, deren Dimension kleiner als n ist, so ist die Behauptung sicher trivial, wenn $M = \mathbb{C} \cdot \mathrm{Id}_V$ gilt. Andernfalls sei $f \notin \mathbb{C} \cdot \mathrm{Id}_V$ und $\lambda \in \mathbb{C}$ ein Eigenwert von f mit Eigenwert λ, sei $V_\lambda := V_\lambda(f)$ der zugehörige Eigenraum. Für $g \in M$ gilt dann wegen der Kommutativität von M nach Lemma 12.10

$$g(V_\lambda) \subseteq V_\lambda, \quad g^*(V_\lambda^\perp) \subseteq V_\lambda^\perp.$$

Da M abgeschlossen unter Adjunktion ist, ist $g^* \in M$, und wir erhalten (mit $(g^*)^* = g$) genauso

$$g^*(V_\lambda) \subseteq V_\lambda, \quad g(V_\lambda^\perp) \subseteq V_\lambda^\perp.$$

In der Zerlegung

$$V = V_\lambda \oplus V_\lambda^\perp$$

operiert also M auf beiden Summanden, und nach Induktionsannahme haben V_λ und $V_\lambda^\perp$ jeweils eine Orthonormalbasis aus gemeinsamen Eigenvektoren aller Elemente von M (Da $M \neq \mathbb{C} \cdot \mathrm{Id}_V$ ist, haben beide Summanden kleinere Dimension als V). Setzt man diese Basen von V_λ und $V_\lambda^\perp$ zu einer Basis von V zusammen, so hat man die gesuchte Orthonormalbasis von V aus gemeinsamen Eigenvektoren aller Elemente von M. $\square$

Bemerkung. Auf ähnliche Weise kann man zeigen: Ist K ein beliebiger Körper, V ein endlichdimensionaler K-Vektorraum, $M \subseteq \mathrm{End}(V)$ eine kommutative Unteralgebra, so dass alle Elemente von M diagonalisierbar sind, so besitzt V eine Basis aus simultanen Eigenvektoren der Elemente von M.

Korollar 12.12. *Ist $A \in M_n(\mathbb{C})$ eine normale Matrix (also ${}^t\overline{A} \cdot A = A \cdot {}^t\overline{A}$), so gibt es $U \in U_n(\mathbb{C})$, so dass ${}^t\overline{U}AU$ Diagonalgestalt hat.*

Beweis. Das ist die Matrixversion des vorigen Satzes, man erhält sie, indem man den Satz auf den Endomorphismus L_A von $\mathbb{C}^n$ anwendet und die Vektoren der danach gefundenen Orthonormalbasis des $\mathbb{C}^n$ aus Eigenvektoren von A als Spaltenvektoren in die unitäre Matrix $\overline{U}$ einträgt. □

Korollar 12.13. *Sei $A \in U_n(\mathbb{C})$. Dann gibt es $U \in U_n(\mathbb{C})$, so dass*

$$ {}^t\overline{U}AU = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \text{ mit } \lambda_j \in \mathbb{C},\ |\lambda_j| = 1 $$

gilt.
Insbesondere gilt: Alle Eigenwerte einer unitären Matrix haben Betrag 1*, alle reellen Eigenwerte einer unitären Matrix sind entweder* 1 *oder* -1.

Beweis. Wegen $A^* = A^{-1}$ ist A offenbar normal, lässt sich also durch Konjugation mit einer unitären Matrix in Diagonalgestalt bringen. Wir müssen nur noch zeigen, dass alle Eigenwerte einer unitären Matrix Betrag 1 haben.
Ist also $\mathbf{x} \in \mathbb{C}^n$ ein Eigenvektor der unitären Matrix A zum Eigenwert λ , so gilt

$$ |\lambda|^2 \langle \mathbf{x}, \mathbf{x} \rangle = \langle A\mathbf{x}, A\mathbf{x} \rangle = \langle \mathbf{x}, A^*A\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle, $$

also $|\lambda|^2 = 1$. □

Bevor wir unsere bisherigen Ergebnisse benutzen, um orthogonale Matrizen bzw. Abbildungen noch genauer zu untersuchen, soll deren geometrische Bedeutung betrachtet werden.
Dafür untersuchen wir zunächst diejenigen Abbildungen von euklidischen Vektorräumen, die sich mit der zusätzlichen Struktur vertragen, die durch das Skalarprodukt gegeben ist.
Der Einfachheit halber behandeln wir nicht abstrakte euklidische Räume sondern den $\mathbb{R}^n$ mit dem Standard-Skalarprodukt $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^n x_j y_j$ und der daraus abgeleiteten Norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{j=1}^n x_j^2}$. Für $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ nennen wir $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|$ den (euklidischen) Abstand von $\mathbf{x}$ und $\mathbf{y}$.

Definition 12.14. *Eine Abbildung* $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *heißt* euklidische Bewegung*, wenn für alle* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$$

gilt, wenn φ *also abstandserhaltend ist.*
φ *heißt eine (lineare) orthogonale Abbildung, wenn* $\varphi = L_A$ *für eine orthogonale Matrix* A *ist.*

Satz 12.15. *Sei* $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *eine Abbildung. Dann sind q̈uivalent:*

a) φ *ist eine euklidische Bewegung mit* $\varphi(\mathbf{0}) = \mathbf{0}$.
b) *Für alle* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *gilt*

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle.$$

c) *Es gibt eine orthogonale Matrix* A*, so dass* $\varphi(\mathbf{x}) = A \cdot \mathbf{x}$ *für alle* $\mathbf{x} \in \mathbb{R}^n$ *gilt, d.h.,* φ *ist eine (lineare) orthogonale Abbildung.*

Beweis. a)⇒b): Sei φ eine euklidische Bewegung, die den Nullpunkt festlässt. Zunächst haben wir für $\mathbf{x} \in \mathbb{R}^n$:

$$\begin{aligned} \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}) \rangle &= \|\varphi(\mathbf{x})\|^2 \\ &= \|\varphi(\mathbf{x}) - \varphi(\mathbf{0})\|^2 \\ &= \|\mathbf{x} - \mathbf{0}\|^2 \\ &= \|\mathbf{x}\|^2 \\ &= \langle \mathbf{x}, \mathbf{x} \rangle. \end{aligned}$$

Für $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ haben wir dann

$$\begin{aligned} -2\langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle &= \langle \varphi(\mathbf{x}) - \varphi(\mathbf{y}), \varphi(\mathbf{x}) - \varphi(\mathbf{y}) \rangle - \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}) \rangle - \langle \varphi(\mathbf{y}), \varphi(\mathbf{y}) \rangle \\ &= \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle \\ &= -2\langle \mathbf{x}, \mathbf{y} \rangle, \end{aligned}$$

die Abbildung φ erhält also wie behauptet das Skalarprodukt.
b)⇒ c): Falls eine Abbildung $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ das Skalarprodukt erhält und jeden der kanonischen Basisvektoren $\mathbf{e}_i$ festlässt und $\mathbf{x} \in \mathbb{R}^n$ ein Vektor mit $\psi(\mathbf{x}) = \mathbf{y}$ ist, so ist

$$\begin{aligned} x_i &= \langle \mathbf{x}, \mathbf{e}_i \rangle \\ &= \langle \psi(\mathbf{x}), \psi(\mathbf{e}_i) \rangle \\ &= \langle \mathbf{y}, \mathbf{e}_i \rangle \\ &= y_i, \end{aligned}$$

also $\psi(\mathbf{x}) = \mathbf{x}$ für alle $\mathbf{x} \in \mathbb{R}^n$, d.h., $\psi = \mathrm{Id}$. Wir betrachten jetzt unsere Abbildung φ, von der wir annehmen, dass sie das Skalarprodukt erhält. Für $1 \leq i \leq n$ sei $\varphi(\mathbf{e}_i) =: \mathbf{e}'_i$, sei A die Matrix, deren Spalten die Vektoren $\mathbf{e}'_1, \ldots, \mathbf{e}'_n$ sind.
Da φ das Skalarprodukt erhält, bilden die $\mathbf{e}'_i$ eine Orthonormalbasis des $\mathbb{R}^n$, die Matrix A ist also eine orthogonale Matrix und die Abbildung L_A ebenso wie ihre inverse $L_{A^{-1}}$ erhält das Skalarprodukt. Daher erhält

auch die Abbildung $\rho := L_A^{-1} \circ \varphi$ das Skalarprodukt; da sie alle $\mathbf{e}_i$ fixiert, ist sie die Identität, es gilt also $\varphi = L_A$ mit der orthogonalen Matrix A, d.h, es gilt c)
c)$\Rightarrow$ a) schließlich ist trivial. $\square$

Bemerkung. Da im Anschauungsraum $\mathbb{R}^3$ der Winkel α zwischen den Vektoren v und w bekanntlich mit Hilfe der Formel

$$\cos(\alpha) = \frac{\langle v, w \rangle}{\|v\| \|w\|}$$

durch die Berechnung von Skalarprodukten bestimmt wird, zeigt der Satz, dass abstandstreue Abbildungen, die den Ursprung fixieren, zusätzlich winkeltreu und linear sind.

Korollar 12.16. *Sei $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ eine euklidische Bewegung, sei $\mathbf{b} := \varphi(\mathbf{0}) \in \mathbb{R}^n$, sei $T_{\mathbf{b}}$ die durch*

$$T_{\mathbf{b}}(x) := \mathbf{x} + \mathbf{b}$$

definierte Translation um den Vektor $\mathbf{b}$.
Dann gibt es eine orthogonale Matrix $A \in O_n(\mathbb{R}) := \{A \in GL_n(\mathbb{R}) \mid A = {}^tA^{-1}\}$, so dass

$$\varphi = T_{\mathbf{b}} \circ L_A$$

gilt, so dass also

$$\varphi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$$

für alle $\mathbf{x} \in \mathbb{R}^n$ ist.
Jede euklidische Bewegung lässt sich also als Komposition einer Translation und einer linearen orthogonalen Abbildung schreiben.

Beweis. Klar. $\square$

Korollar 12.17. *Sei $A \in O_n(\mathbb{R})$. Dann gibt es $U \in O_n(\mathbb{R})$, so dass*

$$U^{-1}AU = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_r \end{pmatrix}$$

mit $D_j = (\pm 1) \in M(1 \times 1, \mathbb{R})$ oder $D_j = \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix} \in M(2 \times 2, \mathbb{R})$, $\theta_j \in \mathbb{R}$ gilt.

Beweis. Eine äquivalente Formulierung der Behauptung ist: Es gibt eine Orthonormalbasis von $\mathbb{R}^n$ bezüglich des Standardskalarprodukts, bezüglich der die Matrix der A zugeordneten lineare Abbildung $f = L_A$ (mit $L_A(\mathbf{x}) = A\mathbf{x}$) die angegebene Gestalt hat.
Wir zeigen diese Behauptung durch Induktion nach n, der Induktionsanfang $n = 1$ ist trivial. Sei also $n > 1$ und die Behauptung für $n' < n$ gezeigt.
Hat A einen reellen Eigenwert λ, so ist $\lambda = \pm 1$. Ist v ein Eigenvektor von f zu diesem Eigenwert, so gibt es zu dem (ebenfalls orthogonalen) Endomorphismus $f|_{(\mathrm{Lin}(v))^\perp}$ eine Orthonormalbasis von $(\mathrm{Lin}(v))^\perp$,

bezüglich der die Matrix von $f|_{(\mathrm{Lin}(v))^\perp}$ die angegebene Gestalt hat. Ergänzt man diese durch $\frac{v}{\|v\|}$ zu einer Orthonormalbasis von $V = \mathbb{R}^n$, so hat f bezüglich dieser Basis die angegebene Gestalt, und wir sind in diesem Fall fertig.
Andernfalls ist keiner der Eigenwerte von A reell. Sei dann $\lambda \in \mathbb{C}$ ein Eigenwert von $f_\mathbb{C} := L_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ und $\mathbf{v} \in \mathbb{C}^n$ ein Eigenvektor von $f_\mathbb{C}$ mit $\|\mathbf{v}\| = 1$. Da A reell ist, gilt für den Vektor $\overline{\mathbf{v}}$, der aus $\mathbf{v}$ durch komponentenweise komplexe Konjugation entsteht,

$$A\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}},$$

der Vektor $\bar{\mathbf{v}}$ ist also ein Eigenvektor von $f_\mathbb{C}$ zum Eigenwert $\bar{\lambda}$, und da nach Voraussetzung λ nicht reell ist, ist $\lambda \neq \bar{\lambda}$.
Ebenso wie für selbstadjungierte Abbildungen gilt auch für unitäre Abbildungen, dass Eigenvektoren zu verschiedenen Eigenwerten orthogonal zueinander sind: Man hat nämlich für unitäres f und Eigenvektoren w, w' zu Eigenwerten $\mu \neq \mu'$ mit $|\mu| = |\mu'| = 1$

$$\langle w, w'\rangle = \langle f(w), f(w')\rangle = \mu\overline{\mu'}\langle w, w'\rangle = \mu(\mu')^{-1}\langle w, w'\rangle,$$

wegen $\mu \neq \mu'$ folgt dann $\langle w, w'\rangle = 0$.
Wir haben also $\langle \mathbf{v}, \bar{\mathbf{v}}\rangle = 0$, für die Vektoren

$$\tilde{\mathbf{w}}_1 := \mathbf{v} + \bar{\mathbf{v}}, \quad \tilde{\mathbf{w}}_2 := i(\mathbf{v} - \bar{\mathbf{v}}) \in \mathbb{R}^n$$

gilt also (man rechne $\langle \tilde{\mathbf{w}}_j, \tilde{\mathbf{w}}_j\rangle$ für $j = 1, 2$ aus) $\|\tilde{\mathbf{w}}_1\| = \|\tilde{\mathbf{w}}_2\| = \sqrt{2}$.
Die Vektoren $\mathbf{w}_1 := \tilde{\mathbf{w}}_1/\sqrt{2}, \mathbf{w}_2 := \tilde{\mathbf{w}}_2/\sqrt{2}$ bilden also eine Orthonormalbasis von $U := \mathrm{Lin}(w_1, w_2)$. Da $|\lambda| = 1$ gilt, kann man $\lambda = \exp(i\theta) = \cos(\theta) + i\sin(\theta)$ mit $\theta \in \mathbb{R}$ schreiben, man rechnet dann leicht nach, dass

$$\begin{aligned} f(\mathbf{w}_1) &= \cos(\theta)\mathbf{w}_1 + \sin(\theta)\mathbf{w}_2 \\ f(\mathbf{w_2}) &= -\sin(\theta)\mathbf{w}_1 + \cos(\theta)\mathbf{w}_2 \end{aligned}$$

gilt, so dass als $f|_U$ bezüglich der Orthonormalbasis $(\mathbf{w}_1, \mathbf{w}_2)$ von U die Matrix

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

hat. Ergänzt man $(\mathbf{w}_1, \mathbf{w}_2)$ mit Hilfe der Induktionsannahme durch eine Orthonormalbasis von $U^\perp$, bezüglich der $f|_{U^\perp}$ eine Matrix vom angegebenen Typ hat, so erhält man eine Basis von V, bezüglich der f eine Matrix vom angegebenen Typ hat. □

Korollar 12.18. *Ist $A \in O_3(\mathbb{R})$, so hat $\det(A) \cdot A$ wenigstens einen Fixvektor ($\neq \mathbf{0}$); $(\det A) \cdot A$ stellt eine Drehung um die Achse in Richtung des Fixvektors dar.*
Insbesondere gilt der Satz vom Fußball*: Auf einem Fußball gibt es wenigstens zwei Punkte, die sich zu Beginn der zweiten Halbzeit des Spiels am gleichen Ort (relativ zum Stadion) befinden wie zu Beginn der ersten Halbzeit.*

Bemerkung. Beim Satz vom Fußball macht man natürlich die idealisierenden Annahmen, dass der Ball beim Anpfiff stets genau auf dem Anstoßpunkt liegt, dass beide Halbzeiten mit dem gleichen Ball gespielt werden und dass der Ball während der ersten Halbzeit nicht deformiert wurde.

Beweis. Der erste Teil des Satzes folgt aus dem vorigen Korollar: Da $\det(A) = \pm 1$ für $A \in O_n(\mathbb{R})$ gilt und $\det(\det(A)A) = (\det(A))^4$ für $A \in O_3(\mathbb{R})$ ist, hat $A_1 := \det(A)A$ Determinante 1. In der Normalgestalt aus dem vorigen Lemma ist A_1 daher entweder eine Diagonalmatrix mit einer geraden Anzahl von Einträgen -1, also wenigstens einem Eintrag $+1$, oder von der Form

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},$$

wobei der Eintrag oben links $+1$ sein muss, damit die Determinante $+1$ wird.
A_1 hat also in jedem Fall den Eigenwert 1, d.h., die durch A gegebene lineare Abbildung hat einen Fixvektor. In der Ebene senkrecht zum Fixvektor wirkt die durch A_1 gegebene lineare Abbildung durch den unteren rechten 2×2-Block der Normalgestalt der Matrix, also durch $\pm\mathrm{Id}$ (= Drehung um $0°$ oder um $180°$) falls diese diagonal ist bzw. durch die Drehung um den Winkel θ andernfalls. □

Zum Beweis des Satzes vom Fußball wird noch ein Lemma gebraucht:

Lemma 12.19. *Sei $0 < t_0 \in \mathbb{R}$, für $t \in [0, t_0] \subseteq \mathbb{R}$ sei $g_t : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ eine abstandstreue Abbildung (euklidische Bewegung), die nach Korollar 12.16 als $g_t(\mathbf{x}) = A_t \cdot \mathbf{x} + g_t(\mathbf{0})$ mit $A_t \in O_n(\mathbb{R})$ für alle $t \in [0, t_0]$ geschrieben sei; die Abbildung $t \longmapsto A_t \in M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ sei dabei stetig und es gelte $g_0 = \mathrm{Id}$.*
Dann ist $\det(A_t) = 1$ für alle $t \in [0, t_0]$.

Beweis. Die Abbildung $A \mapsto \det(A)$ von $M_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ nach $\mathbb{R}$ ist auf Grund der Formel von Leibniz für die Determinante stetig, daher ist die zusammengesetzte Abbildung $t \mapsto \det(A_t) : [0, t_0] \longrightarrow \mathbb{R}$ stetig. Sie hat in $t = 0$ den Wert 1 und kann nur die Werte 1 und -1 annehmen. Nach dem Zwischenwertsatz muss sie dann konstant gleich 1 sein. □

Beweis des Satzes vom Fußball. Die Bewegung des Balls ist eine euklidische Bewegung, bei der bei jedem Anstoß der Ballmittelpunkt an der gleichen Stelle im Stadion ist (nämlich senkrecht über dem Anstoßpunkt in der durch den Radius des Balls gegebenen Höhe). Wählen wir diesen Punkt als Ursprung des Koordinatensystems, so geht also die Position $\mathbf{y}$ eines Punktes auf dem Ball, der sich beim Anpfiff des Spiels in $\mathbf{x}$ befand, bei Beginn der zweiten Halbzeit aus $\mathbf{x}$ durch

$\mathbf{y} = A\mathbf{x}$ mit $A \in SO_3(\mathbb{R})$ hervor. Nach dem ersten Teil des Satzes hat A einen Fixvektor, ist also die Drehung um die Achse durch diesen Vektor. Die beiden Punkte, in denen diese Achse durch die Oberfläche des Balls geht, befinden sich daher beim Anpfiff zur zweiten Halbzeit an der gleichen Stelle wie beim Anpfiff zur ersten Halbzeit. □

Beispiel. Zwei Drehungen $f \neq \mathrm{Id} \neq g$ im $\mathbb{R}^3$ sind genau dann miteinander vertauschbar, wenn sie entweder die gleiche Drehachse haben oder wenn es Drehungen um zueinander orthogonale Achsen um jeweils 180° sind. Beweis: Übung.

Zusammenfassung

Für einen beliebigen Körper K und $A \in M_n^{\mathrm{sym}}(K)$ symmetrisch gibt es $T \in \mathrm{GL}_n(K)$, so dass ${}^tTAT$ Diagonalgestalt hat (Gram-Schmidt).
Dabei ist im allgemeinen ${}^tT \neq T^{-1}$, die Einträge der Diagonalmatrix sind in der Regel keine Eigenwerte von A und A ist nicht notwendig diagonalisierbar.
Ist $A \in M_n^{\mathrm{sym}}(\mathbb{R})$, so gibt es dagegen $U \in O_n(\mathbb{R})$ mit ${}^tUAU = D$ diagonal. Da hier ${}^tU = U^{-1}$ gilt, sind die Einträge der Diagonalmatrix die Eigenwerte von A, A ist diagonalisierbar. Die Spalten der Transformationsmatrix U bilden eine Orthonormalbasis des $\mathbb{R}^n$ aus Eigenvektoren von A.
Verzichtet man auf die Bedingung $U \in O_n(\mathbb{R})$, so erreicht man hier (falls $\det(A) \neq 0$ ist)

$$
{}^tTAT = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{pmatrix}
$$

mit einer oberen Dreiecksmatrix $T \in \mathrm{GL}_n(\mathbb{R})$, die Anzahl p der Einträge $+1$ bzw. q der Einträge -1 ist dabei nach dem Trägheitssatz von Sylvester eindeutig bestimmt und gleich der Anzahl (mit Vielfachheit) der positiven bzw. negativen Eigenwerte von A (aber ± 1 ist im allgemeinen kein Eigenwert von A), (p, q) (oder gelegentlich $p - q$) heißt die Signatur von A.

Ist $A \in M_n(\mathbb{C})$ hermitesch, so gibt es $U \in U_n(\mathbb{C})$, so dass ${}^t\overline{U}AU$ eine Diagonalmatrix ist. Da hier ${}^t\overline{U} = U^{-1}$ gilt, sind die Einträge der Diagonalmatrix die Eigenwerte von A, A ist diagonalisierbar, die Eigenwerte von A sind überdies reell. Ist $A \in M_n(\mathbb{C})$ nur normal, so kann A wie oben diagonalisiert werden, die Eigenwerte brauchen dann aber nicht reell zu sein (sind sie reell, so ist A schon hermitesch). Die Spalten der

Matrix $\bar{U}$ bilden in jedem dieser Fälle eine Orthonormalbasis des $\mathbb{C}^n$ aus Eigenvektoren von A.

Da unitäre und orthogonale Matrizen insbesondere normal sind, gilt die Aussage für normale Matrizen von oben insbesondere auch für unitäre und für orthogonale Matrizen. Für beide sind alle Eigenwerte vom Betrag 1. Für orthogonale Matrizen folgt, daß sie sich durch Konjugation $A \longmapsto T^{-1}AT$ mit $T \in O_n(\mathbb{R})$ in Blockdiagonalgestalt bringen lassen, wobei die Blöcke ± 1 oder 2-dimensionale Drehmatrizen sind.

13. Minimalpolynom und Satz von Hamilton-Cayley

Wir kehren in diesem und dem nächsten Abschnitt zurück zu dem Problem, quadratische Matrizen bis auf Ähnlichkeit zu klassifizieren. Wir wissen bereits, dass das charakteristische Polynom für dieses Problem eine wichtige Rolle spielt. Um das ebenfalls wichtige Minimalpolynom zu definieren, brauchen wir noch etwas Vorbereitung über das Rechnen im Polynomring. Für die grundlegenden Definitionen und Eigenschaften des Polynomrings $K[X]$ über dem Körper K erinnern wir an Beispiel e) nach Satz 5.2 und an Definition und Satz 8.3 sowie Satz 8.4.

Definition 13.1. *Sei R ein kommutativer Ring.*
*Eine R-*Algebra *ist ein (nicht notwendig kommutativer) Ring A mit Einselement, so dass gilt:*

a) *A ist ein R- Modul, d.h., man hat eine als $(r, a) \mapsto ra$ geschriebene Verknüpfung $R \times A \to A$, für die die Vektorraumaxiome SM1 bis SM4 (siehe Definition 3.1) gelten.*
b) *Für $a, b \in A, \lambda \in R$ gilt $\lambda(a \cdot b) = (\lambda a) \cdot b = a \cdot (\lambda b)$.*

Beispiel:

a) Ist K ein Körper und $L \supseteq K$ ein Oberkörper von K, so ist L eine K-Algebra.
b) Ist R ein kommutativer Ring und $M_n(R)$ die Menge der $n \times n$-Matrizen über R, so ist $M_n(R)$ eine R-Algebra.
c) Ist K ein Körper, V ein K-Vektorraum, so ist $\mathrm{End}(V)$ eine K-Algebra.

Definition 13.2. *Sei R ein kommutativer Ring mit* 1.
Eine Teilmenge $I \subseteq R$ heißt ein Ideal *in R wenn gilt:*

a) *$I \neq \emptyset$*
b) *Für alle $a, b \in I$ ist $a + b \in I$*
c) *Ist $a \in I, r \in R$, so ist $ra \in I$.*

Satz 13.3. *Sei K ein Körper. Dann ist im Polynomring $K[X]$ jedes Ideal ein Hauptideal; man sagt, $K[X]$ sei ein* Hauptidealring.
Ein Ideal $I \neq \{0\}$ wird erzeugt von dem (eindeutig bestimmten) normierten Polynom kleinsten Grades in I.

Beweis. Sei $I \neq \{0\}$ ein Ideal in $K[X]$ und g ein normiertes Polynom vom kleinstmöglichen Grad in I. Ist $f \in I$, so kann man $f = qg + r$ mit $q, r \in K[X]$ und $r = 0$ oder $\deg(r) < \deg(g)$ schreiben. Da $r = f - qg \in I$ aus der Idealeigenschaft von I folgt und $\deg(g)$ der kleinstmögliche Grad eines Polynoms $\neq 0$ in I ist, muss $r = 0$ gelten, also ist f im von g erzeugten Hauptideal (g). Wir haben also $I \subseteq (g)$, und da offenbar $(g) \subseteq I$ gilt, ist $I = (g)$ wie behauptet. Ist g_1 ebenfalls ein normiertes Polynom vom kleinsten möglichen Grad in I, so ist nach dem eben gezeigten $g_1 = gh$ mit $h \in K[X]$, wegen $\deg(g) = \deg(g_1)$ muss dann

$\deg(h) = 0$ sein, d.h., $h = c \in K$ ist konstant. Da g und g_1 normiert sind, ist $h = 1$, also $g = g_1$, die Eindeutigkeitsaussage ist also auch klar. □

Definition und Korollar 13.4. *Sei S eine K-Algebra, $s \in S$,*

$$I_s := \{f \in K[X] \mid f(s) = 0\} \neq \{0\}$$

das Verschwindungsideal *(der* Annullator*) von s in $K[X]$. (Dabei ist $f(s) = \sum_{j=1}^{n} c_j s^j$ (mit $s^0 = 1_S$) für $f = \sum_{j=1}^{n} c_j X^j \in R[X]$.)*
Dann ist $I_s = (g)$, wo g das normierte Polynom kleinsten Grades in I_s ist.
g heißt das Minimalpolynom *von s über K; es teilt alle Polynome $f \in K[X]$ mit $f(s) = 0$.*

Beweis. Klar. □

Bemerkung. Man

Lemma 13.5. *Sei $A \in M_n(K)$. Dann gilt:*

a) *Es gibt $0 \neq f \in K[X]$ mit $f(A) = 0$. Das Minimalpolynom von A über K wird mit $\mu_{A,K}$ oder μ_A bezeichnet.*
b) *Ist $\mu_{A,K}$ das Minimalpolynom von A über K, $L \supseteq K$ ein Erweiterungskörper, so ist $\mu_{A,L} = \mu_{A,K}$.*

Beweis. a) In dem n^2-dimensionalen K-Vektorraum $M_n(K)$ können die $n^2 + 1$ Elemente $E_n = A^0, A, A^2, \ldots, A^{n^2}$ nicht linear unabhängig sein. Ist $\sum_{i=0}^{n^2} c_i A^i = 0_n$ eine nichttriviale lineare Relation zwischen ihnen, so ist $f := \sum_{i=0}^{n^2} c_i X^i \in K[X]$ ein von 0 verschiedenes Polynom in $K[X]$ mit $f(A) = 0_n$.
b) Zunächst ist zu bemerken, dass es wegen $M_n(K) \subseteq M_n(L)$ möglich ist, A auch als Element der L-Algebra $M_n(L)$ aufzufassen und es daher sinnvoll ist, vom Minimalpolynom von A über L zu sprechen.
Hat die Matrix A^i die Koeffizienten $a_{jk}^{(i)}$, so ist für $m \in \mathbb{N}_0, c_0, \ldots, c_m \in L$ die Matrixgleichung $\sum_{i=0}^{m} c_i A^i = 0_n$ nichts anderes als ein lineares Gleichungssystem aus den n^2 Gleichungen

$$\sum_{i=0}^{m} c_i a_{jk}^{(i)} = 0 \quad (1 \leq j, k \leq n)$$

mit Koeffizienten $a_{jk}^{(i)} \in K$. Es gibt also genau dann ein Polynom $0 \neq f \in L[X]$ vom Grad $\leq m$, wenn dieses lineare Gleichungssystem in $m + 1$ Variablen eine nichttriviale Lösung $\begin{pmatrix} c_0 \\ \vdots \\ c_m \end{pmatrix} \in L^{m+1}$ hat.
Da ein homogenes lineares Gleichungssystem mit Koeffizienten in K genau dann im Oberkörper L eine nichttriviale Lösung hat, wenn es

bereits in K eine nichttriviale Lösung hat, sieht man, dass das Minimalpolynom von A über K den gleichen Grad hat wie das Minimalpolynom von f über L, wegen $K[X] \subseteq L[X]$ müssen beide also wie behauptet gleich sein. $\square$

Beispiel.

- Ist $A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \in M(3 \times 3, \mathbb{R})$, so ist $\mu_A = (X-1)(X-2)$.
- Ist $A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \in M(3 \times 3, \mathbb{R})$,
  so ist $\mu_{A_2} = (X-1)(X-2)(X-3)$.
- Ist $A_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix} \in M(3 \times 3, \mathbb{R})$, so ist $\mu_{A_3} = (X-1)(X-2)^2$.

Satz 13.6. (Cayley-Hamilton) *Für das charakteristische Polynom χ_A von $A \in M_n(K)$ gilt:*

$$\chi_A(A) = 0_n.$$

Insbesondere ist das Minimalpolynom μ_A der Matrix A ein Teiler des charakteristischen Polynoms χ_A.

Beweis. Setzt man A in das Polynom $\chi_A(X) = \det(XE_n - A) \in K[X]$ ein, so erhält man dasselbe Ergebnis, wie wenn man in der Matrix

$$XE_n - A = \begin{pmatrix} X - a_{11}X^0 & \dots & -a_{1n}X^0 \\ \vdots & & \vdots \\ -a_{n1}X^0 & \dots & X - a_{nn}X^0 \end{pmatrix} \in M_n(K[X])$$

die Variable X durch A (also insbesondere X^0 durch $A^0 = E_n$) ersetzt und anschließend die Determinante der so erhaltenen Matrix

$$C := \begin{pmatrix} A - a_{11}E_n & \dots & -a_{1n}E_n \\ \vdots & & \vdots \\ -a_{n1}E_n & \dots & A - a_{nn}E_n \end{pmatrix}$$

berechnet. Die Einträge dieser Matrix C sind Elemente des kommutativen Teilrings

$$K[A] := \{\sum_{i=0}^{m} c_i A^i \mid m \in \mathbb{N}_0, c_0, \dots, c_m \in K\} \subseteq M_n(K)$$

des Matrizenrings $M_n(K)$, insbesondere ist also $\chi_A(A) = \det(C) \in K[A] \subseteq M_n(K)$ selbst wieder eine $n \times n$-Matrix.
Der ganz einfache Beweisversuch
$\chi_A(A) = \det(AE_n - A) = \det(A - A) = 0$
geht also in die falsche Richtung.

Stattdessen gehen wir wie folgt vor:
Mit $C = (c_{ij})$ wie oben gilt offenbar für jedes j für die Standardbasisvektoren $\mathbf{e}_i$ von K^n die Gleichung

$$\begin{aligned}\sum_{i=1}^n c_{ij}\mathbf{e}_i &= \sum_{i=1}^n (\delta_{ij}A - a_{ij}E_n)\mathbf{e}_i \\ &= A\mathbf{e}_j - \sum_{i=1}^n a_{ij}\mathbf{e}_i \\ &= \mathbf{0}.\end{aligned}$$

Wir multiplizieren diese Gleichung mit dem jk-Koeffizienten $\tilde{c}_{jk}$ der Komplementärmatrix (Satz 7.20) $\tilde{C}$ von C, summieren über j und erhalten für $1 \leq k \leq n$

$$\begin{aligned}\mathbf{0} &= \sum_{j=1}^n \tilde{c}_{jk} \sum_{i=1}^n c_{ij}\mathbf{e}_i \\ &= \sum_{i=1}^n (\sum_{j=1}^n c_{ij}\tilde{c}_{jk})\mathbf{e}_i \\ &= \sum_{i=1}^n \delta_{ik}\det(C)\mathbf{e}_i \\ &= \chi_A(A)\mathbf{e}_k,\end{aligned}$$

da $C \cdot \tilde{C} = \det(C)E_n$ nach Satz 7.20 gilt und $\det(C) = \chi_A(A)$ ist. Das heißt aber, dass Multiplikation mit der Matrix $\chi_A(A) \in M_n(K)$ die Nullabbildung von K^n in sich liefert, dass also $\chi_A(A) = 0_n$ gilt. □

Korollar 13.7. *Das Minimalpolynom μ_A von $A \in M_n(K)$ hat die gleichen Nullstellen wie das charakteristische Polynom χ_A.*

Beweis. Nach dem vorigen Satz ist klar, dass das Minimalpolynom μ_A ein Teiler von χ_A in $K[X]$ ist und daher alle Nullstellen von μ_A auch Nullstellen von χ_A sind.
Ist umgekehrt $\lambda \in K$ eine Nullstelle von χ_A, so ist λ ein Eigenwert von A, es gibt also einen Vektor $\mathbf{x} \neq \mathbf{0}$ in K^n mit $A\mathbf{x} = \lambda\mathbf{x}$; es gilt dann offenbar auch $A^j\mathbf{x} = \lambda^j\mathbf{x}$ für alle $j \in \mathbb{N}_0$. Ist $\mu_A = \sum_{i=1}^r c_i X^i$, so haben wir wegen $\mu_A(A) = 0_n$ daher

$$\begin{aligned}\mathbf{0} &= \mu_A(A)\mathbf{x} \\ &= \sum_{i=1}^r c_i A^i\mathbf{x} \\ &= \sum_{i=1}^r c_i \lambda^i\mathbf{x} \\ &= \mu_A(\lambda)\mathbf{x}\end{aligned}$$

und daher $\mu_A(\lambda) = 0$. $\square$

Lemma 13.8. *Sind $A, A' \in M_n(K)$ zueinander ähnliche Matrizen (also $A' = S^{-1}AS$ mit $S \in \mathrm{GL}_n(K)$), so haben sie das gleiche Minimalpolynom.*

Beweis. Übung, man zeige zunächst $(S^{-1}AS)^j = S^{-1}A^jS$. $\square$

Lemma 13.9. *Das Minimalpolynom des Endomorphismus $f \in \mathrm{End}(V)$ ist gleich dem Minimalpolynom seiner Matrix A bezüglich einer beliebigen Basis von V.*

Beweis. Klar. $\square$

Satz 13.10. *Sei $f \in \mathrm{End}(V)$ so, dass das charakteristische Polynom von f über K als*

$$\chi_f = \prod_{i=1}^{r}(X - \beta_i)^{e_i}$$

mit paarweise verschiedenen β_i und $e_i \in \mathbb{N}$ in Linearfaktoren zerfällt. Dann gilt:
f (bzw. die zugehörige Matrix $A \in M_n(K)$) ist genau dann diagonalisierbar, wenn das Minimalpolynom μ_f ($= \mu_A$) nur einfache Nullstellen hat, wenn also

$$\mu_f = \prod_{i=1}^{r}(X - \beta_i)$$

gilt.

Beweis. Ist f diagonalisierbar, so ist V die direkte Summe der Eigenräume V_i zu den Eigenwerten β_i, und jeder Eigenraum V_i ist f-invariant (d.h., $f(V_i) \subseteq V_i$). Da offenbar $(f - \beta_i Id_V)|_{V_i} = 0$ gilt, ist

$$\prod_{i=1}^{r}(f - \beta_i Id_V)|_{V_j} = 0$$

für alle $1 \leq j \leq r$, d.h., f wird von dem Polynom $\prod_{i=1}^{r}(X - \beta_i)$ annulliert. Da nach Korollar 13.7 das Minimalpolynom μ_f von f durch alle $X - \beta_i$ teilbar ist, folgt $\mu_f = \prod_{i=1}^{r}(X - \beta_i)$ wie behauptet.
Die Gegenrichtung zeigen wir durch Induktion nach der Anzahl r der verschiedenen Eigenwerte von f, indem wir versuchen $V = V_1 \oplus W$ mit einem f-invarianten Unterraum W zu schreiben und die Induktionsannahme auf W anzuwenden.
Wir nehmen also an, dass μ_f nur einfache Nullstellen hat, dass also $\mu_f = \prod_{i=1}^{r}(X - \beta_i)$ gilt. Ist $r = 1$, so ist $f = \beta_1 Id_V$, f ist also diagonalisierbar.
Sei jetzt $r \geq 2$ und die Behauptung für Endomorphismen mit weniger als r verschiedenen Eigenwerten bereits bewiesen (Induktionsannahme).

Um den gesuchten f invarianten Unterraum W von V zu konstruieren, betrachten wir $g := \prod_{i=2}^{r}(f - \beta_i Id_V)$. Nach Satz 8.4 können wir das Polynom $p = \prod_{i=2}^{r}(X - \beta_i)$ mit Rest durch $X - \beta_1$ teilen und erhalten $p = q(X - \beta_1) + r$ mit $q, r \in K[X]$ und $r = 0$ oder $\deg(r) < \deg(X - \beta_1) = 1$. Dabei ist $r = 0$ nicht möglich, denn sonst hätte man nach Einsetzen von β_1 für X die Gleichung $\prod_{i=2}^{r}(\beta_1 - \beta_i) = q(\beta_1) \cdot 0 = 0$, im Widerspruch dazu, dass nach Voraussetzung $\beta_i \neq \beta_1$ für $i \geq 2$ gilt. Der hier auftretende Rest r ist also ein von 0 verschiedenes konstantes Polynom $r \in K, r \neq 0$.
Damit haben wir

$$\frac{p}{r} - (X - \beta_1)\frac{q}{r} = 1(= X^0),$$

nach Einsetzen von f für X ergibt sich also

$$\frac{g}{r} - (f - \beta_1 Id_V)\frac{q(f)}{r} = f^0 = Id_V.$$

Wir können daher jeden Vektor $v \in V$ als $v = v_1 + v_2$ mit $v_1 = g(\frac{v}{r}) \in \mathrm{Im}(g)$ und $v_2 = (f - \beta_1 Id_V)(q(f)(\frac{v}{r})) \in \mathrm{Im}(f - \beta_1 Id_V)$ schreiben.
Wegen $\prod_{i=1}^{r}(f - \beta_i Id_V) = 0$ ist $\mathrm{Im}(g) \subseteq \mathrm{Ker}(f - \beta_1 Id_V)$, es gilt also

$$V = \mathrm{Im}(f - \beta_1 Id_V) + \mathrm{Ker}(f - \beta_1 Id_V)$$

und wegen $\dim(\mathrm{Im}(f - \beta_1 Id_V)) + \dim(\mathrm{Ker}(f - \beta_1 Id_V) = \dim(V)$ folgt, dass die Summe direkt ist, dass also $\mathrm{Im}(f - \beta_1 Id_V) \cap \mathrm{Ker}(f - \beta_1 Id_V) = \mathbf{0}$ gilt.
Der Teilraum $W = \mathrm{Im}(f - \beta_1 Id_V)$ von V ist f-invariant, und $f|_W \in \mathrm{End}(W)$ hat wegen $W \cap \mathrm{Ker}(f - \beta_1 Id_V) = \{\mathbf{0}\}$ nicht den Eigenwert β_1. Ferner ist $\chi_f = \chi_{f|_{V_1}} \cdot \chi_{f|_W}$, also muss $\chi_{f|_{V_1}} = (X - \beta_1)^{e_1}$ und $\chi_{f|_W} = \prod_{i=2}^{r}(X - \beta_i)_i^e$ sein. Ferner teilt $\mu_{f|_W}$ sowohl $\mu_f = \prod_{i=1}^{r}(X - \beta_i)$ als auch $\chi_{f|_W}$, ist also gleich $\prod_{i=2}^{r}(X - \beta_i)$.
Auf den Endomorphismus $f|_W$ von W können wir also die Induktionsannahme anwenden, er ist also diagonalisierbar.
Der Teilraum $\mathrm{Ker}(f - \beta_1 Id_V) = V_1$ ist der Eigenraum von f zum Eigenwert β_1; er ist ebenfalls f-invariant und $f|_{V_1} = \beta_1 Id_{V_1}$ ist diagonalisierbar.
Wegen $V = W \oplus V_1$ ist dann auch f diagonalisierbar (man wähle in jedem dieser Teilräume eine Basis aus Eigenvektoren von f, die Vereinigung dieser Basen ist dann eine Basis aus Eigenvektoren von f für V).

$\square$

14. Jordansche Normalform

In diesem Abschnitt ist stets K ein Körper, V ein n-dimensionaler K-Vektorraum mit Basis $\mathcal{B}$, $f \in \mathrm{End}(V)$ mit Matrix $A = M_{\mathcal{B}}(f)$.
Ziel dieses Abschnitts ist der Beweis des folgenden Satzes:

Satz 14.1. (Jordan'sche Normalform) *Sei $f \in \mathrm{End}(V)$ so, dass das charakteristische Polynom χ_f in Linearfaktoren zerfällt, $\chi_f = \prod_{i=1}^{r}(X-\beta_i)^{e_i}$ mit paarweise verschiedenen β_i und $e_i \in \mathbb{N} \setminus \{0\}$. Dann gibt es eine Basis $\mathcal{B}$ von V, bezüglich der die Matrix von f Blockgestalt*

$$\begin{pmatrix} \beta_1 E_{e_1} + N_1 & & 0 \\ & \ddots & \\ 0 & & \beta_r E_{e_r} + N_r \end{pmatrix},$$

hat, wobei jedes N_i $(1 \le i \le r)$ für ein $d = d_i$ die Gestalt

$$N = \begin{pmatrix} J_d & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & J_d & & & & & & & \\ & & & J_{d-1} & & & & & & \\ & & & & \ddots & & & 0 & & \\ & & & & & J_{d-1} & & & & \\ & & 0 & & & & \ddots & & & \\ & & & & & & & J_1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & J_1 \end{pmatrix}$$

mit jeweils s_ν Jordan-Kästchen

$$J_\nu = \begin{pmatrix} 0 & 1 & & & & \\ & \cdot & \cdot & & & \\ & & \cdot & \cdot & 0 & \\ & 0 & & \cdot & \cdot & \\ & & & & \cdot & 1 \\ & & & & & 0 \end{pmatrix}$$

der Größe $\nu \times \nu$ in der Diagonale $(1 \le \nu \le d)$ hat.

Ein Block der Gestalt

$$\begin{pmatrix} \beta_i & 1 & & & & \\ & \cdot & \cdot & & & \\ & & \cdot & \cdot & 0 & \\ & 0 & & \cdot & \cdot & \\ & & & & \cdot & 1 \\ & & & & & \beta_i \end{pmatrix} \in M_\nu(K)$$

(oder dessen Transponierte) heißt auch Jordanblock *der Größe ν zu β_i.*

Diese Blockgestalt heißt die Jordan'sche Normalform *der Matrix von f; sie ist bis auf die Anordnung der Blöcke auf der Diagonalen eindeutig bestimmt.*

Zwei Matrizen $A, B \in M_n(K)$ mit in Linearfaktoren zerfallendem charakteristischen Polynom sind genau dann zueinander ähnlich (konjugiert), wenn sie (bis auf die Anordnung der Blöcke) die gleiche Jordan'sche Normalform haben.
Insbesondere gilt: Die Matrix $A \in M_n(K)$ ist genau dann diagonalisierbar, wenn ihr charakteristisches Polynom in ein Produkt von (nicht notwendig verschiedenen) Linearfaktoren zerfällt und ihre Jordan'sche Normalform Diagonalgestalt hat.

Korollar 14.2. *Für V, f wie im Satz gilt:*

a) *Die Vielfachheit der Nullstelle β_i des Minimalpolynoms von f ist das Maximum der d_i.*
b) *Die geometrische Vielfachheit des Eigenwerts β_i von f ist gleich der Anzahl der Jordanblöcke zum Eigenwert β_i.*

Bemerkung. Ist $K = \mathbb{C}$ (oder ein anderer algebraisch abgeschlossener Körper), so zerfällt jedes nicht konstante Polynom über K in Linearfaktoren, der Satz gilt dann also für beliebige Endomorphismen bzw. Matrizen.

Es gibt zwei grundsätzlich verschiedene Beweise für diesen Satz, die im folgenden beide durchgeführt werden: Der eine betrachtet die Operation des Polynomrings $K[X]$ auf dem Vektorraum V, die durch $p * v := p(f)(v)$ gegeben ist und benutzt Sätze über den Polynomring, die bei einem weiteren Studium der Algebra ohnehin betrachtet werden. Dieser Ansatz liefert auch eine allgemeinere Aussage, bei der man nicht darauf angewiesen ist, dass das charakteristische Polynom über dem Körper K in Linearfaktoren zerfällt und die daher für alle Endomorphismen eines K-Vektorraums über einem beliebigen Körper K gültig ist.
Der andere Ansatz geht quasi mit Bordmitteln vor und untersucht explizit die Eigenwerte und Eigenvektoren, zunächst für den Fall, dass 0 der einzige Eigenwert ist. Dieser Ansatz ist weniger abstrakt und daher bei vielen Studierenden zumindest zunächst beliebter. Er funktioniert aber nur, wenn man genug Eigenwerte hat, und ist daher auf den Fall $K = \mathbb{C}$ (bzw. von Endomorphismen, deren charakteristisches Polynom in Linearfaktoren zerfällt) beschränkt; er liefert auch weniger Einsicht in die Struktur des Problems. Anders gesagt: Mit dem zweiten konkreteren Ansatz kommt man für Rechenzwecke schon ziemlich weit, wer die Sache aber wirklich verstehen will, sollte sich jetzt oder auch später die Mühe machen, den abstrakten ersten Ansatz zu studieren und die Eleganz und Schönheit des allgemeinen Arguments zu würdigen.
Für diesen erstgenannten Ansatz definieren wir:

Definition 14.3. *Sei R ein kommutativer Ring mit Einselement $1 = 1_R$. Eine kommutative (=abelsche) Gruppe M mit einer Verknüpfung (Skalarmultiplikation) $R \times M \to M$ heißt ein R-Modul, wenn für diese Verknüpfung die Forderungen der Axiome SM1 bis SM4 aus der Definition eines Vektorraums (siehe Definition 3.1) erfüllt sind.*
Eine Untergruppe U von M heißt ein R-Untermodul, wenn $a \cdot u \in U$ für alle $a \in R, u \in U$ gilt.

Lemma 14.4. *Ist V ein K-Vektorraum und $f \in \mathrm{End}(V)$, so wird durch $(P, v) \mapsto P *_f v = P(f)(v)$ (für $P \in K[X]$ und $v \in V$) eine R-Modul-Struktur auf V definiert. Ein Unterraum U von V ist genau dann f-invariant, wenn U ein $K[X]$-Untermodul von V ist.*

Beweis. Beide Aussagen rechnet man leicht nach. □

Bemerkung. Fasst man in der Basis von V, bezüglich der der Endomorphismus f die in Satz 14.1 angegebene Gestalt hat, die Basisvektoren, die zu einem Jordanblock gehören, zusammen und betrachtet den von ihnen erzeugten Unterraum U, so ist dieser offenbar f-invariant, also ein $K[X]$-Untermodul von V. Die Jordan'sche Normalform liefert also eine Zerlegung von V in eine direkte Summe von besonders einfachen $K[X]$-Untermoduln.

Wir beginnen jetzt aber zunächst mit dem zweiten Ansatz, beginnen also, den Satz mit „Bordmitteln“ zu beweisen:

Lemma 14.5. (Fitting) *Sei $g \in \mathrm{End}(V)$,*

$$d := \min\{\ell \in \mathbb{N} \mid \mathrm{Ker}(g^\ell) = \mathrm{Ker}(g^{\ell+1})\},$$

sei $\chi_g = X^r p$ mit $p \in K[X]$, $X \not| p$. Dann gilt:

a) $d = \min\{\ell \in \mathbb{N} \mid \mathrm{Im}(g^\ell) = \mathrm{Im}(g^{\ell+1})\}$.
b) *Für alle $i \in \mathbb{N}$ ist $\mathrm{Ker}(g^{d+i}) = \mathrm{Ker}(g^d)$, $\mathrm{Im}(g^{d+i}) = \mathrm{Im}(g^d)$.*
c) *$U := \mathrm{Ker}(g^d)$ und $W := \mathrm{Im}(g^d)$ sind g-invariante Unterräume, es gilt $(g|_U)^d = 0$, und $g|_W$ ist bijektiv (ist also ein Automorphismus von W).*
d) *Das Minimalpolynom $\mu_{g|_U}$ von $g|_U$ ist X^d.*
e) *Es ist $V = U \oplus W$ mit $\dim U = r \geq d$.*

Beweis. Zunächst ist klar, dass $\mathrm{Ker}(g^\ell) \subseteq \mathrm{Ker}(g^{\ell+1})$ und $\mathrm{Im}(g^{\ell+1}) \subseteq \mathrm{Im}(g^\ell)$ für alle $\ell \in \mathbb{N}_0$ gilt. Wegen der Dimensionsformel für Kern und Bild ist ferner klar, dass $\mathrm{Ker}(g^\ell) = \mathrm{Ker}(g^{\ell+1})$ und $\mathrm{Im}(g^{\ell+1}) = \mathrm{Im}(g^\ell)$ zueinander äquivalent sind. Damit folgen a) und b), Aussage c) ist ist dann unmittelbar klar. Für d) müssen wir zeigen, dass $(g|_U)^{d-1} \neq 0$ ist. Das ist aber klar, da sonst $\mathrm{Ker}(g^d) \subseteq \mathrm{Ker}(g^{d-1})$ im Widerspruch zur Minimalität von d gelten würde.
Für e) schließlich sei $v \in U \cap W$, also $v = g^d(x)$ mit $x \in V$ und $g^d(v) = \mathbf{0}$. Dann ist $\mathbf{0} = g^{2d}(x)$, also $x \in \mathrm{Ker}(g^{2d}) = \mathrm{Ker}(g^d)$, also $v = g^d(x) = \mathbf{0}$. Also ist $U \cap W = \{\mathbf{0}\}$, und aus der Dimensionsformel für Kern und Bild folgt $V = U \oplus W$.

Dass $\dim(U) \geq d$ gilt, folgt aus der Definition von d. Für das charakteristische Polynom χ_g gilt $\chi_g = \chi_{g|_U} \cdot \chi_{g|_W} = X^d \cdot \chi_{g|_W}$ mit $X \nmid \chi_{g|_W}$, da $g|_W$ injektiv ist, also nicht den Eigenwert 0 hat. Damit sehen wir auch $d = r$. □

Definition 14.6. *Sei V ein K-Vektorraum. Der Endomorphismus $f \in \mathrm{End}(V)$ heißt* nilpotent*, wenn es ein $m \in \mathbb{N}$ gibt mit $f^m = 0$. Das kleinste derartige m heißt dann der* Nilpotenzindex *von f, man sagt auch, f sei m-*stufig nilpotent*.*

Bemerkung. f ist genau dann m-stufig nilpotent, wenn sein Minimalpolynom gleich X^m ist.

Beispiel. Multiplikation eines Vektors aus $\mathbb{C}^d$ mit der Matrix J_d aus Satz 14.1 ist eine nilpotente lineare Abbildung vom Index d.

Wir können das Lemma von Fitting jetzt anwenden, um ein beliebiges $f \in \mathrm{End}(V)$ durch Rückführung auf den nilpotenten Fall zu behandeln:

Satz 14.7. (Hauptraumzerlegung) *Sei $f \in \mathrm{End}(V)$ so, dass das charakteristische Polynom χ_f in Linearfaktoren zerfällt:*
$\chi_f = \prod_{i=1}^r (X - \beta_i)^{e_i}$ mit paarweise verschiedenen β_i und $e_i \in \mathbb{N} \setminus \{0\}$.
Sei $V_i := \mathrm{Ker}(f - \beta_i \mathrm{Id})^{e_i}$ der Hauptraum *zum Eigenwert β_i von f. Dann gilt:*

a) $V = \bigoplus_{i=1}^r V_i$
b) *Die V_i sind f-invariante Teilräume mit $\dim(V_i) = e_i$.*
c) *Es ist $f = f_d + f_n$ mit $f_d, f_n \in \mathrm{End}(V)$, f_d diagonalisierbar, f_n nilpotent und $f_d f_n = f_n f_d$.*

Korollar 14.8. *Setzt man eine Basis $\mathcal{B}$ von V aus Basen der Haupträume V_i zusammen, so hat f bezüglich $\mathcal{B}$ die Blockmatrix*

$$\begin{pmatrix} \beta_1 E_{e_1} + N_1 & & 0 \\ & \ddots & \\ 0 & & \beta_r E_{e_r} + N_r \end{pmatrix},$$

wo die $N_i \in M(e_i \times e_i, K)$ nilpotente Matrizen mit $N_i^{e_i} = 0$ sind.

Beweis des Satzes. Induktion nach der Anzahl r der verschiedenen Eigenwerte von f. Ist $r = 1$, so sind a) und b) trivial, und in c) setzt man $f_d = \beta_1 \mathrm{Id}_V$ und $f_n = f - f_d$.
Ist $r > 1$ und die Behauptung bewiesen für Endomorphismen mit weniger als r verschiedenen Eigenwerten, so wenden wir auf $g := f - \beta_1 \mathrm{Id}_V$ das Lemma von Fitting an und erhalten $V = V_1 \oplus W$, wobei $f|_W$ nur noch die Eigenwerte $\beta_2, \ldots, \beta_r$ hat. Die Behauptung folgt dann aus der Induktionsannahme. □

Satz 14.9. (Normalform für nilpotente Endomorphismen) *Sei $g \in \mathrm{End}(V)$ nilpotent vom Index d. Dann gibt es eindeutig bestimmte $s_1, \ldots, s_d \in \mathbb{N}$ mit*

$$d \cdot s_d + (d-1)s_{d-1} + \cdots + s_1 = \dim(V) = n$$

und eine Basis $\mathcal{B}$ von V, bezüglich der g die Blockmatrix

$$\begin{pmatrix} J_d & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & J_d & & & & & & & & \\ & & & J_{d-1} & & & & & & & \\ & & & & \ddots & & & & 0 & & \\ & & & & & J_{d-1} & & & & & \\ & & & 0 & & & \ddots & & & & \\ & & & & & & & J_1 & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & J_1 & \end{pmatrix}$$

mit jeweils s_ν Jordan-Kästchen

$$J_\nu = \begin{pmatrix} 0 & 1 & & & & \\ & \cdot & \cdot & & & \\ & & \cdot & \cdot & 0 & \\ & 0 & & \cdot & \cdot & \\ & & & & \cdot & 1 \\ & & & & & 0 \end{pmatrix}$$

der Größe $\nu \times \nu$ in der Diagonale ($1 \leq \nu \leq d$) hat.
Durch Umnummerieren der Basisvektoren lässt sich hier auch

$${}^tJ_\nu = \begin{pmatrix} 0 & & & & & \\ 1 & 0 & \cdot & \cdot & & \\ 0 & 1 & 0 & \cdot & \cdot & \\ & 0 & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot & \\ & & & 0 & 1 & 0 \end{pmatrix} \in M(\nu \times \nu, K)$$

erreichen.

Beweis. (Skizze, siehe auch das Buch von Fischer) Man benutzt die aufsteigende *Filtrierung* von V durch die g-invarianten Unterräume $U_\nu := \mathrm{Ker}(g^\nu)$ (mit $U_\nu \subseteq U_{\nu+1}$) für $0 \leq \nu \leq d$, in der die Inklusionen $U_\nu \subseteq U_{\nu+1}$ strikte Inklusionen sind (nach dem Lemma von Fitting) und in der $g^{-1}(U_{\nu-1}) = U_\nu$ für $1 \leq \nu \leq d$ gilt. Man wählt dann W_d als einen zu U_{d-1} komplementären Unterraum in $U_d = V$, stellt fest, dass $g(W_d) \subseteq U_{d-1}$ mit $U_{d-2} \cap g(W_d) = \{\mathbf{0}\}$ gilt und ergänzt $g(W_d)$ zu einem zu U_{d-2} komplementären Unterraum von U_{d-1}. Indem man

dieses Verfahren iteriert erhält man direkte Summenzerlegungen

$$V = U_d = U_{d-1} \oplus W_d = U_{d-2} \oplus W_{d-1} \oplus W_d = \cdots = W_1 \oplus \cdots \oplus W_d$$

in Teilräume W_j mit $U_j = U_{j-1} \oplus W_j$, für die $g|_{W_j}$ für $j > 1$ injektiv ist und W_j nach $W_{j-1} \subseteq U_{j-1}$ mit $g(W_j) \cap U_{j-2} = \{\mathbf{0}\}$ abbildet.

Die gesuchte Basis erhält man dann, indem man eine Basis $w_1^{(d)}, \ldots, w_{s_d}^{(d)}$ von W_d wählt, die (linear unabhängigen) Vektoren $g(w_1^{(d)}), \ldots, g(w_{s_d}^{(d)})$ durch $w_1^{(d-1)}, \ldots, w_{s_{d-1}}^{(d-1)}$ zu einer Basis von W_{d-1} ergänzt und so fortfährt bis schließlich die Bilder

$$g^{d-1}(w_1^{(d)}), \ldots, g^{d-1}(w_{s_d}^{(d)}), \ldots, g(w_1^{(2)}), \ldots, g(w_{s_2}^{(2)})$$

aller Basisvektoren von W_2 unter g durch $w_1^{(1)}, \ldots, w_{s_1}^{(1)}$ zu einer Basis von $W_1 = U_1$ ergänzt werden. Dabei ist W_j isomorph zum Faktorraum U_j/U_{j-1} und hat Dimension $\sum_{\nu=j}^{d} s_\nu$.
Ordnet man diese Basisvektoren in der Reihenfolge

$$\begin{gathered} g^{d-1}(w_1^{(d)}), g^{d-2}(w_1^{(d)}), \ldots, w_1^{(d)}, \ldots, g^{d-1}(w_{s_d}^{(d)}), \ldots, w_{s_d}^{(d)}, \\ g^{d-2}(w_1^{(d-1)}), \ldots, w_1^{(d-1)}, \ldots, w_1^{(1)}, \ldots, w_{s_1}^{(1)} \end{gathered}$$

an, so ist die Matrix von f in der gewünschten Gestalt.

Sei umgekehrt $v_1, \ldots, v_n$ eine Basis von V, bezüglich der die Matrix von f die angegebene Gestalt hat.
Dann spannen die s_d Vektoren

$$v_d, v_{2d}, \ldots, v_{s_d d}$$

einen Raum W_d auf, die s_{d-1} Vektoren

$$v_{s_d d+d-1}, v_{s_d d+2(d-1)}, \ldots, v_{s_d d+s_{d-1}(d-1)}$$

spannen zusammen mit den Vektoren

$$g(v_d) = v_{d-1}, \ldots, g(v_{s_d d}) = v_{s_d d-1}$$

einen Raum W_{d-1} auf, und so fort, bis die s_1 Vektoren

$$v_{1+\sum_{j=2}^{d} j s_j}, \ldots, v_{s_1+\sum_{j=2}^{d} j s_j}$$

zusammen mit den Bildern aller Basisvektoren von W_2 unter g den Raum $W_1 = \mathrm{Ker}(g)$ aufspannen.
Man hat dann genau wie oben, dass W_j isomorph zum Faktorraum U_j/U_{j-1} ist und Dimension $\sum_{\nu=j}^{d} s_\nu$ hat. Insbesondere sieht man, dass man die Anzahl s_ν der Kästchen J_ν aus den von Basiswahlen unabhängigen Zahlen $\dim(U_j/U_{j-1})$ für $1 \leq j \leq d$ berechnen kann, die Normalform also in der Tat eindeutig ist. □

Beweis von Satz 14.1. Wir wählen für jeden der Haupträume $V_i = \mathrm{Ker}(f - \beta_i \mathrm{Id}_V)^{e_i}$ von f gemäß Satz 14.9 eine Basis $\mathcal{B}_i$, bezüglich der der nilpotente Endomorphismus $(f - \beta_i \mathrm{Id}_V)|_{V_i}$ die dort angegebene Matrix

in Normalgestalt hat. Bezüglich der aus diesen Basen der V_i zusammengesetzten Basis von $V = \oplus_i V_i$ hat dann f die im Satz angegebene Gestalt (mit den β_i auf den Diagonalen der Jordanblöcke). $\square$

Für den anderen Ansatz ist es zweckmäßig, sich zunächst noch einen Überblick über ein paar grundlegende Eigenschaften des Polynomrings $K[X]$ zu verschaffen

Definition und Lemma 14.10. *Ein Polynom $q \in K[X]$, das nicht konstant ist (also Grad ≥ 1 hat), heißt irreduzibel, wenn gilt:*
Ist $q = h_1 h_2$ mit $h_1, h_2 \in K[X]$, so ist h_1 oder h_2 konstant.
Ist $q \in K[X]$ irreduzibel, so gilt:

a) *Ist $h \in K[X]$ mit $q \nmid h$, so gibt es $f_1, f_2 \in K[X]$ mit $f_1 q + f_2 h = 1$, das von den Polynomen q und h erzeugte Ideal $(q, h) = \{f_1 q + f_2 h \mid g_1, g_2 \in K[X]\}$ ist also gleich $K[X]$. Man sagt dann, q und h seien teilerfremd oder hätten größten gemeinsamen Teiler 1 und schreibt* $\mathrm{ggT}(q, h) = 1$.
b) *Sind $h_1, h_2 \in K[X]$ mit $q \mid h_1 h_2$, so ist $q \mid h_1$ oder $q \mid h_2$ (man sagt, q sei ein Primelement des Ringes $K[X]$).*
c) *Ist q normiert und $q_2 \neq q$ ein weiteres normiertes irreduzibles Polynom, so sind q und q_2 teilerfremd.*

Beweis. a): Sei $I := (q, h) := \{qf_1 + hf_2 \mid h_1, h_2 \in K[X]\}$ das von den Polynomen q und h erzeugte Ideal. Da $K[X]$ ein Hauptidealring ist (d.h., jedes Ideal ist ein Hauptideal), gibt es ein $g \in K[X]$, das I erzeugt, für das also $I = \{gf \mid f \in K[X]\}$ ist und für das daher $g \mid q, g \mid h$ gilt. Da q irreduzibel ist, folgt aus $g \mid q$, dass g konstant ist oder $g = cq$ mit $c \in K, c \neq 0$ gilt. Wäre $g = cq$ mit $c \in K$, so wäre auch $q = c^{-1} g$ im Widerspruch zur Annahme $q \nmid h$ ein Teiler von h.
Im verbleibenden Fall $g = c \in K, c \neq 0$ ist aber $c^{-1} g = 1 \in I$, besitzt also eine Darstellung $f_1 q + f_2 h = 1$ wie behauptet.
b): Ist $q \nmid h_1$, so finden wir nach a) Polynome g_1, g_2 mit $g_2 q + g_1 h_1 = 1$. Wir multiplizieren diese Gleichung mit h_2 und erhalten

$$g_2 q h_2 + g_1 h_1 h_2 = h_2.$$

Da auf der linken Seite beide Summanden durch q teilbar sind, muss auch h_2 durch q teilbar sein.
c): Da q_2 irreduzibel ist und die Möglichkeit $q_2 = cq$ mit $c \in K$ durch die Normiertheitsbedingung ausgeschlossen ist, kann q kein Teiler von q_2 sein, nach a) sind also q und q_2 teilerfremd. $\square$

Bemerkung. In der Situation von a) des vorigen Lemmas erhält man eine Darstellung $1 = f_1 q + f_2 h$ durch wiederholte Divison mit Rest wie folgt:
Man hat zunächst $h = p_1 q + r_1$ mit $0 \leq \deg(r_1) < \deg(q)$, wobei $r_1 = 0$ nicht möglich ist, da q kein Teiler von h ist. Ist $\deg(r_1) \neq 0$, so hat man weiter $q := r_0 = p_2 r_1 + r_2$ mit $0 \leq \deg(r_2) < \deg(r_1)$, und $r_2 \neq 0$,

da sonst r_1 ein nicht trivialer Teiler des irreduziblen Polynoms q wäre. Man fährt fort, indem man jeweils r_{j-1} mit Rest durch r_j dividiert, bis man bei einem konstanten Rest $r_n \neq 0$ ankommt. Die Gleichungskette $r_n = r_{n-2} - p_n r_{n-1} = r_{n-2} - p_n(r_{n-3} - p_{n-1}r_{n-2}) = r_{n-2}(1 + p_n p_{n-1}) - r_{n-3}p_n = ... = f_1' q + f_2' h$ liefert dann mit $f_1 = f_1'(r_n)^{-1}, f_2 = f_2'(r_n)^{-1}$ die gewünschte Darstellung.
Dieses Verfahren nennt man den euklidischen Algorithmus, es liefert allgemeiner bei beliebigen h_1, h_2 eine Darstellung des normierten Erzeugers g des Ideals (h_1, h_2) als Linearkombination von h_1 und h_2 und kann analog auch im Ring $\mathbb{Z}$ der ganzen Zahlen angewendet werden, um den größten gemeinsamen Teiler zweier ganzer Zahlen zu bestimmen. Der gesuchte Erzeuger g ist dann der letzte von 0 verschiedene Rest; Betrachten der entstehenden Gleichungskette zeigt, dass dieser alle vorherigen Reste und damit auch die ursprünglichen Polynome h_1, h_2 teilt. Er teilt damit alle Polynome $f_1h_1 + f_2h_2$ im von h_1, h_2 erzugten Ideal (h_1, h_2) und hat somit den kleinstmöglichen Grad in diesem Ideal, erzeugt es also.

Satz 14.11. *Im Polynomring $K[X]$ hat jedes nicht konstante normierte Polynom h eine (bis auf Reihenfolge) eindeutige Zerlegung*

$$h = \prod_{j=1}^{r} q_j^{e_j} \quad e_j \in \mathbb{N}, q_j \text{ irreduzibel und paarweise verschieden.}$$

(Man sagt, der Ring $K[X]$ sei faktoriell oder besitze eindeutige Primfaktorzerlegung).

Beweis. Dieser Satz wird in der Vorlesung EAZ allgemein für Hauptidealringe bewiesen. In der Vorlesung wurde ein Beweis für den hier vorliegenden Fall skizziert. □

Bemerkung. Ist $K = \mathbb{C}$, so hat (Fundamentalsatz der Algebra) jedes nicht konstante Polynom $h \in \mathbb{C}[X]$ eine Nullstelle $a \in \mathbb{C}$ und ist daher durch $X - a$ teilbar. Daraus folgt, dass die irreduziblen Polynome in $\mathbb{C}[X]$ genau die linearen Polynome $X - a$ sind. Die Primfaktorzerlegung in $\mathbb{C}[X]$ wird dann die schon früher betrachtete Zerlegung

$$h = \prod_{j=1}^{r} (X - a_j)^{e_j},$$

wo $a_1, \ldots, a_r$ die verschiedenen Nullstellen von h sind.

Satz 14.12. *Seien $p_1, p_2 \in K[X]$ teilerfremd (d.h., das von p_1 und p_2 erzeugte Ideal $(p_1, p_2) \in K[X]$ ist gleich $K[X]$), $f \in \mathrm{End}(V)$ mit $p_1(f)p_2(f) = 0$, seien*

$$V_1 = \mathrm{Ker}(p_1(f)),\ V_2 = \mathrm{Ker}(p_2(f)).$$

Dann sind V_1 und V_2 f-invariante Unterräume von V mit $V = V_1 \oplus V_2$.

Allgemeiner gilt: Sind $p_1, \ldots, p_r$ paarweise teilerfremde Polynome mit $p_1(f) \cdots p_r(f) = 0$, so hat man eine Zerlegung

$$V = V_1 \oplus \cdots \oplus V_r$$

in die f-invarianten Teilräume $V_i := \mathrm{Ker}(p_i(f))$ $(1 \leq i \leq r)$.
Sind die $p_j = q_j^{e_j}$ Potenzen verschiedener irreduzibler Polynome und hat das charakteristische Polynom χ_f von f die Primfaktorzerlegung $\chi_f = \prod_{j=1}^r q_j^{e_j}$, so heißt diese Zerlegung auch die Primärzerlegung *(oder verallgemeinerte Hauptraumzerlegung) von V bezüglich f.*

Beweis. Zunächst ist wegen $f \circ p_j(f) = p_j(f) \circ f$ klar, dass V_1 und V_2 invariant unter f sind.
Wir finden nun Polynome g_1, g_2 mit $g_1 p_1 + g_2 p_2 = 1$, also

$$g_1(f) \circ p_1(f) + g_2(f) \circ p_2(f) = \mathrm{Id}_V.$$

Für $v \in V_1 \cap V_2$ ist dann

$$v = \mathrm{Id}_V(v) = (g_1(f) \circ p_1(f))(v) + (g_2(f) \circ p_2(f))(v) = \mathbf{0},$$

also ist $V_1 \cap V_2 = \{\mathbf{0}\}$. Ist $v \in V$ beliebig, so ist

$$\begin{aligned} v &= \mathrm{Id}_V(v) \\ &= (g_1(f) \circ p_1(f))(v) + (g_2(f) \circ p_2(f))(v) \\ &= (p_1(f) \circ g_1(f))(v) + (p_2(f) \circ g_2(f))(v) \\ &= v_2 + v_1 \end{aligned}$$

wobei

$$\begin{aligned} v_2 &:= (p_1(f) \circ g_1(f))(v) \in V_2 = \mathrm{Ker}(p_2(f)) \\ v_1 &:= (p_2(f) \circ g_2(f))(v) \in V_1 = \mathrm{Ker}(p_1(f)) \end{aligned}$$

wegen $p_1(f) \circ p_2(f) = p_2(f) \circ p_1(f) = 0$ gilt.
Wir haben also $V_1 + V_2 = V$ und damit insgesamt

$$V = V_1 \oplus V_2.$$

Die Aussage für Polynome $p_1, \ldots, p_r$ folgt hieraus leicht durch vollständige Induktion nach r (Übung). Man benutzt dabei, dass aus der paarweisen Teilerfremdheit der p_i folgt, dass auch p_1 und $p_2 \ldots p_r$ zueinander teilerfremd sind. Dies zeigt man z.B., indem man zunächst $1 = h_{1,2}p_1 + h_2 p_2 = \cdots = h_{1,r}p_1 + h_r p_r$ schreibt und dann das Produkt $1 = \prod_{i=2}^r (h_{1,i}p_1 + h_i p_i)$ distributiv ausmultipliziert; der einzige Term, der nicht durch p_1 teilbar ist, ist dabei ein Vielfaches von $p_2 \ldots p_r$. Alternativ überlegt man sich zunächst, dass zwei Polynome p_1, p_2 genau dann im oben angegebenen Sinne teilerfremd sind, wenn ihre Zerlegungen in Produkte irreduzibler Polynome keinen gemeinsamen irreduziblen Faktor enthalten (Übung). □

Bemerkung. Der Satz verallgemeinert offenbar Satz 14.7 über die Hauptraumzerlegung für den Fall, dass das charakteristische Polynom

nicht in Linearfaktoren zerfällt. Im Fall $q_j = X - \beta_j$ erhält man erneut genau die Aussage dieses Satzes

Definition 14.13. *Sei R ein kommutativer Ring mit 1. Ein R-Modul M heißt ein Torsionsmodul, wenn es zu jedem $x \in M$ ein $a \in R$ gibt mit $a \neq 0, ax = 0$.*

Beispiel. a) Sei $R = \mathbb{Z}, M = \mathbb{Z}/2\mathbb{Z}$. Wegen $2 \cdot \bar{1} = \bar{0}$ ist M ein R-Torsionsmodul.

b) Der endlich dimensionale K-Vektorraum V werde mit einem $f \in \mathrm{End}(V)$ wie üblich zu einem $K[X]$-Modul gemacht. Dann folgt aus $\mu_f(f) = 0$, dass $\mu_f *_f v = 0$ für alle $v \in V$ gilt, V ist also ein Torsionsmodul über $K[X]$ (aber natürlich nicht über dem Körper K!).

Bemerkung. Ist der Modul endlich erzeugt, etwa von $x_1, \ldots, x_n$ und Torsionsmodul und sind $a_i \in R$ mit $a_i x_i = 0$ sowie $a = \prod_i a_i$, so gilt offenbar $ax = 0$ für alle $x \in M$. Für einen endlich erzeugten Torsionsmodul gibt es also ein einheitliches $a \in R$ mit $a \neq 0, ax = 0$ für alle $x \in M$.

Den Beweis des folgenden Satzes werden wir im nächsten Abschnitt behandeln.

Satz 14.14. *Sei $R = K[X]$ oder $R = \mathbb{Z}$, sei V ein endlich erzeugter R-Torsionsmodul.*
Dann gibt es $r \in \mathbb{N}$, Elemente $v_1, \ldots, v_r \in V$ und im Fall $R = K[X]$ eindeutig bestimmte normierte irreduzible Polynome $q_1, \ldots, q_r \in K[X]$ bzw. im Fall $R = \mathbb{Z}$eindeutig bestimmte Primzahlen $q_1, \ldots, q_r \in \mathbb{Z}$ (die jeweils nicht notwendig paarweise verschieden sind), sowie (ebenfalls eindeutig bestimmte) $\mu_j \in \mathbb{N}(1 \leq j \leq r)$, so dass gilt:

a) *Ist $a \in R$, so ist genau dann $av_j = \mathbf{0}$, wenn $q_j^{\mu_j}$ ein Teiler von a in R ist.*

b) *Ist $v \in V$, so kann man*

$$v = \sum_{j=1}^{r} a_j v_j$$

mit Elementen $a_j = a_j^{(v)} \in R$ schreiben, dabei sind für jedes $v \in V$ die Elemente $a_j = a_j^{(v)}$ modulo $q_j^{\mu_j} R$ (d.h., bis auf Addition von Vielfachen von $q_j^{\mu_j}$) eindeutig bestimmt. Insbesondere sind die a_j eindeutig bestimmt, wenn man im Fall $R = \mathbb{Z}$ zusätzlich $0 \leq a_j < q_j^{\mu_j}$ verlangt und im Fall $R = K[X]$ zusätzlich verlangt, dass $a_j = 0$ oder $\deg(a_j) < \deg(q_j^{\mu_j} = \mu_j \deg(q_j)$ für alle j gilt.

Insbesondere gilt mit $V_j := Rv_j := \{av_j \mid a \in R\}$:

$$V = V_1 \oplus \cdots \oplus V_r,$$

und für $1 \leq j \leq r$ *ist*

$$V_j \cong R/q_j^{\mu_j} R := \{a + q_j^{\mu_j} R \mid a \in R\},$$

wobei die letzte Isomorphie als Isomorphie von R*-Moduln zu verstehen ist und der Faktormodul* $R/q_j^{\mu_j} R$ *analog zum Faktorraum (siehe Satz 10.3) definiert ist.*
Speziell haben wir:
Sei V *ein endlichdimensionaler* K*-Vektorraum,* $f \in \mathrm{End}(V)$*.*
Dann gibt es $r \in \mathbb{N}, r \leq n = \dim(V)$*, Vektoren* $v_1, \ldots, v_r \in V$ *und eindeutig bestimmte normierte irreduzible Polynome* $q_1, \ldots, q_r \in K[X]$ *(die nicht notwendig paarweise verschieden sind) sowie (ebenfalls eindeutig bestimmte)* $\mu_j \in \mathbb{N}(1 \leq j \leq r)$*, so dass gilt:*

a) *Ist* $p \in K[X]$*, so ist genau dann* $p(f)(v_j) = \mathbf{0}$*, wenn* $q_j^{\mu_j}$ *ein Teiler von* p *in* $K[X]$ *ist.*
b) *Ist* $v \in V$*, so kann man*

$$v = \sum_{j=1}^{r} p_j(f)(v_j)$$

mit Polynomen $p_j = p_j^{(v)} \in K[X]$ *schreiben, dabei sind für jedes* $v \in V$ *die Polynome* $p_j = p_j^{(v)}$ *modulo* $q_j^{\mu_j} K[X]$ *eindeutig bestimmt.*

Insbesondere gilt mit $V_j := K[X]v_j := \{p(f)(v_j) \mid p \in K[X]\}$*:*

$$V = V_1 \oplus \cdots \oplus V_r,$$

und für $1 \leq j \leq r$ *ist*

$$V_j \cong K[X]/q_j^{\mu_j} K[X],$$

wobei der letzte Isomorphismus ein Isomorphismus von $K[X]$*-Moduln ist.*

Korollar 14.15. *Mit den Bezeichnungen des Satzes gilt:*

a) *Ist* $\deg(q_j^{\mu_j}) = t_j$ *und* $\lambda \in K$*, so bilden die Vektoren* $v_j, (f - \lambda \,\mathrm{Id}_V)v_j, \ldots, (f - \lambda \,\mathrm{Id}_V)^{t_j - 1} v_j$ *eine Basis des* K*-Vektorraums* $K[X]v_j$*.*
b) *Es gilt*

$$\chi_f = \prod_{j=1}^{r} q_j^{\mu_j}.$$

c) *Sind die* q_j *so nummeriert, dass* $\{q_1, \ldots, q_r\} = \{q_1, \ldots, q_t\}$ *mit einem* $t \leq r$ *und paarweise verschiedenen* $q_1, \ldots, q_t$ *sowie* $\mu_i = \max\{\mu_j \mid 1 \leq j \leq r, q_j = q_i\}$ *für* $1 \leq i \leq t$ *gilt, so gilt für das Minimalpolynom* μ_f *von* f

$$\mu_f = \prod_{j=1}^{t} q_j^{\mu_j}.$$

Insbesondere hat das Minimalpolynom die gleichen irreduziblen Faktoren wie das charakteristische Polynom (siehe Korollar 13.7 für den Fall, dass das charakteristische Polynom in ein Produkt von Linearfaktoren zerfällt).

Beweis des Korollars. a) ist klar für $\lambda = 0$. Für beliebiges λ expandiert man $(f - \lambda \operatorname{Id}_V)^k$ für $1 \leq k \leq t_j$ nach dem binomischen Lehrsatz und sieht, dass die Übergangsmatrix zwischen den Vektoren $v_j, \ldots, f^{t_j-1}(v_j)$ und den Vektoren $v_j, (f - \lambda \operatorname{Id}_V)v_j, \ldots, (f - \lambda \operatorname{Id}_V)^{t_j-1}v_j$ eine Dreicksmatrix mit Determinante 1 ist. Die letzteren Vektoren bilden daher ebenfalls eine Basis des Raums $K[X]v_j$.
b): Offenbar reicht es, die Behauptung für die Räume $V_j = K[X]v_j = \{p(f)(v_j) \mid p \in K[X]\}$ zu zeigen (einen solchen Unterraum nennt man einen *f-zyklischen Unterraum*). Ist (mit $q := q_j, \mu := \mu_j, w := v_j$) $q^\mu(X) = \sum_{i=1}^t a_i X^i$ mit $a_t = 1$, so bilden die Vektoren $v, f(w), \ldots, f^{t-1}(w)$ eine Basis von $V_j =: W$, bezüglich der $f|_W$ die Matrix

$$\begin{pmatrix} 0 & & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & & \ddots & 0 & \vdots \\ 0 & & \ldots & 1 & -a_{t-1} \end{pmatrix} \in M_t(K)$$

hat.
Man rechne als Übung nach, dass $q^\mu = \sum_{i=1}^t a_i X^i$ das charakteristische Polynom dieser Matrix ist (sie wird auch die *Begleitmatrix* von f genannt).
Die Aussage c) über das Minimalpolynom ist trivial. □

Wir kommen jetzt zurück zum zweiten Beweis von Satz 14.9. Da g nilpotent vom Index d ist, ist das Minimalpolynom μ_g von g gleich X^d, nach Teil c) des vorigen Korollars folgt, dass $q_j = X$ für alle j gilt.
Da $g^d = 0$ ist, sind die Exponenten μ_j alle zwischen 1 und d, und für $\mu_j = \nu$ hat $g|_{K[X]v_j}$ bezüglich der Basis dieses Teilraums $V_j = K[X]v_j$ aus den Vektoren $v_j, g(v_j), \ldots, g^{\nu-1}v_j$ die Matrix ${}^tJ_\nu$ (bzw. bezüglich der Basis $g^{\nu-1}v_j, \ldots, g(v_j), v_j$ die Matrix J_ν).
Bezeichnen wir mit s_ν die Anzahl der j mit $\mu_j = \nu$, so erhalten wir bezüglich der aus diesen Basen der V_j zusammengesetzten Basis von V wieder die Matrix von g in der behaupteten Normalgestalt.

Die Eindeutigkeit folgt in diesem Fall daraus, dass man aus der Basis von V, bezüglich der die Matrix von g die Normalform annimmt, wieder eine Zerlegung von V gemäß Satz 14.14 gewinnt, indem man als Vektoren v_j aus dieser Basis zu jedem Kästchen J_ν den letzten Basisvektor aus dem zugehörigen Abschnitt der Basis wählt. Die Eindeutigkeit der Zerlegung von V nach Satz 14.14 impliziert dann die Eindeutigkeit der Normalform.

Satz 14.16. *Sei V ein endlich dimensionaler K-Vektorraum, $f \in \mathrm{End}(V)$. Seien q_j, μ_j wie in Satz 14.14. Dann ist $\chi_f = \prod_j q_j^{\mu_j}$, und V hat eine Basis, bezüglich der die Matrix von f Blockgestalt*

$$\begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_r \end{pmatrix}$$

hat, wobei B_j die Begleitmatrix von $q_j^{\mu_j}$ ist. Die A_j sind dabei bis auf die Reihenfolge eindeutig bestimmt.
Diese Matrix in Blockgestalt heißt die rationale Normalform *von f.*
Ist $A \in M_n(K)$, so ist A zu (bis auf Vertauschung der Blöcke) genau einer Matrix in obiger Blockgestalt ähnlich, diese heißt die rationale Normalform *von A.*

Beweis. Klar nach Satz 14.14 und Korollar 14.15. Die gewünschte Basis setzt man zusammen aus den Basen $v_j, f(v_j), \ldots, f^{t_j-1}(v_j)$ der Räume $V_j = K[X]v_j$. □

Bemerkung. a) Im Falle, dass die q_j alle lineare Polynome $q_j = X - \lambda_j$ sind, erhält man aus der obigen Darstellung die Jordan'sche Normalform, indem man für V_j statt der Basis

$$(v_j, f(v_j), \ldots, f^{t_j-1}(v_j))$$

die Basis

$$((f - \lambda \mathrm{Id}_V)^{t_j-1}(v_j), \ldots, (f - \lambda \mathrm{Id}_V)(v_j), v_j)$$

wählt.

b) Man kann zeigen: Sind $A, A' \in M_n(K)$, K ein beliebiger Körper, und $L \supseteq K$ ein Erweiterungskörper, in dem χ_A und $\chi_{A'}$ in Linearfaktoren zerfallen, so sind A und A' genau dann in $M_n(K)$ zueinander konjugiert, wenn sie über L die gleiche Jordan'sche Normalform haben.

c) Die Bestimmung der Jordan'schen Normalform ist zur algorithmischen Klärung der Frage, ob zwei gegebene Matrizen zueinander konjugiert sind, nur in Grenzen geeignet, da dafür die Nullstellen des charakteristischen Polynoms bestimmt werden müssen, was algorithmisch schwierig ist.

15. Elementarteilersatz und Moduln über Polynomringen

In diesem Abschnitt ist, sofern nicht ausdrücklich etwas anderes vorausgesetzt wird, stets $R = \mathbb{Z}$ oder $R = K[X]$ mit einem Körper K.
Die Hauptaussagen dieses Paragraphen gelten allgemeiner auch für einen Hauptidealring R, einige Beweise vereinfachen sich aber in der angegebenen Situation deutlich.
Wir nennen $a \in R$ normiert, wenn $a > 0$ für $R = \mathbb{Z}$ gilt bzw. wenn a ein normiertes Polynom (im Fall $R = K[X]$) ist.
Wir setzen für $a \in R$:

$$N(a) = \begin{cases} |a| & R = \mathbb{Z} \\ 2^{\deg(a)} & R = K[X], a \neq 0 \\ 0 & a = 0 \end{cases} \tag{15.1}$$

und nennen $N(a)$ die Norm von a.
Diese Funktion $N : R \longrightarrow \mathbb{N}_0$ ist multiplikativ, erfüllt also

$$N(ab) = N(a)N(b) \quad \text{für alle } a, b \in R.$$

Ferner ist $N(a) = 1$ genau dann, wenn a in R ein multiplikatives Inverses hat (Einheit im Ring R ist), wenn also

$$a = \begin{cases} \pm 1 & \text{falls } R = \mathbb{Z} \\ c \in K, c \neq 0 & \text{falls } R = K[X] \end{cases}$$

gilt.

Lemma 15.1. *Seien $a, b \in R$ mit $a \mid b$ (a ist ein Teiler von b, es gibt ein $c \in R$ mit $b = ac$). Dann ist $N(a) \leq N(b)$, mit $N(a) = N(b)$ genau dann, wenn $a = \epsilon b$ gilt mit einem in R invertierbaren $\epsilon \in R$.*

Beweis. Mit $b = ac$ wie oben ist $N(b) = N(a)N(c) \geq N(a)$, da $N(c) \geq 1$ für $c \neq 0$ gilt. Da $N(c) = 1$ genau dann gilt, wenn c invertierbar ist, folgt auch der Zusatz über die Gleichheit. □

Wir wissen weiter, dass folgendes gilt: In jedem der beiden Fälle hat man eine Division mit Rest in R (auch euklidischer Algorithmus genannt):
Sind $a, b \in R, b \neq 0$, so gibt es $q, r \in R$, mit $N(r) < N(b)$, so dass $a = qb + r$ gilt.
Daraus folgt (siehe Beweis von Satz 13.3 für $R = K[X]$, der Beweis für $R = \mathbb{Z}$ geht genauso mit $N(a)$ statt $\deg(a)$), dass jedes Ideal in R ein Hauptideal ist, also von der Form $I = (g) = \{cg \mid c \in R\}$ für ein $g \in R$. Verlangt man, dass das Element g normiert ist, so ist es dadurch eindeutig bestimmt.

Definition und Lemma 15.2. *Seien $a_1, \ldots, a_n \in R$ gegeben, nicht alle Null. Dann gibt es genau ein normiertes $d \in R$, so dass*

$$(a_1, \ldots, a_n) := \{x_1 a_1 + \cdots + x_n a_n \mid x_i \in R\} = (d) := \{xd \mid x \in R\}$$

gilt. Dieses d heißt der größte gemeinsame Teiler von $a_1, \dots, a_n$, man schreibt $d = \mathrm{ggT}(a_1, \dots, a_n)$.
Es gibt $x_1, \dots, x_n \in R$ mit $d = x_1 a_1 + \cdots + x_n a_n$.
Es gilt: Ist d' irgendein gemeinsamer Teiler aller a_i, so ist $N(d') \leq N(d)$. Ist hier d' ebenfalls normiert, so ist $d = d'$ oder $N(d') < N(d)$. Der ggT *der a_i ist also in Bezug auf die Norm der größte normierte aller gemeinsamen Teiler.*

Beweis. Dass $(a_1, \dots, a_n)$ in der Tat ein Ideal ist, ist ist klar, also gibt es nach Satz 13.3 (bzw. dessen Analogon für $\mathbb{Z}$) ein (eindeutig bestimmtes) normiertes $d \in R$, so dass $(a_1, \dots, a_n) = (d)$ gilt; für dieses d hat man wegen $d \in (a_1, \dots, a_n)$ eine Darstellung $d = x_1 a_1 + \cdots + x_n a_n$ mit $x_i \in R$.
Ein gemeinsamer Teiler d' der a_i teilt dann auch $d = x_1 a_1 + \cdots + x_n a_n$, also gilt $N(d') \leq N(d)$. Normiertheit von d, d' mit $d' \neq d$ schliesst hier aus, dass $d = \epsilon d'$ mit invertierbarem ϵ gilt, also ist dann sogar $N(d') < N(d)$.

□

Lemma 15.3. *Eine Matrix $A \in M_n(R)$ ist genau dann in $M_n(R)$ invertierbar, wenn* $\det(A)$ *eine Einheit in R ist.*
Die Menge der invertierbaren Matrizen in $M_n(R)$ wird mit $GL_n(R)$ bezeichnet.

Beweis. Satz 7.20 gilt auch, wenn man den dort betrachteten Körper K durch einen kommutativen Ring R mit 1 ersetzt und dort in der Aussage über Invertierbarkeit die Bedingung $\det(A) \neq 0$ durch die Bedingung $\det(A)$ ist invertierbar in R ersetzt (siehe die zweite Bemerkung nach Satz 7.21). □

Satz 15.4 (Elementarteilersatz, Smith-Normalform)**.** *Sei $A \in M(p \times n, R)$, $A \neq 0$. Dann gibt es Matrizen $S \in GL_p(R), T \in GL_n(R)$, so dass*

$$SAT = \begin{pmatrix} d_1 & \dots & 0 & & \\ & \ddots & & & 0 \\ 0 & \dots & d_r & & \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} \tag{15.2}$$

mit normierten $d_j \neq 0$ und $d_j \mid d_{j+1}$ für $1 \leq j \leq r-1$ gilt.
Die Diagonalelemente $d_1, \dots, d_r$ sind eindeutig bestimmt, sie heißen Elementarteiler *oder* Invariantenteiler *der Matrix A, die Matrix (15.2) heißt* Elementarteilerform (Smith-Normalform) *von A.*
Der erste Elementarteiler d_1 ist dabei der größte gemeinsame Teiler der Einträge a_{ij} der Matrix A.

Beweis. Bevor wir den eigentlichen Beweis beginnen, erinnern wir daran, dass die elementaren Zeilenumformungen einer Matrix $A \in M(p \times n, R)$ der drei Typen

i) Addition der mit $\lambda \in R$ multiplizierten j-ten Zeile zur i-ten Zeile (also ${}^t\mathbf{z}_i \longmapsto {}^t\mathbf{z}'_i = {}^t\mathbf{z}_i + \lambda {}^t\mathbf{z}_j$) für $i \neq j$.
ii) Multiplikation der i-ten Zeile mit einer Einheit $\lambda \in R^\times$.
iii) Vertauschen von i-ter Zeile und j-ter Zeile.

durch Multiplikation von links mit einer Matrix aus $GL_p(R)$ realisiert werden können (nämlich mit einer Elementarmatrix $T_{ij}(\lambda)$, einer Diagonalmatrix $D_i(\lambda)$ bzw. einer Permutationsmatrix P_{ij}). Genauso werden die elementaren Spaltenumformungen durch Multiplikation von rechts mit der entsprechenden Matrix aus $GL_n(R)$ realisiert.
Wir können also die Behauptung beweisen, indem wir zeigen, dass A sich durch elementare Zeilen- und Spaltenumformungen der angegebenen Typen in die angegebene Gestalt bringen lässt.

Das zeigen wir jetzt durch Induktion nach der Anzahl p der Zeilen von A (wie schon beim Gauß - Algorithmus über einem Körper K kann man den Beweis auch als Angabe eines rekursiven Algorithmus auffassen). Für $p = 1$ nehmen wir an, dass A nicht die Nullzeile ist (sonst ist nichts zu zeigen) und erreichen durch Spaltenvertauschungen, dass $a_{11} \neq 0$ die kleinste Norm unter allen $a_{1j} \neq 0$ hat. Anschließend teilen wir alle a_{1j} mit Rest durch a_{11}, schreiben also $a_{1j} = \lambda_j a_{11} + a'_{1j}$ mit $N(a'_{1j}) < N(a_{11})$ (und ziehen die mit λ_j multiplizierte 1-te Spalte von A von der j-ten ab.
Wir erhalten eine Zeile, in der entweder alle Einträge außer a_{11} gleich 0 sind oder die minimale Norm eines von 0 verschiedenen Eintrags kleiner als $N(a_{11})$ ist, im letzteren Fall platzieren wir ein Element minimaler Norm durch Spaltenvertauschungen in Position $1, 1$ und beginnen von vorn. Da die Norm eines Elements in $\mathbb{N}_0$ liegt, kann diese minimale Norm nur endlich oft verkleinert werden, nach endlich vielen Schritten erhalten wir also eine Zeile der Form $(d_1, 0, \ldots, 0)$.
In dieser ist offenbar d_1 der größte gemeinsame Teiler aller Einträge. Da eine Umformung $a_{1j} \mapsto a'_{1j} = a_{1j} - \lambda_j a_{11}$ den größten gemeinsamen Teiler aller Einträge nicht ändert, ist $d_1 = \mathrm{ggT}(a_{11}, \ldots, a_{1n})$.

Sei jetzt $p > 1$ und die Behauptung für Matrizen mit weniger als p Zeilen gezeigt.
Wir bringen zunächst durch Zeilen - und Spaltenvertauschungen einen Eintrag minimaler Norm in die Position $1, 1$ und erreichen dann in der gleichen Weise wie eben durch Zeilen- und Spaltenumformungen, dass in der ersten Zeile und der ersten Spalte alle Elemente außer $a_{11} =: d_1$ gleich 0 sind; die minimale Norm eines Eintrags der Matrix hat sich dabei vermindert oder ist gleich geblieben, und $N(d_1)$ ist nicht größer als die anfängliche minimale Norm eines Eintrags der Matrix.

Falls jetzt alle Einträge der Matrix durch d_1 teilbar sind, führt man die Matrix $A' \in M((p-1) \times (n-1), R)$, die man durch Streichen der ersten Zeile und Spalte erhält, mit Hilfe der Induktionsannahme in die Form

$$\begin{pmatrix} d_2 & \dots & 0 & & \\ & \ddots & & & 0 \\ 0 & \dots & d_r & & \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$$

mit $d_j \neq 0$ und $d_j \mid d_{j+1}$ für $2 \leq j \leq r-1$ über, dabei ist d_2 als größter gemeinsamer Teiler der Einträge von A' durch d_1 teilbar.

Andernfalls sei etwa a_{ij} nicht durch $a_{11} = d_1$ teilbar. Man addiert dann die erste Zeile zur i-ten und dividiert a_{ij} mit Rest durch d_1. Mit $a_{ij} = \lambda_j d_1 + a'_{ij}$ subtrahiert man die mit λ_j multiplizierte (neue) 1-te Spalte von der j-ten und hat einen Eintrag a'_{ij} erzeugt, dessen Norm kleiner als $N(d_1)$ und damit kleiner als die anfängliche minimale Norm eines Eintrags der Matrix ist. Man beginnt dann das Verfahren von Neuem. Da die Norm Werte in $\mathbb{N}_0$ nimmt, kann die minimale Norm nur endlich oft vermindert werden, nach endlich vielen Schritten muss also der Fall erreicht werden, in dem alle Einträge durch den Eintrag d_1 in Position $1,1$ teilbar sind und man die Induktionsannahme anwenden kann. $\square$

Die Eindeutigkeit werden wir im nächsten Satz beweisen. **Bemerkung:**

a) Der Satz ist für einen beliebigen Hauptidealring R richtig, allerdings ohne das Konzept der Norm und die Division mit Rest etwas schwieriger zu beweisen.
b) Lässt man nur Multiplikation von links bzw. von rechts mit einer invertierbaren Matrix zu, so erreicht man untere bzw. obere Dreiecksgestalt (Hermite-Normalform)
c) Für Matrizen in $M(p \times n, R)$ kann man Äquivalenz (über R) genauso wie in Definition 6.7 für $M(p \times n, K)$ definieren; der Elementarteilersatz sagt dann aus, dass jede Matrix aus $M(p \times n, R)$ zu (im wesentlichen genau) einer Matrix in Elementarteilergestalt äquivalent ist.

Satz 15.5. *Sei $A \in M(p \times n, R), T \in M(p \times p, R)$. Dann gilt für $1 \leq r \leq p$:*
Die $r \times r$ Unterdeterminanten ($r \times r$ Minoren) von TA sind Linearkombinationen (mit Koeffizienten in R) der $r \times r$ Unterdeterminanten von A.

Das Gleiche gilt für AS mit $S \in M(n \times n, R)$.
Insbesondere gilt für $S \in GL_p(R), T \in GL_n(R)$:

a) *Der größte gemeinsame Teiler der $r \times r$ Unterdeterminanten von A ist (bis auf Multiplikation mit Einheiten) gleich dem größten gemeinsamen Teiler der $r \times r$ Unterdeterminanten von SAT.*
b) *Ist*

$$SAT = \begin{pmatrix} d_1 & \dots & 0 & & \\ & \ddots & & & 0 \\ 0 & \dots & d_r & & \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} \tag{15.3}$$

in Elementarteilergestalt, so ist für $1 \leq j \leq r$ der größte gemeinsame Teiler der $j \times j$ Unterdeterminanten von A gleich $d_1 \dots d_j$; er heißt der j-te Determinantenteiler *von A.*
c) *Die Elementarteiler $d_1, \dots, d_j$ der Matrix A sind eindeutig bestimmt.*

Beweis. Für einen Beweis der ersten Aussage dieses Satzes sei für den Augenblick auf das Buch von Lorenz verwiesen, wir kommen bei der Behandlung der multilinearen Algebra noch einmal darauf zurück. Die Aussagen a)-c) folgen daraus direkt. □

Beispiel:
Sei $R = \mathbf{Q}[X]$,

$$A = \begin{pmatrix} X^3 + X^2 - 2X - 2 & X^5 - 4X \\ X^5 - 4X & X^5 - X^4 - 4X + 4 \end{pmatrix}.$$

Wir bringen A durch elementare Umformungen über $R = \mathbf{Q}[X]$ in Elementarteilergestalt:

$$\begin{pmatrix} X^3+X^2-2X-2 & X^5-4X \\ X^5-4X & X^5-X^4-4X+4 \end{pmatrix} \xrightarrow{Z_{II}\mapsto Z_{II}-(X^2-X+3)Z_I}$$
$$\begin{pmatrix} X^3+X^2-2X-2 & X^5-4X \\ -3X^2+6 & -X^7+X^6-2X^5-X^4+4X^3-4X^2+8X+4 \end{pmatrix}$$

$$\xrightarrow{Z_I\leftrightarrow Z_{II},Z_{II}\mapsto 3Z_{II}}$$

$$\begin{pmatrix} -3X^2+6 & -X^7+X^6-2X^5-X^4+4X^3-4X^2+8X+4 \\ 3(X^3+X^2-2X-2) & 3(X^5-4X) \end{pmatrix}$$

$$\xrightarrow{Z_{II}\mapsto Z_{II}+(X+1)Z_I,S_I\mapsto -S_I/3}$$

$$\begin{pmatrix} X^2-2 & -X^7+X^6-2X^5-X^4+4X^3-4X^2+8X+4 \\ 0 & -X^8-X^6+3X^4+4X^2+4 \end{pmatrix}$$

$$\xrightarrow{S_{II}\mapsto S_{II}-(-X^5+X^4-4X^3+X^2-4X-2)S_I}$$

$$\begin{pmatrix} X^2-2 & 0 \\ 0 & -(X^2-2)(X^2+2)(X^2-X+1)(X^2+X+1) \end{pmatrix}$$

Dabei kommt z. B. der erste Umformungsschritt dadurch zustande, dass man X^3+X^2-2X-2 als den Eintrag mit der kleinsten Norm identifiziert und X^5-4X mit Rest durch X^3+X^2-2X-2 teilt, dabei erhält man $X^5-4X=(X^2-X+3)(X^3+X^2-2X-2)+(-3X^2+6)$, analog für die weiteren Schritte.

Der Elementarteilersatz wird häufig auch in einer Form gebraucht, in der er Aussagen über endlich erzeugte R- Moduln und deren Untermoduln macht. Dafür fassen wir zunächst zusammen, wie sich die Begriffe der Vektorraumtheorie auf Moduln über einem Ring übertragen:

Bemerkung. Die Begriffe *lineare Abbildung, Kern, linear abhängig/ unabhängig, Erzeugendensystem, Basis* sind für R-Moduln genauso definiert wie für Vektorräume über einem Körper. Auch für den R-Modul ist ein linear unabhängiges Erzeugendensystem eine Basis (und umgekehrt), im Gegensatz zur Vektorraumsituation muss aber weder ein minimales Erzeugendensystem noch ein maximales linear unabhängiges System eine Basis sein, und es gibt (endlich erzeugte) Moduln über Ringen, die überhaupt keine Basis haben. Das einfachste Beispiel hierfür ist der $\mathbb{Z}$-Modul $\mathbb{Z}/2\mathbb{Z}$, in dem es wegen $2\cdot\bar{1}=\bar{0}$ überhaupt keine linear unabhängigen Vektoren gibt.
Ein R-Modul, der eine Basis hat, heißt *frei*, das einfachste Beispiel hierfür ist $M=R^n$ für $n\in\mathbb{N}$, in diesem Fall hat man wieder die Standardbasis aus den $\mathbf{e}_i$, in denen die i-te Komponente 1 ist und alle anderen Komponenten 0. Hat der R-Modul M eine Basis aus n Elementen,

so ist er isomorph zu R^n (man bilde die Elemente der Standardbasis von R^n auf die Basisvektoren von M ab und setze linear fort).
Der Faktormodul M/N ist für einen Untermodul N des R-Moduls M genauso als $M/N := \{x + N \mid x \in M\}$ mit den Verknüpfungen $(x + N) + (y + N) = (x + y) + N, \lambda(x + N) = \lambda x + N$ definiert wie der Faktorraum in Satz 10.3, auch der Beweis für die Wohldefiniertheit dieser Verknüpfungen überträgt sich ohne jede Änderung.
Auch der Homomorphiesatz (Satz 10.6) überträgt sich ohne Änderung, insbesondere wird für eine lineare Abbildung (auch Modulhomomorphismus genannt) $f : M \to N$ von R-Moduln durch $\bar{f}(x + \mathrm{Ker}(f)) := f(x)$ ein Isomorphismus $\bar{f} : M/\mathrm{Ker}(f) \to \mathrm{Im}(f)$ gegeben.

Satz 15.6. a) *Sei $M \subseteq R^p$ ein (endlich erzeugter) R-Untermodul. Dann gibt es Elemente $x_1, \ldots, x_p \in R^p, r \in \mathbb{N}$, $d_1, \ldots, d_r \in R$ mit $d_j \neq 0$ für $1 \leq j \leq r$ und $d_j \mid d_{j+1}$, so dass gilt:*
i) *$(x_1, \ldots, x_p)$ ist Basis von R^p.*
ii) *$(d_1x_1, \ldots, d_rx_r)$ ist Basis von M.*
Insbesondere ist M ein freier Modul.

Die Elemente $d_1, \ldots, d_r$ heißen die Elementarteiler *(oder Invariantenteile) von M in R^n; nimmt man sie als normiert an, so sind sie eindeutig bestimmt. Die $\delta_i := \prod_{j=1}^{i} d_i$ heißen die* Determinantenteiler.

Basen $(x_1, \ldots, x_p)$ von R^p, $(d_1x_1, \ldots, d_rx_r)$ von M wie oben nennt man Elementarteilerbasen *oder angepasste Basen von $M \subseteq R^n$.*

b) *Ist M ein endlich erzeugter R-Modul, so gibt es $x_1, \ldots, x_n \in M, c_1, \ldots, c_n \in R$, die nicht Einheiten in R sind, mit $c_1, \ldots, c_r \neq 0, c_{r+1} = \cdots = c_n = 0$ (für ein $r \leq n$) und $c_i \mid c_{i+1}$ für $i < r$, so dass jedes $v \in M$ sich als*

$$v = \sum_{i=1}^{n} a_i x_i$$

mit modulo c_i (d.h. bis auf Addition von Vielfachen von c_i) eindeutig bestimmten a_i schreiben lässt.

Beweis. a) Im folgenden Lemma werden wir sehen, dass ein beliebiger Untermodul von R^p zwangsläufig endlich erzeugt ist (diese Aussage gilt nicht über einem beliebigen kommutativen Ring R, die Ringe, für die sie gilt, heißen *noethersch*).
Sei also $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$ ein Erzeugendensystem von M und $A \in M(p \times n, R)$ die Matrix mit Spalten $\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(n)}$.

Nach dem Elementarteilersatz für Matrizen (Satz 15.4) findet man $S \in GL_p(R), T \in GL_n(R)$, so dass SAT die Elementarteilergestalt

$$\begin{pmatrix} d_1 & \dots & 0 & & \\ & \ddots & & & 0 \\ 0 & \dots & d_r & & \\ 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$$

hat. Wir setzen $\tilde{S} := S^{-1}$ und bezeichnen die Spalten von $\tilde{S}$ mit $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(p)}$; diese Vektoren bilden wegen $S \in GL_p(R)$ eine Basis von R^p.

Ebenso erzeugen die Vektoren $\mathbf{u}^{(k)} := \sum_{l=1}^{n} t_{lk}\mathbf{w}^{(l)}$ für $1 \leq k \leq n$ wegen $T \in GL_n(R)$ den gleichen Untermodul von R^p wie die Vektoren $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(n)}$, nämlich M. Da die Koeffizienten b_{ik} von $B = SAT$ die Vektoren $\mathbf{u}^{(k)}$ als

$$\mathbf{u}^{(k)} = \sum_{i=1}^{p} b_{ik}\mathbf{v}^{(i)}$$

durch die $\mathbf{v}^{(i)}$ ausdrücken, haben wir schließlich

$$\mathbf{u}^{(k)} = \begin{cases} d_k\mathbf{v}^{(k)} & k \leq r \\ \mathbf{0} & k > r \end{cases}$$

wie behauptet.

Für b) sei $\{\mathbf{0}\} \neq M$ erzeugt von $y_1, \dots, y_m$ und $f : R^m \longrightarrow M$ die durch

$$f\left(\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}\right) := \sum_{i=1}^{m} a_i y_i$$

gegebene lineare Abbildung; diese ist surjektiv, da die y_i den Modul M erzeugen.

Wir finden dann nach a) eine Basis $(x'_1, \dots, x'_m)$ von R^m und $r \in \mathbb{N}, r \leq m$ sowie $d_1, \dots d_r \in R$ mit $d_i \mid d_{i+1}$ für $1 \leq< i \leq r$, so dass $(d_1x'_1, \dots, d_rx'_r)$ eine Basis des Untermoduls $N := \mathrm{Ker}(f) \subseteq R^m$ ist. Für $r < i \leq m$ setzen wir $d_i = 0$.

Ist $0 \leq s < r$ so, dass $d_1, \dots, d_s$ invertierbar in R sind (Einheiten in R) und $d_{s+1}, \dots d_r$ nicht invertierbar in R sind, so ist $f(x'_1) = \dots = f(x'_s) = \mathbf{0}$ (da die $x'_i = (d_i)^{-1} \cdot d_ix_i$ in N sind). Da f surjektiv ist, erzeugen also bereits die $x_j = f(x'_{s+j})$ für $1 \leq j \leq r-s =: n$ den Modul M, also lässt sich jedes $x \in M$ schreiben als $x = \sum_{j=1}^{n} a_jx_j$. Ist $x = \sum_{j=1}^{n} a'_jx_j$ eine weitere solche Darstellung, so haben wir $f(\sum_{j=1}^{n}(a_j - a'_j)x'_{s+j}) = \mathbf{0}$, also $\sum_{j=1}^{n}(a_j - a'_j)x'_{s+j} \in N$. Mit $c_j := d_{s+j}$ für $1 \leq j \leq r - s = n$ ist dann $a_j - a'_j$ für $1 \leq j \leq n$ durch c_j teilbar, in der

Darstellung $x = \sum_{j=1}^{n} a_j x_j$ sind also wie behauptet die a_j eindeutig modulo den c_j (d.h., bis auf Addition von Vielfachen von c_j), was die Behauptung beweist.
Mittels des Homomorphiesatzes für Moduln können wir den letzten Teil des Beweises eleganter auch so formulieren:
Der Homomorphiesatz für Moduln liefert einen Isomorphismus $R^n/\mathrm{Ker}(f) \cong M$, da f surjektiv ist. Dann hat man (mit c_j wie oben)

$$\begin{aligned} M &\cong R^m/\mathrm{Ker}(f) \\ &\cong Rx'_1/Rd_1x'_1 \oplus \cdots \oplus Rx'_m/Rd_mx'_m \\ &\cong R/c_1R \oplus \cdots \oplus R/c_nR. \end{aligned}$$

□

Bemerkung. Teil a) des Satzes kann man als die für Moduln über R gültige Version des Basisergänzungssatzes aus der Theorie von Vektorräumen über Körpern ansehen. Zwar kann man eine beliebige Basis des Untermoduls $M \subseteq R^p$ nicht mehr unbedingt zu einer Basis von R^p ergänzen, aber man kann immerhin eine Basis von M finden, die aus Vielfachen eines Teils der Vektoren einer geeigneten Basis von R^p besteht.
Teil b) gibt die für einen beliebigen endlich erzeugten R-Modul gültige Version des Satzes von der Existenz von Basen in K- Vektorräumen: Die Koeffizienten in der Schreibweise eines beliebigen Vektors aus M als Linearkombination der Erzeugenden $x_1, \ldots, x_n$ sind zwar nicht mehr wie bei einer Basis eindeutig bestimmt, aber immerhin eindeutig modulo den c_i. Mehr lässt sich hier, wie das Beispiel des Z-Moduls $\mathbb{Z}/2\mathbb{Z}$ zeigt, nicht erreichen.

Bemerkung. Mit Hilfe eines Satzes der Algebra (chinesischer Restsatz, siehe Übungsblatt 9, Aufgaben 2 und 3 für die gegenwärtige Situtation $R = \mathbb{Z}$ oder $R = K[X]$) kann man die Behauptung b) auch in etwas modifizierter Gestalt beweisen:

b') *Ist M ein endlich erzeugter R-Modul, so gibt es $y_1, \ldots, y_m \in M, c_1, \ldots, c_m \in R$ mit Potenzen von Primelementen $c_1, \ldots, c_r$ und $c_{r+1} = \cdots = c_n = 0$ (für ein $r \leq n$), so dass jedes $v \in M$ sich als*

$$v = \sum_{i=1}^{m} a_i y_i$$

mit modulo c_i eindeutig bestimmten a_i schreiben lässt.

Satz 15.7. *Jeder Untermodul von R^n ($n \in \mathbb{N}$) ist endlich erzeugt.*

Beweis. Wir schreiben für $1 \leq r \leq n$

$$F_r := \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in R^n \right\},$$

also $R^n := F_n$, und setzen $M_r := F_r \cap M$, ferner betrachten wir für $1 \leq j \leq n$ die j-te Koordinatenabbildung $\pi_j : R^n \longrightarrow R$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_j.$$

Wir zeigen durch Induktion nach r, dass M_r ein Erzeugendensystem mit $m(r) \leq r$ Elementen hat, insbesondere also endlich erzeugt ist (schaut man im Beweis genauer hin, so sieht man, dass dieses Erzeugendensystem sogar eine Basis ist).
Für alle j und r ist $\pi_j(M_r)$ offenbar ein Ideal in R, also (da in R jedes Ideal ein Hauptideal ist) ein Hauptideal.
Induktionsanfang: Ist $\pi_1(M_1)$ erzeugt von a_1, so ist also

$$M_1 = \left\{ \begin{pmatrix} xa_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in R^n \mid x \in R \right\},$$

d.h., der Vektor $\begin{pmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ist eine Basis (und damit ein Erzeugendensystem) von M_1.
Ist jetzt $r > 1$ und die Behauptung für M_s mit $s < r$ gezeigt, so betrachten wir das Hauptideal $\pi_r(M_r) = (a_r)$ mit einem $a_r \in R$, und es gibt einen Vektor $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in M_r$.

Ist dann $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in M_r$, so ist $x_r = ca_r$ mit $c \in R$, also ist $\mathbf{x} - c\mathbf{a} \in M_{r-1}$. In M_{r-1} gibt es nach Induktionsannahme ein Erzeugendensystem $\{\mathbf{y}_1, \ldots, \mathbf{y}_s\} \subseteq M_{r-1}$ mit $s \leq r-1$, und man sieht, dass $\{\mathbf{y}_1, \ldots, \mathbf{y}_s, \mathbf{y}_{s+1} := \mathbf{a}\}$ ein Erzeugendensystem von M_r mit $s+1 \leq r$

Elementen ist (in der Tat sogar eine Basis, wenn $\{\mathbf{y}_1, \dots \mathbf{y}_s\} \subseteq M_{r-1}$ eine Basis war). $\square$

Beweis von Satz 14.14. (Für den Fall $R = K[X]$):
Sei $\chi_f = \prod_{j=1}^r q_j^{\mu_j}$ die Zerlegung von χ_f in ein Produkt von Potenzen paarweise verschiedener normierter irreduzibler Polynome und $W_j := \mathrm{Ker}(q_j^{\mu_j}(f))$.
Nach Satz 14.12 über die verallgemeinerte Hauptraumzerlegung ist dann $V = \oplus_j W_j$, und wir können jedes W_j als Modul über dem Ring $R = K[X]$ auffassen, indem wir setzen:

$$P \cdot v := P(f)(v) \quad (v \in W_j, P \in K[X]).$$

Da $W := W_j$ schon als K- Vektorraum endlich erzeugt ist, ist erst recht der $K[X]$-Modul $W = W_j$ endlich erzeugt, und wir haben $q_j^{\mu_j} W = \{\mathbf{0}\}$.
Für W finden wir jetzt Vektoren $x_1^{(j)}, \dots, x_{m_j}^{(j)}$ und normierte $c_1^{(j)}, \dots, c_{m_j}^{(j)} \in K[X]$ wie in Satz 15.6b). Da alle Vektoren in W von $q_j^{\mu_j}$ annulliert werden, muss $c_i^{(j)}$ für jedes i ungleich 0 und ein Teiler von $q_j^{\mu_j}$ sein, also (wegen der Eindeutigkeit der Zerlegung in Potenzen irreduzibler Polynome) von der Form $q_j^{\nu_j}$ mit $\nu_j \leq \mu_j$ sein.
Für $W = W_j$ haben wir damit Vektoren wie in Satz 14.14 gefunden; diese Teilsysteme fügen sich dann zu der in diesem Satz geforderten Menge von Vektoren für ganz V zusammen.
Alternativ kann man einen für die Fälle $R = \mathbb{Z}$ und $R = K[X]$ (und sogar für alle Hauptidealringe) einheitlichen Beweis führen, indem man zunächst V als $V \cong \oplus R/c_j R$ mit $c_j \mid c_{j+1}$ zerlegt und dann mit Hilfe der Zerlegung von c_j in ein Produkt von Potenzen $(q_i^{(j)})^{\nu_{i,j}}$ von Primelementen und des oben erwähnten chinesischen Restsatzes eine Zerlegung wie in Satz 14.14 daraus ableitet (Übung). $\square$

Wir können die Ergebnisse über Elementarteiler auch noch verwenden, um ein Kriterium für Ähnlichkeit von Matrizen in $M_n(K)$ für einen beliebigen Körper K herzuleiten.

Satz 15.8. *Sei K ein Körper, seien $A, B \in M(n \times n, K)$ gegeben. Dann sind äquivalent:*

a) *A und B sind ähnlich (konjugiert) zueinander.*
b) *Die charakteristischen Matrizen $XE_n - A, XE_n - B \in M(n \times n, K[X])$ von A, B sind äquivalent über $K[X]$ (also $XE_n - A = S(XE_n - B)T$ mit $S, T \in GL_n(K[X])$).*
c) *Der durch die Multiplikation*

$$(\sum_i a_i X^i).\mathbf{v} \mapsto \sum_i a_i (A^i \mathbf{v})$$

von Elementen von $K[X]$ mit Elementen von K^n definierte $K[X]$-Modul M_A (mit zu Grunde liegender abelscher Gruppe K^n) ist isomorph zum analog definierten $K[X]$-Modul M_B.

Beweis. Für den Beweis sei auf das Buch von Lorenz verwiesen, wir kommen später (wenn das Tensorprodukt zur Verfügung steht) noch einmal darauf zurück. □

Korollar 15.9. *Sei K ein Körper, seien $A, B \in M(n \times n, K)$ gegeben. Dann gilt: A und B sind genau dann ähnlich (konjugiert) zueinander, wenn ihre charakteristischen Matrizen (bis auf Multiplikation mit Einheiten in $K[X]$) die gleichen Determinantenteiler haben.*

Beweis. Das folgt aus dem vorigen Satz und Satz 15.4. □

Bemerkung. Ob zwei $n \times n$- Matrizen über K ähnlich (konjugiert) zueinander sind, kann also im Prinzip dadurch entschieden werden, dass man alle $j \times j$- Unterdeterminanten der jeweiligen charakteristischen Matrizen berechnet. In der Regel wird es für praktische Zwecke einfacher sein, den in Satz 15.4 beschriebenen modifizierten Gauß-Algorithmus durchzuführen.

16. Multilineare Algebra und Tensorprodukt

In diesem Abschnitt geht es darum, multiplikative Strukturen auf Vektorräumen und allgemeiner Moduln über kommutativen Ringen zu beschreiben.
Im Weiteren ist stets R ein kommutativer Ring mit 1, mit U, V, W werden R-Moduln bezeichnet.

Definition 16.1. *Eine Abbildung $\beta : U \times V \to W$ heißt* bilinear*, falls sie linear in jedem Argument ist, falls also für alle $u, u_1, u_2 \in U, v, v_1, v_2 \in V, \lambda \in R$ gilt:*

$$\begin{aligned} \beta(\lambda u_1 + u_2, v) &= \lambda\beta(u_1, v) + \beta(u_2, v) \\ \beta(u, \lambda v_1 + v_2) &= \lambda\beta(u, v_1) + \beta(u, v_2). \end{aligned}$$

*Analog sind k-fach lineare Abbildungen (*multilineare Abbildungen*) für beliebiges $k \in \mathbb{N}$ definiert.*

Lemma 16.2. *Sind $(u_i)_{i\in I}, (v_j)_{j\in J}$ Basen von U bzw. V, so gibt es zu jeder Familie $(w_{ij})_{i\in I, j\in J}$ von Elementen w_{ij} in W genau eine bilineare Abbildung $\beta : U \times V \to W$ mit*

$$\beta((u_i, v_j)) = w_{ij} \text{ für alle } i \in I, j \in J.$$

Beweis. Man setzt die Vorgabe $\beta((u_i, v_j)) = w_{ij}$ für alle $i \in I, j \in J$ bilinear fort, was wegen der Basiseigenschaft in eindeutiger Weise möglich ist. □

Beispiel:

a) Sei K ein Körper, $V = K[X]$. Man hat die bilineare Abbildung

$$(\sum_{i=1}^{n} a_i X^i, \sum_{j=1}^{m} b_j X^j) \mapsto \sum_{k=1}^{m+n} c_k X^k \text{ mit } c_k = \sum_{i+j=k} a_i b_j \tag{16.1}$$

von $K[X] \times K[X]$ in $K[X]$.

b) Mit K und V wie oben hat man die bilineare Abbildung

$$(\sum_{i=1}^{n} a_i X^i, \sum_{j=1}^{m} b_j X^j) \mapsto \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j X_1^i X_2^j \in K[X_1, X_2] \tag{16.2}$$

von $K[X] \times K[X]$ in den Polynomring $K[X_1, X_2] = (K[X_1])[X_2]$ in zwei Variablen X_1, X_2.

c) Sei jetzt $V = K^3$. Man hat das aus der analytischen Geometrie der Oberstufe bekannte *Kreuzprodukt*

$$\mathbf{x} \times \mathbf{y} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}, \tag{16.3}$$

das eine bilineare Abbildung $K^3 \times K^3 \to K^3$ definiert.

d) U_1, V_1 seien K-Vektorräume mit Dualräumen $U := U_1^*, V := V_1^*$. Bezeichnet man mit $\mathrm{Bil}_K(U_1 \times V_1)$ den Vektorraum der bilinearen Abbildungen von $U_1 \times V_1$ nach K (Bilinearformen auf $U_1 \times V_1$), so hat man die folgende bilineare Abbildung

$$\begin{aligned} T \ : \ & U \times V \longrightarrow \mathrm{Bil}_K(U_1 \times V_1) \\ & (f,g) \longmapsto T(f,g) \quad \text{mit } T(f,g)(u,v) = f(u)g(v). \end{aligned}$$

Seien jetzt U_1 und V_1 endlichdimensional mit Basen $(u_1, \ldots, u_m), (v_1, \ldots, v_n)$; wir haben dann in U und V die dazu dualen Basen $(u_1^*, \ldots, u_m^*), (v_1, \ldots, v_n^*)$ Im Bild von T befinden sich dann insbesondere die $T(u_i^*, v_j^*) =: B_{ij}$, für die

$$T(u_i^*, v_j^*)(u_k, v_l) = B_{ij}(u_k, v_l) = \begin{cases} 1 & \text{falls } i = k, j = l \\ 0 & \text{sonst} \end{cases}$$

gilt.
Da diese Bilinearformen B_{ij} offenbar eine Basis von $\mathrm{Bil}_K(U_1 \times V_1)$ bilden, wird dieser Raum vom Bild von T erzeugt. Man beachte, dass das Bild einer bilinearen Abbildung im Allgemeinen kein Vektorraum ist, im hier betrachteten Fall besteht das Bild aus allen Bilinearformen, deren Matrix $(\beta(u_i, v_j))$ sich als $\mathbf{x}\,{}^t\mathbf{y} \in M(m \times n, \mathbb{R})$ mit $\mathbf{x} \in K^m, \mathbf{y} \in K^n$ schreiben lässt, also Rang 1 hat.

Die erste dieser bilinearen Abbildungen ist symmetrisch, die dritte antisymmetrisch (sogar alternierend), die zweite ist weder symmetrisch noch antisymmetrisch, auf die letzte lassen sich diese Begriffe nicht anwenden, da die zugrundeliegenden Vektorräume U und V nicht gleich sind.

Definition 16.3. *Seien R-Moduln U, V gegeben. Ein R-Modul X zusammen mit einer bilinearen Abbildung $T := \otimes : U \times V \to X$, heißt* Tensorprodukt *von U und V, falls das Paar $(X, \otimes)$ folgende (universelle) Eigenschaft hat:*
Ist W irgendein R-Modul und $\beta : U \times V \to W$ eine bilineare Abbildung, so gibt es genau eine lineare Abbildung $\tilde{\beta} : X \to W$, die das Diagramm

(16.4)
$$\begin{tikzcd} & X \arrow[d, "\tilde{\beta}"] \\ U \times V \arrow[ur, "\otimes"] \arrow[r, "\beta"] & W \end{tikzcd}$$

kommutativ macht.
Man schreibt dann $X = U \otimes V = U \otimes_R V$ und notiert die Abbildung $T = \otimes$ als $(u, v) \mapsto T(u, v) = u \otimes v$.

Satz 16.4. *Seien R-Moduln U, V gegeben. Dann existiert das Tensorprodukt von U und V und ist bis auf (eindeutige) Isomorphie eindeutig bestimmt. Genauer:*

Sind $(W_1, \otimes_1), (W_2, \otimes_2)$ *beide wie in Definition 16.3, so gibt es genau einen Isomorphismus* $\varphi : W_1 \to W_2$, *so dass das Diagramm*

(16.5)
$$\begin{tikzcd} & W_1 \arrow[dd, "\varphi"] \\ U \times V \arrow[ur, "\otimes_1"] \arrow[dr, "\otimes_2"'] & \\ & W_2 \end{tikzcd}$$

kommutativ ist. Man spricht daher von dem *Tensorprodukt von* U *und* V.

Beweis. Zunächst zur Eindeutigkeit:
Sind W_1, W_2 mit bilinearen Abbildungen $T_j : U \times V \longrightarrow W_j$ Tensorprodukte von U und V, so gibt es nach Definition lineare Abbildungen $\varphi_1 : W_1 \longrightarrow W_2, \varphi_2 : W_2 \longrightarrow W_1$ mit $\varphi_1 \circ T_1 = T_2, \varphi_2 \circ T_2 = T_1$. Dann ist $\varphi_1 \circ \varphi_2 \circ T_2 = T_2, \varphi_2 \circ \varphi_1 \circ T_1 = T_1$, und die Eindeutigkeitsanforderung in der Definition des Tensorprodukts impliziert $\varphi_1 \circ \varphi_2 = \mathrm{Id}_{W_2}, \varphi_2 \circ \varphi_1 = \mathrm{Id}_{W_1}$, die Abbildungen φ_1, φ_2 sind also zueinander inverse Isomorphismen, die (wiederum wegen der Eindeutigkeitsanforderung in der Definition des Tensorprodukts) eindeutig bestimmt sind.
Zum Nachweis der Existenz eines Tensorprodukts gibt es im Wesentlichen zwei Varianten:
Variante 1 (für freie Moduln, also Moduln, die eine Basis besitzen):
Ist $(u_i)_{i \in I}$ eine Basis von U und $(v_j)_{j \in J}$ eine Basis von V, so sei X ein R-Modul mit einer Basis $(x_{ij})_{(i,j) \in I \times J}$ (etwa $X = R^{(I \times J)}$). Man definiert dann $T : U \times V \longrightarrow X$ als die eindeutig bestimmte bilineare Abbildung mit $T(u_i, v_j) = x_{ij}$ für alle $i \in I, j \in J$ und rechnet mit Hilfe von Lemma 16.2 nach, dass das Paar (X, T) in der Tat die charakteristische (universelle) Eigenschaft des Tensorprodukts hat.
Variante 2 (basisunabhängig, für beliebige Moduln): Sei X' ein R-Modul mit einer Basis $(x_{(u,v)})_{(u,v) \in U \times V}$, etwa $X' = R^{(U \times V)}$. In X' sei N der Untermodul, der von allen Elementen der Form

$$\begin{aligned} x_{(u+u',v)} - x_{(u,v)} - x_{(u',v)} \\ x_{(u,v+v')} - x_{(u,v)} - x_{(u,v')} \\ x_{(\lambda u,v)} - \lambda x_{(u,v)} \\ x_{(u,\lambda v)} - \lambda x_{(u,v)} \end{aligned}$$

erzeugt wird, sei $X = X'/N$ der Faktormodul von X' nach N.
Die Abbildung $T(u, v) := x_{(u,v)} + N$ ist dann bilinear und man rechnet wiederum (jetzt mit Hilfe des Homomorphiesatzes für Moduln) nach, dass das Paar (X, T) in der Tat die charakteristische (universelle) Eigenschaft des Tensorprodukts hat. $\square$

Beispiel: Sind $U_1, V_1, U = U_1^*, V = V_1^*$ wie in d) des vorigen Beispiels von endlicher Dimension, so hat $X := \mathrm{Bil}_K(U_1 \times V_1)$ mit der Abbildung $T : U \times V \longrightarrow X := \mathrm{Bil}_K(U_1 \times V_1)$ die in der Definition eines Tensorprodukts von U und V geforderte Eigenschaft.
Dass T bilinear ist, haben wir bereits gesehen. Sind W und $\beta : U \times V \longrightarrow W$ wie in der Definition und ist $\beta(u_i^*, v_j^*) =: w_{ij} \in W$, so definieren wir $\tilde{\beta} : \mathrm{Bil}_K(U_1 \times V_1) \longrightarrow W$ als die (eindeutig bestimmte) lineare Abbildung, die auf den Basisvektoren B_{ij} von $\mathrm{Bil}_K(U_1 \times V_1)$ durch

$$\tilde{\beta}(B_{ij}) = w_{ij}$$

gegeben ist, für diese gilt offenbar $\tilde{\beta} \circ T = \beta$. Sie ist auch die einzige lineare Abbildung von $\mathrm{Bil}_K(U_1 \times V_1)$ nach W, die das Diagramm kommutativ macht, denn für jede derartige Abbildung $\hat{\beta}$ muss

$$\hat{\beta}(B_{ij}) = \hat{\beta}(T((u_i^*, v_j^*))) = w_{ij}$$

gelten.

Bemerkung. In beiden Beweisen sieht man, dass die Elemente $u \otimes v$ mit $u \in U, v \in V$ ein Erzeugendensystem des Raums $U \otimes V$ bilden; diese Elemente werden auch *reine Tensoren* oder Tensoren vom Rang 1 genannt. Ist keiner der Räume U, V eindimensional, so gibt es Elemente von $U \otimes V$, die nicht von dieser Form sind, siehe das Beispiel nach dem nächsten Korollar.

Korollar 16.5. *Seien freie R-Moduln U, V gegeben, seien $(u_i)_{i \in I}, (v_j)_{j \in J}$ Basen von U bzw. V.*

a) *Die Familie der $(u_i \otimes v_j)_{i \in I, j \in J}$ ist eine Basis von $U \otimes V$.*
b) *Ist $\dim(U) = m, \dim(V) = n$, so ist $\dim(U \otimes V) = mn$.*
c) *Ist $w \in U \otimes V$, so gibt es eindeutig bestimmte Vektoren $x_j \in U (j \in J), y_i \in V (i \in I)$, so dass gilt:*

$$w = \sum_{j \in J} x_j \otimes v_j = \sum_{i \in I} u_i \otimes y_i.$$

Beweis. Übung. □

Beispiel: Als Übung zeige man, dass sich das Element

$$\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 \in \mathbb{R}^2 \otimes \mathbb{R}^2$$

nicht als $u \otimes v$ mit $u, v \in \mathbb{R}^2$ schreiben lässt.

Korollar 16.6. *Seien endlichdimensionale K-Vektorräume U, V gegeben, seien $(u_i)_{1 \le i \le n}, (u_i')_{1 \le i \le n}, (v_j)_{1 \le j \le m}, (v_j')_{1 \le j \le m}$ Basen von U bzw. V mit $u_i = \sum_{k=1}^n t_{ki} u_k', v_j = \sum_{l=1}^m s_{lj} v_l'$, T, S die zugehörigen Matrizen. Dann gilt:*

a) *Ist* $x = \sum_{i,j} a_{ij} u_i \otimes v_j = \sum_{k,l} a'_{kl} u'_k \otimes v'_l \in U \otimes V, A = (a_{ij}), A' = (a'_{kl}) \in M(m \times n, K)$, *so ist*

$$A' = TA^tS.$$

b) *Sind* $(u_i^*), (u'^*_i)$ *die zugehörigen dualen Basen von* U^*, *und* $F = \sum_{i,j} a_{ij} u_i^* \otimes v_j = \sum_{k,l} a'_{kl} u'^*_k \otimes v'_l \in U^* \otimes V$, *so ist*

$$A' = {}^tT^{-1} A^t S.$$

Beweis. a) rechnet man direkt nach.
b) folgt aus a), wenn man weiß, dass $u'^*_i = \sum_k t_{ik} u_k^*$ für $1 \leq i \leq n$ gilt. Das sieht man mit Hilfe von $u'^*_i(u_j) = \sum_\ell u'^*_i(t_{\ell j} u'_\ell) = t_{ij} = \sum_k t_{ik} u_k^*(u_j)$ für alle j. □

Bemerkung. In der Physik werden häufig Tensoren als Koeffizientenschemata mit gewissen Transformationseigenschaften bei Wechsel des Koordinatensystems definiert (etwa in den "Feynman Lectures on Physics"). Das obige Korollar zeigt, dass solche Koeffizientenschemata gerade als die Koeffizienten bezüglich der angebenen Basen der Tensorprodukte auftreten. Für Verwirrung sorgt dabei gelegentlich die Tatsache, dass sich für eine orthogonale Matrix T der Unterschied zwischen $U \otimes V$ und $U^* \otimes V$ wegen ${}^tT^{-1} = T$ nicht in den Transformationseigenschaften auswirkt.

Beispiel:

a) Koeffizientenerweiterung:
Sei V ein R-Modul, $S \supseteq R$ ein Ring, der R enthält (eine Ringerweiterung), man denke etwa an $R = K = \mathbb{R}, S = L = \mathbb{C}$. Der Ring S kann auch als R-Modul aufgefasst werden (bezüglich der in S definierten Multiplikation von Elementen von R mit Elementen in S), man kann also das Tensorprodukt von R-Moduln $S \otimes_R V$ bilden. Das ist zunächst ein R-Modul.
Man kann jetzt aber auch eine multiplikative Verknüpfung von Elementen des Rings S mit Elementen von $S \otimes_R V$ definieren: Für $\lambda \in S$ wird durch $(a, v) \mapsto (\lambda a) \otimes v \quad (a \in S, v \in V)$ eine bilineare Abbildung $M_\lambda : S \times V \to S \otimes_R V$ definiert, die auf Grund der universellen Eigenschaft des Tensorprodukts zu einer linearen Abbildung $\overline{M_\lambda} : S \otimes_R V \to S \otimes_R V$ mit $\overline{M_\lambda}(a \otimes v) = (\lambda a) \otimes v \quad (a \in S, v \in V)$ führt. Man prüft leicht nach, dass durch

$$\lambda.w := \overline{M_\lambda}(w) \quad (\lambda \in S, w \in S \otimes_R V)$$

eine Struktur eines R-Moduls auf der abelschen Gruppe $S \otimes_R V$ eingeführt wird, bezüglich der $\lambda. \sum_i a_i \otimes v_i = \sum_i (\lambda a_i) \otimes v_i$ gilt. Man nennt $S \otimes_R V$ mit dieser S-Modulstruktur die *Koeffizientenerweiterung* V_S von V nach S. Sind R und S Körper, so hat V_S als S-Vektorraum die gleiche Dimension wie sie V als R-Vektorraum hat, eine Basis (v_i) von V über R führt zu der Basis $(1 \otimes v_i)$ von V_S

über S. Speziell für $R = \mathbb{R}, S = \mathbb{C}$ heißt $V_{\mathbb{C}}$ die *Komplexifizierung* von V.

Durch die Konstruktion mittels des Tensorprodukts liefert die Koeffizientenerweiterung eine basisfreie Verallgemeinerung der natürlichen Inklusion $\mathbb{R}^n \subseteq \mathbb{C}^n$, die wir schon mehrfach benutzt haben. Ein häufig vorkommender Spezialfall ist $V = K[X]$ mit $V_L \cong L[X]$ (in natürlicher Weise).

b) Sei $U = V = K[X]$. Man rechnet leicht nach, dass die zu Anfang dieses Abschnitts betrachtete bilineare Abbildung $\varphi : K[X] \times K[X] \to K[X_1, X_2]$, die durch $(\sum_i a_i X^i, \sum_j b_j X^j) \mapsto \sum_i \sum_j a_i b_j X_1^i X_2^j$ gegeben ist, zu einem Isomorphismus $\bar{\varphi} : K[X] \otimes K[X] \to K[X_1, X_2]$ führt.

c) Auf $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ kann man in ähnlicher Weise wie in a) eine (assoziative und distributive) Multiplikation definieren, für die $(z_1 \otimes z_2)(z_1' \otimes z_2') = z_1 z_1' \otimes z_2 z_2'$ gilt, damit wird $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ zu einem Ring. Man überlege sich als Übung, ob dieser Ring ein Körper ist (Warnung: Der offensichtliche Versuch, die Inversenbildung durch $(z_1 \otimes z_2)^{-1} = z_1^{-1} \otimes z_2^{-1}$ zu definieren, stößt zumindest auf Schwierigkeiten, weil sich nicht jedes Element des Tensorprodukts in dieser Form schreiben lässt.)

Bemerkung. Sind $U_1, \ldots, U_n$ Moduln über dem Ring R, so kann man ganz analog zu $U \otimes V$ ein n-faches Tensorprodukt $U_1 \otimes \cdots \otimes U_n$ mit einer n-fach linearen Abbildung $T_n : U_1 \times \cdots \times U_n \to U_1 \otimes \cdots \otimes U_n$, geschrieben als $T_n(u_1, \ldots, u_n) = u_1 \otimes \cdots \otimes u_n$ definieren. Es hat die universelle Eigenschaft, dass jede n-fach lineare Abbildung $g : U_1 \times \cdots \times U_n \to W$ über $U_1 \otimes \cdots \otimes U_n$ faktorisiert, also genau eine lineare Abbildung $\gamma : U_1 \otimes \cdots \otimes U_n \to W$ liefert, für die $\gamma \circ T_n = g$ gilt.

Satz 16.7. *U, V, W seien R-Moduln. Dann gibt es natürliche Isomorphismen*

a)

$$\begin{array}{rcl} U \otimes V & \xrightarrow{\cong} & V \otimes U \quad \textit{mit} \\ u \otimes v & \mapsto & v \otimes u \end{array}$$

b)

$$\begin{array}{rcl} (U \otimes V) \otimes W & \xrightarrow{\cong} & U \otimes (V \otimes W) \xrightarrow{\cong} U \otimes V \otimes W \quad \textit{mit} \\ (u \otimes v) \otimes w & \mapsto & u \otimes (v \otimes w) \mapsto u \otimes v \otimes w \end{array}$$

c)

$$\begin{array}{rcl} (U \oplus V) \otimes W & \xrightarrow{\cong} & (U \otimes W) \oplus (V \otimes W) \quad \textit{mit} \\ (u + v) \otimes w & \mapsto & u \otimes w + v \otimes w \end{array}$$

$$\begin{array}{rcl} R \otimes U & \stackrel{\cong}{\to} & U \quad \textit{mit} \\ a \otimes u & \mapsto & au \end{array}$$

Satz 16.8. *Seien R-Moduln U, V, U', V' gegeben, seien $f : U \longrightarrow U'$, $g : V \longrightarrow V'$ lineare Abbildungen.*

a) *Es gibt genau eine lineare Abbildung $f \otimes g : U \otimes V \longrightarrow U' \otimes V'$ mit*
$$(f \otimes g)(u \otimes v) = f(u) \otimes g(v) \textit{ für alle } u \in U, v \in V.$$
b) *Sind f, g beide surjektiv (bijektiv), so auch $f \otimes g$.*
c) *Ist $R = K$ ein Körper und sind f, g beide injektiv, so auch $f \otimes g$.*

Beweis. Sei $f \times g : U \times V \longrightarrow U' \otimes V'$ durch $(f \times g)(u, v) = f(u) \otimes g(v)$ gegeben. Die Abbildung $f \times g$ ist, wie man nachrechnet, bilinear, liefert also ein (eindeutig bestimmtes) lineares $(f \otimes g) : U \otimes V \longrightarrow U \otimes U'$ mit
$$(f \otimes g)(u \otimes v) = (f \times g)(u, v) = f(u) \otimes g(v)$$
für alle $u \in U, v \in V$, dieses ist die gesuchte Abbildung.
b) und c) zeige man als Übung. □

Satz 16.9. *U, V, W seien R-Moduln. Dann gibt es die folgenden natürlichen Isomorphismen:*

a)
$$\mathrm{Hom}(U \otimes V, W) \stackrel{\cong}{\to} \mathrm{Bil}_R(U \times V, W) \stackrel{\cong}{\to} \mathrm{Hom}(U, \mathrm{Hom}(V, W)),$$
dabei ist $\mathrm{Bil}_R(U \times V, W)$ der R-Modul der bilinearen Abbildungen von $U \times V$ in W.

b)
$$(U \otimes V)^* \stackrel{\cong}{\to} \mathrm{Bil}_R(U \times V, R)$$

c)
$$U^* \otimes V \stackrel{\cong}{\to} \mathrm{Hom}(U, V),$$
falls $R = K$ ein Körper ist und U, V endliche Dimension haben.

Beweis. a): Nach Definition des Tensorprodukts gibt es zu jedem $\beta \in \mathrm{Bil}_R(U \times V, W)$ genau eine zugehörige lineare Abbildung $\bar{\beta} : U \otimes V \to W$, und die Zuordnung $\beta \mapsto \bar{\beta}$ ist offenbar linear und injektiv. Sie ist auch surjektiv, denn $g : U \otimes V \to W$ ist das Bild der durch $(u, v) \mapsto g(u \otimes v)$ gegebenen bilinearen Abbildung. Das zeigt die erste Isomorphie. Für die zweite bilden wir $\beta \in \mathrm{Bil}_R(U \times V, W)$ auf die Abbildung $F_\beta : U \to \mathrm{Hom}(V, W)$ ab, die durch $F_\beta(u)(v) = \beta(u, v)$ gegeben ist; offenbar ist $F_\beta(u)$ linear, und auch die Abbildung $\beta \mapsto F_\beta$ ist linear. Umgekehrt bilden wir $F \in \mathrm{Hom}(U, \mathrm{Hom}(V, W))$ auf die durch $\beta_F(u, v) := F(u)(v)$ gegebene Abbildung $\beta_F : U \times V \to W$ ab. Man rechnet nach, dass β_F bilinear ist, dass $F \mapsto \beta_F$ linear ist, und dass $\beta \mapsto F_\beta$ und $F \mapsto \beta_F$ zueinander invers sind.

b): Ist der Spezialfall $W = R$ des ersten Teils von a).
c): Für $\varphi \in U^*, v \in V$ sei $f_{\varphi,v}$ die durch $f_{\varphi,v}(u) = \varphi(u)v$ gegebene Abbildung von U nach V. Man rechnet nach, dass $f_{\varphi,v}$ linear ist und dass die Abbildung $(\varphi, v) \mapsto f_{\varphi,v}$ eine bilineare Abbildung von $U \times V$ in $\mathrm{Hom}(U, V)$ ist. Es gibt also genau eine lineare Abbildung $F : U^* \otimes V \to \mathrm{Hom}(U, V)$, so dass $F(\varphi \otimes v)(u) = \varphi(u)v$ für alle $\varphi \in U^*, v \in V, u \in U$ gilt.
Die Abbildung F ist injektiv, denn ist $(v_1, \ldots, v_n)$ eine R-Basis von V und $x = \sum_{i=1}^n \varphi_i \otimes v_i \in U^* \otimes V$ mit $F(x) = 0$, so ist $\sum_{i=1}^n \varphi_i(u)v_i = 0$ für alle $u \in U$, also ist $\varphi_i = 0$ für alle i und damit $x = 0$. Da $U^* \otimes V$ und $\mathrm{Hom}(U, V)$ als R-Vektorräume die gleiche Dimension haben, ist F auch surjektiv. □

Bemerkung. Wir hatten im Beispiel d) nach Lemma 16.2 für endlich dimensionale K-Vektorräume eine surjektive bilineare Abbildung $U_1^* \times V_1^* \to \mathrm{Bil}_K(U_1 \times V_1, K)$ konstruiert. Aus Dimensionsgründen ist dann die zugehörige lineare Abbildung $U_1^* \otimes V_1^* \to \mathrm{Bil}_K(U_1 \times V_1, K)$ ebenfalls ein (in natürlicher Weise definierter) Isomorphismus.

Satz 16.10. *Sei V ein endlich dimensionaler Vektorraum über dem Körper K, $U := K[X] \otimes V$ die Koeffizientenerweiterung zu $K[X]$, $f \in \mathrm{End}(V)$ und V_f der mittels $p *_f v := p(f)(v)$ für $p \in K[X], v \in V$ als $K[X]$-Modul aufgefasste K-Vektorraum V.*

a) *Die K-linearen Abbildungen $X \otimes \mathrm{Id}_V$ und $1 \otimes f$ von U in sich sind sogar $K[X]$-linear, und $N_f := \mathrm{Im}(X \otimes \mathrm{Id}_V - 1 \otimes f)$ ist ein $K[X]$-Untermodul von U.*
b) *Es gibt genau eine $K[X]$-lineare Abbildung $\Phi_f : U \to V_f$ mit $\Phi_f(p \otimes v) = p(f)(v)$ für alle $p \in K[X], v \in V$. Für diese gilt $N_f = \mathrm{Ker}(\Phi_f)$, und Φ_f induziert einen Isomorphismus $\overline{\Phi_f} : W_f := U/N_f \to V_f$ von $K[X]$-Moduln.*

Beweis. Dass $1 \otimes f$ eine $K[X]$-lineare Abbildung ist, ist klar, und für $X \otimes \mathrm{Id}_V$ folgt die $K[X]$-Linearität daraus, dass $K[X]$ kommutativ ist, also insbesondere $X \cdot p = p \cdot X$ für alle $p \in K[X]$ gilt. Damit ist natürlich auch N_f als Bild einer $K[X]$-linearen Abbildung ein $K[X]$-Untermodul.
Für b) bekommt man Φ_f als die zur K-bilinearen Abbildung $(p, v) \mapsto p(f)(v)$ gehörige lineare Abbildung von $U = K[X] \otimes V \to V = V_f$. Dass Φ_f dann auch $K[X]$-linear ist, liegt an der Definition der $K[X]$-Modulstruktur von V_f.
Ist $w := (X \otimes \mathrm{Id}_V - 1 \otimes f)(p \otimes v) = X \cdot p \otimes v - p \otimes f(v) \in N_f$, so ist $\Phi_f(w) = f \circ p(f)(v) - p(f)(f(v)) = \mathbf{0}$, also ist $N_f \subseteq \mathrm{Ker}(\Phi_f)$, und nach dem Homomorphiesatz für Moduln bekommen wir eine $K[X]$-lineare Abbildung $\overline{\Phi_f} : W_f = U/N_f \to V_f$, die ebenso wie Φ_f surjektiv ist.
Umgekehrt können wir durch $\Psi(v) := 1 \otimes v + N_f$ eine K-lineare Abbildung $\Psi : V_f \to W_f$ definieren, von der man mit Hilfe der Definition von N_f leicht nachrechnet, dass sie sogar $K[X]$-linear ist. Ψ und $\overline{\Phi_f}$ sind

aber, wie man ebenfalls sofort nachrechnet, zueinander invers, also ist $\overline{\Phi_f}$ wie behauptet ein $K[X]$-Isomorphismus.

Alternativ können wir auch ein Dimensionsargument für den Beweis der Injektivität verwenden:
Als K-Vektorraum wird W_f von den $(1\otimes v_i)+N_f$ erzeugt, wo $(v_1,\ldots,v_n)$ eine beliebige Basis des K-Vektorraums V ist: Wir können nämlich jedes Element u von U als $u=\sum_{i=1}^n p_i\otimes v_i$ mit $p_i\in K[X]$ schreiben und haben dann

$$u=\sum_{i=1}^n 1\otimes p_i(f)(v_i)+(\sum_{i=1}^n(p_i\otimes v_i-1\otimes p_i(f)(v_i)))$$

mit $\sum_{i=1}^n(p_i\otimes v_i-1\otimes p_i(f)(v_i))\in N_f$ und $\sum_{i=1}^n 1\otimes p_i(f)(v_i)$ im K-Erzeugnis $\{1\otimes v\mid v\in V\}$ der $1\otimes v_i$.
Also ist $\dim_K(W_f)\le n=\dim_K(V)$, und die surjektive K-lineare Abbildung Φ_f muss auch injektiv sein (und die $(1\otimes v_i)+N_f$ sind sogar eine Basis von W_f). □

Korollar 16.11. *Sei K ein Körper, $n\in\mathbb{N}$, seien $A,B\in M_n(K)$. Dann sind A und B genau dann ähnlich in $M_n(K)$, wenn es $S,T\in GL_n(K[X])$ gibt mit $S(XE_n-A)=(XE_n-B)T$.*

Beweis. Sind A und B ähnlich in $M_n(K)$, so sind sie das erst recht in $M_n(K[X])$, und man hat trivialerweise S und T wie gewünscht.

Umgekehrt seien $S,T\in GL_n(K[X])$ mit $S(XE_n-A)=(XE_n-B)T$ gegeben. Wir betrachten die Situation des vorigen Satzes mit $V=K^n$ für die Endomorphismen $f=L_A$ und $g=L_B$, wir schreiben dann V_A,N_A,W_A,Φ_A statt der entsprechenden Notationen mit dem Index f und entsprechend für B.
Ein Element $u=(X\otimes\mathrm{Id}_V-1\otimes L_A)(1\otimes v)$ von N_A können wir dann als $(XE_n-A)v$ schreiben, wobei hier $v\in K^n$ als Element von $K[X]^n$ aufgefasst wird. Dann ist $L_Su=S(XE_n-A)v=(XEn-B)Tv\in N_B$, und da N_A als $K[X]$-Modul von Elementen u dieses Typs erzeugt wird, folgt $L_S(N_A)\subseteq N_B$. Wegen der Invertierbarkeit von S und T können wir genauso zeigen, dass $L_S(N_A)\supseteq N_B$ gilt und folgern, dass L_S nach dem Homomorphiesatz für $K[X]$-Moduln einen $K[X]$-Isomorphismus $\sigma:W_A\to W_B$ von $K[X]$-Moduln induziert.
Nach dem vorigen Satz sind dann auch die $K[X]$-Moduln V_A und V_B isomorph. Da ein solcher Isomorphismus erst recht eine K-lineare Abbildung ist, können wir ihn als Multiplikation mit einer geeigneten invertierbaren Matrix $R\in GL_n(K)$ schreiben; für diese gilt dann $R\cdot(X*_A v)=X*_B(Rv)$, d.h., $R\cdot Av=B\cdot Rv$ für alle $v\in V$, also $RA=BR$, also $B=RAR^{-1}$. □

Lemma 16.12. *Sei V ein R-Modul, $k \in \mathbb{N} \setminus \{0\}, \sigma \in S_k$ eine Permutation, sei*

$$V^{\otimes k} := \underbrace{V \otimes \cdots \otimes V}_{k\text{-mal}}.$$

Dann gibt es genau eine lineare Abbildung $L_\sigma \in End(V^{\otimes k})$ mit

$$L_\sigma(v_1 \otimes \cdots \otimes v_k) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \text{ für alle } v_1, \ldots, v_k \in V.$$

Ist $f \in \mathrm{End}(V)$ und $f^{\otimes k} \in \mathrm{End}(V^{\otimes k})$ die zugehörige lineare Abbildung von $V^{\otimes k}$ in sich, so gilt $L_\sigma \circ f^{\otimes k} = f^{\otimes k} \circ L_\sigma$.

Beweis. Die Abbildung $L'_\sigma : V^k \longrightarrow V^{\otimes k}$, die durch

$$L_\sigma(v_1, \ldots, v_k) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \text{ für alle } v_1, \ldots, v_k \in V$$

gegeben ist, ist, wie man nachrechnet, k-fach multilinear, liefert also eine eindeutig bestimmte lineare Abbildung $L_\sigma \in \mathrm{End}(V^{\otimes k})$ mit

$$L_\sigma(v_1 \otimes \cdots \otimes v_k) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \text{ für alle } v_1, \ldots, v_k \in V.$$

Die zweite Aussage ist klar. □

Definition 16.13. *Sei V ein R-Modul, $k \in \mathbb{N} \setminus \{0\}$.*

a) *Sei W_0 der von den $w - L_\sigma(w)$ ($w \in V^{\otimes k}, \sigma \in S_k$) erzeugte Untermodul von $V^{\otimes k}$. Dann heißt*

$$Sym^k(V) := V^{\otimes k}/W_0$$

die k-te symmetrische Potenz *von V. Die Klasse von $v_1 \otimes \cdots \otimes v_k$ in $Sym^k(V)$ wird mit $v_1 \vee \cdots \vee v_k$ bezeichnet.*

b) *Sei W_1 der von den $v_1 \otimes \cdots \otimes v_k$, in denen ein Vektor wenigstens zweimal vorkommt ($v_i = v_j$ für ein Paar (i, j) mit $i \neq j$), erzeugte Untermodul von $V^{\otimes k}$. Dann heißt*

$$\bigwedge^k V := V^{\otimes k}/W_1$$

die k-te äußere Potenz *(das k-fache* Graßmann-Produkt*) von V. Die Klasse von $v_1 \otimes \cdots \otimes v_k$ in $\bigwedge^k V$ wird mit $v_1 \wedge \cdots \wedge v_k$ bezeichnet.*

Bemerkung: Der Untermodul W_1 von $V^{\otimes k}$ enthält alle $w - \mathrm{sgn}(\sigma)(L_\sigma(w))$ mit $w \in V^{\otimes k}, \sigma \in S_k$, ist 2 invertierbar in R, so wird W_1 auch von diesen Elementen erzeugt.

Satz 16.14. *Sei V ein R-Modul, $k \in \mathbb{N} \setminus \{0\}$.*

a) *Ist $\beta : V^k \to X$ eine k-fach lineare symmetrische Abbildung in einen R-Modul X, so gibt es genau eine lineare Abbildung $\check{\beta} : Sym^k V \to X$ mit*

$$\check{\beta}(v_1 \vee \cdots \vee v_k) = \beta(v_1, \ldots, v_k) \text{ für alle } v_1, \ldots, v_k \in V.$$

b) *Ist $\alpha : V^k \to X$ eine k-fach lineare alternierende Abbildung in einen R-Modul X, so gibt es genau eine lineare Abbildung $\hat{\alpha} : \bigwedge^k V \to X$ mit*

$$\hat{\alpha}(v_1 \wedge \cdots \wedge v_k) = \alpha(v_1, \ldots, v_k) \text{ für alle } v_1, \ldots, v_k \in V.$$

Die Moduln $\mathrm{Sym}^k V$, $\bigwedge^k V$ mit den zugehörigen Abbildungen $(v_1, \ldots, v_k) \mapsto v_1 \vee \cdots \vee v_k$ and $(v_1, \ldots, v_k) \mapsto v_1 \wedge \cdots \wedge v_k$ sind durch diese universellen Eigenschaften bis auf eindeutige Isomorphie eindeutig bestimmt.

Beweis. b): Ist $\alpha : V^k \to X$ eine k-fach lineare Abbildung, so induziert sie eine lineare Abbildung $\tilde{\alpha} : V^{\otimes k} \to X$ mit $\tilde{\alpha}(v_1 \otimes \cdots \otimes v_k) = \alpha(v_1, \ldots, v_k)$ für alle $v_1, \ldots, v_k \in V$. Ist α zudem alternierend, so ist $W_1 \subseteq \mathrm{Ker}(\alpha)$, nach dem Homomorphiesatz gibt es also genau eine lineare Abbildung $\hat{\alpha} : V^{\otimes k}/W_1 = \bigwedge^k V \to X$, die $v_1 \wedge \cdots \wedge v_k$ für alle $v_1, \ldots, v_k$ auf $\alpha(v_1, \ldots, v_k)$ abbildet.
a) beweist man analog. □

Lemma 16.15. *Sei V ein R-Modul mit Basis $(v_1, \ldots, v_n)$, $k \in \mathbb{N}$, sei W ein weiterer R-Modul.*

a) *Für $i_1, \ldots, i_k \in \mathbb{N}$ mit $1 \leq i_1 \leq \cdots \leq i_k \leq n$ und beliebiges $w \in W$ gibt es genau eine symmetrische k-fach multilineare Abbildung $M^{sym}_{(i_1,\ldots,i_k)} : V^k \longrightarrow W$, so dass für $1 \leq j_1 \leq \cdots \leq j_k \leq n$ gilt:*

$$M^{sym}_{(i_1,\ldots,i_k)}((v_{j_1}, \ldots, v_{j_k})) = \begin{cases} w & \text{falls } i_1 = j_1, \ldots, i_k = j_k \\ \mathbf{0} & \text{sonst.} \end{cases}$$

b) *Für $i_1, \ldots, i_k \in \mathbb{N}$ mit $1 \leq i_1 < \ldots < i_k \leq n$ und beliebiges $w \in W$ gibt es genau eine alternierende k-fach multilineare Abbildung $M^{alt}_{(i_1,\ldots,i_k)} : V^k \longrightarrow W$, so dass für $1 \leq j_1 < \cdots < j_k \leq n$ gilt:*

$$M^{alt}_{(i_1,\ldots,i_k)}((v_{j_1}, \ldots, v_{j_k})) = \begin{cases} w & \text{falls } i_1 = j_1, \ldots, i_k = j_k \\ \mathbf{0} & \text{sonst.} \end{cases}$$

Beweis. Man beweise das als Übung durch multilineare und symmetrische bzw. alternierende Fortsetzung (in der Vorlesung im alternierenden Fall durchgeführt). □

Satz 16.16. *Sei V ein endlich erzeugter freier R-Modul mit Basis $(v_1, \ldots, v_n)$, $k \in \mathbb{N} \setminus \{0\}$.*

a) *Die $v_{i_1} \vee \cdots \vee v_{i_k}$ mit $1 \leq i_1 \leq \cdots \leq i_k \leq n$ bilden eine Basis von $\mathrm{Sym}^k V$. Ist $R = K$ ein Körper, so hat der Vektorraum $\mathrm{Sym}^k V$ die Dimension $\binom{n+k-1}{k}$.*
b) *Die $v_{i_1} \wedge \cdots \wedge v_{i_k}$ mit $1 \leq i_1 < \cdots < i_k \leq n$ bilden eine Basis von $\bigwedge^k V$. Ist $R = K$ ein Körper, so hat der Vektorraum $\bigwedge^k V$ die Dimension $\binom{n}{k}$.*

Beweis. b): Dass die $v_{i_1} \wedge \cdots \wedge v_{i_k}$ mit $1 \leq i_1 < \cdots < i_k \leq n$ den Modul $\bigwedge^k V$ erzeugen, folgt direkt daraus, dass die Abbildung $(w_1, \ldots, w_k) \mapsto w_1 \wedge \cdots \wedge w_k$ von V^k in $\bigwedge^k V$ alternierend und k-fach linear ist (hierfür würde es reichen, dass $(v_1, \ldots, v_n)$ den Modul V erzeugen). Dass diese Elemente linear unabhängig sind, folgt aus Teil b) von Lemma 16.15.
a) beweist man genauso. □

Bemerkung: Insbesondere ist $\bigwedge^k V = \{\mathbf{0}\}$, falls $k > n$ gilt, und der Vektorraum (falls $R = K$ ein Körper ist) $\bigwedge^n V$ hat für n-dimensionales V die Dimension 1.

Korollar 16.17. *Ist V ein K-Vektorraum der endlichen Dimension n, so hat der Vektorrraum $\mathrm{Alt}_n(V)$ der alternierenden n-fachen Multilinearformen auf V^n Dimension 1.*
Insbesondere: Ist $(v_1, \ldots, v_n)$ eine Basis von V, so gibt es genau eine alternierende n-fache Multilinearform δ auf V^n mit $\delta(v_1, \ldots, v_n) = 1$.

Beweis. Die universelle Eigenschaft des äußeren Produkts $\bigwedge^n V$ impliziert insbesondere, dass $\mathrm{Alt}_n(V)$ der Dualraum $(\bigwedge^n V)^*$ des eindimensionalen Raums $\bigwedge^n V$ ist. □

Beispiel Sei $V = \mathbb{R}^3$ mit der Standardbasis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Der Raum $\bigwedge^2 V = V \wedge V$ hat die Basis $w_1 = \mathbf{e}_2 \wedge \mathbf{e}_3, w_2 = \mathbf{e}_3 \wedge \mathbf{e}_1, w_3 = \mathbf{e}_1 \wedge \mathbf{e}_2$. Man rechnet nach:

$$\mathbf{x} \wedge \mathbf{y} = (x_2y_3 - x_3y_2)w_1 + (x_3y_1 - x_1y_3)w_2 + (x_1y_2 - x_2y_1)w_3.$$

Die Koordinaten von $\mathbf{x} \wedge \mathbf{y}$ bezüglich der Basis (w_1, w_2, w_3) sind also gerade die Komponenten des Kreuzprodukts (Vektorprodukts) $\mathbf{x} \times \mathbf{y}$ der Vektoren $\mathbf{x}, \mathbf{y}$.

Satz 16.18. *Sei K ein Körper, $V = K^n$ mit der Standardbasis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, sei $A = (a_{ij}) \in M_n(K)$. Dann gilt:*

a) $A\mathbf{e}_1 \wedge \cdots \wedge A\mathbf{e}_n = \det(A)\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n$
b) *Sind $r < n \in \mathbb{N}$ und $1 \leq j_1 < \cdots < j_r \leq n, 1 \leq i_1 < \cdots < i_r \leq n$ sowie $1 \leq k_1 < \cdots < k_{n-r} \leq n$ so, dass $\{k_1, \ldots, k_{n-r}, i_1, \ldots, i_r\} = \{1, \ldots, n\}$ gilt, so ist $(-1)^p A\mathbf{e}_{j_1} \wedge \cdots \wedge A\mathbf{e}_{j_r} \wedge \mathbf{e}_{k_1} \wedge \cdots \wedge \mathbf{e}_{k_{n-r}} = A^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n$, wo $A^{i_1,\ldots,i_r}_{j_1,\ldots,j_r}$ die Unterdeterminante zu den Zeilen $i_1, \ldots, i_r$ und den Spalten $j_1, \ldots, j_r$ der Matrix A ist und $p = i_1 + \cdots + i_r + \frac{r(r+1)}{2}$ gilt.*

Beweis. Man erhält beide Teile des Satzes, indem man $A\mathbf{e}_{j_1} \wedge \cdots \wedge A\mathbf{e}_{j_r} \wedge \mathbf{e}_{k_1} \wedge \cdots \wedge \mathbf{e}_{k_{n-r}}$ mittels des Distributivgesetzes (=Multilinearität) für $\wedge$ und der Regel $w_{\sigma(1)} \wedge \cdots \wedge w_{\sigma(r)} = \mathrm{sgn}(\sigma)w_1 \wedge \cdots \wedge w_r$ für $\sigma \in S_r, w_1, \ldots, w_r \in K^r$ expandiert (wobei im Fall $r = n$ der Anteil $\mathbf{e}_{k_1} \wedge \cdots \wedge \mathbf{e}_{k_{n-r}}$ entfällt). Der Vorzeichenfaktor $(-1)^p$ tritt dabei zunächst als das Signum der Permutation auf, die das Tupel $(i_1, \ldots, i_r, k_1, \ldots, k_{n-r})$

der Größe nach ordnet; Abzählen der Fehlstellen dieser Permutation ergibt die angegebene Formel für p. $\square$

Bemerkung. Die Aussage und der Beweis des Satzes sind unverändert gültig, wenn man den Körper K durch einen beliebigen kommutativen Ring R mit 1 ersetzt.

Lemma 16.19. *Sei R ein kommutativer Ring mit 1, sei $A \in M(m \times n, R$, seien $S \in M_m(R), T \in M_n(R)$. Dann sind für $1 \leq k \leq \min(m,n)$ die $k \times k$-Unterdeterminanten von SAT Linearkombinationen mit Koeffizienten in R der $k \times k$-Unterdeterminanten von A.*

Beweis. Es reicht, die Behauptung für AT zu zeigen, sie folgt dann für SA durch Transponieren und für SAT durch Zusammensetzen beider Schritte.
In AT sind die Spalten $\mathbf{s}'_j$ Linearkombinationen der Spalten $\mathbf{s}_j$ von A. Für $1 \leq j_1 < \cdots < j_r \leq n$ ist daher $\mathbf{s}'_{j_1} \wedge \cdots \wedge \mathbf{s}'_{j_r}$ eine Linearkombination der $\mathbf{s}_{k_1} \wedge \cdots \wedge \mathbf{s}_{k_r}$ mit beliebigen $1 \leq k_1 < \cdots < k_r \leq n$. Aus Teil b) des vorigen Satzes folgt die Behauptung. $\square$

Satz 16.20. *Seien V, W Moduln über R und $f : V \to W$ eine lineare Abbildung, sei $k \in \mathbb{N}$. Dann gibt es genau eine lineare Abbildung $f^{\wedge k} : \bigwedge^k V \to \bigwedge^k W$ mit $f(v_1 \wedge \cdots \wedge v_k) = f(v_1) \wedge \cdots \wedge f(v_k)$ für alle $v_1, \ldots, v_k \in V$.*
Insbesondere gilt: Ist $R = K$ ein Körper und $\dim(V) = n < \infty$, so ist $f^{\wedge n}$ Multiplikation mit dem Skalar $\det(f)$.

Beweis. Die Abbildung $(v_1, \ldots, v_k) \mapsto f(v_1) \wedge \cdots \wedge f(v_k)$ von V^k in $\bigwedge^k V$ ist alternierend und k-fach multilinear; $f^{\wedge k}$ ist die hiervon induzierte lineare Abbildung $\bigwedge^k V \to \bigwedge^k V$. $\square$

Bemerkung. Der Satz kann auch benutzt werden, um eine von vornherein invariante und von Basiswahlen unabhängige Definition von $\det(f)$ zu geben.

Satz 16.21. *Sei K ein Körper, $A \in M_n(K)$ und $\chi_A = X^n + \sum_{j=0}^{n-1} c_j X^j$ das charakteristische Polynom von A.*

a) *Es gilt $c_{n-j} = (-1)^j \mathrm{tr}(L_A^{\wedge j})$ für $1 \leq j \leq n$, wobei die Spur $\mathrm{tr}(g) = \mathrm{Spur}(g)$ eines Endomorphismus g als die Summe der Diagonalkoeffizienten der Matrix von g bezüglich einer beliebigen Basis des zu Grunde liegenden Vektorraums definiert ist.*
b) *Für $B \in M_n(K)$ ist $\chi_{AB} = \chi_{BA}$.*

Beweis. Mit Hilfe von Teil b) von Satz 16.18 sieht man, dass $\mathrm{tr}(L_A^{\wedge j})$ die Summe der $j \times j$- Hauptminoren von A ist, d.h., derjenigen $j \times j$-Unterdeterminanten, bei denen man die gleichen Zeilen- und Spaltenindizes ausgewählt hat. Andererseits ist klar, dass der mit $(-1)^j$ multiplizierte Koeffizient c_{n-j} ebenfalls gleich dieser Summe ist. Teil b) zeige man als Übung. $\square$

Bemerkung. a) AB ist i.a. nicht ähnlich zu BA, Beispiele (etwa für 2×2-Matrizen) überlege man sich als Übung.

b) Ebenfalls mit Teil b) von Satz 16.18 kann man jetzt für $1 \leq j_1 < \cdots < j_r \leq n$ den verallgemeinerten Laplace'schen Entwicklungssatz zeigen:

$$\det(A) = \sum_{1 \leq i_1 < \cdots < i_r \leq n} (-1)^{i_1 + \cdots + i_r + j_1 + \cdots + j_r} A^{i_1, \ldots, i_r}_{j_1, \ldots, j_r} A^{i'_1, \ldots, i'_{n-r}}_{j'_1, \ldots, j'_{n-r}},$$

wo

$$\{i_1, \ldots, i_r, i'_1, \ldots, i'_{n-r}\} = \{j_1, \ldots, j_r, j'_1, \ldots, j'_{n-r}\} = \{1, \ldots, n\}$$

mit $1 \leq i'_1 < \cdots < i'_{n-r} \leq n, 1 \leq j'_1 < \cdots < j'_{n-r} \leq n$ ist.

Definition und Lemma 16.22. *Sei V ein Modul über dem Ring R, seien $k_1, k_2 \in \mathbb{N}$ und $k = k_1 + k_2$.*
Dann gibt es genau eine bilineare Abbildung

$$\beta : \bigwedge^{k_1} V \times \bigwedge^{k_2} V \longrightarrow \bigwedge^{k} V,$$

für die

$$\beta((v_1 \wedge \cdots \wedge v_{k_1}), (v_{k_1+1} \wedge \cdots \wedge v_k)) = v_1 \wedge \cdots \wedge v_k$$

für alle $v_1, \ldots, v_k \in V$ gilt.
Diese wird mit $w_1 \in \bigwedge^{k_1} V =: W_1, w_2 \in \bigwedge^{k_2} V =: W_2$ auch als

$$\begin{aligned} (w_1, w_2) &\longmapsto w_1 \wedge w_2 := \beta(w_1, w_2) \\ W_1 \times W_2 &\longrightarrow \bigwedge^{k} V \end{aligned}$$

geschrieben.

Beweis. Die Abbildung

$$\beta_1 : ((v_1, \ldots, v_{k_1}), (v_{k_1+1}, \ldots, v_k)) \longmapsto v_1 \wedge \cdots \wedge v_k \in \bigwedge^{k} V$$

von $V^{k_1} \times V^{k_2}$ in $\bigwedge^k V$ ist offenbar bilinear. Da sie sowohl als Funktion der ersten k_1 Einträge $v_1, \ldots, v_{k_1}$ als auch als Funktion der folgenden k_2 Einträge $v_{k_1+1}, \ldots, v_k$ eine alternierende k_1- bzw. k_2-fache Multilinearform ist, induziert sie für jedes Tupel $(v_{k_1+1}, \ldots, v_k)$ eine lineare Abbildung $\beta_{(v_{k_1+1}, \ldots, v_k)} : \bigwedge^{k_1} V \to \bigwedge^k V$ mit $\beta_{(v_{k_1+1}, \ldots, v_k)}(v_1 \wedge \cdots \wedge v_{k_1}) = v_1 \wedge \cdots \wedge v_k$ für alle $(v_1, \ldots, v_{k_1}) \in V^{k_1}$), wobei die Abbildung $(v_{k_1+1}, \ldots, v_k) \mapsto \beta_{(v_{k_1+1}, \ldots, v_k)}$ alternierend k_2-fach linear ist. Sie induziert also eine lineare Abbildung $\tilde{\beta} : \bigwedge^{k_2} V \to \mathrm{Hom}(\bigwedge^{k_1} V, \bigwedge^k V)$ mit $\tilde{\beta}(v_{k_1+1} \wedge \cdots \wedge v_k) = \beta_{(v_{k_1+1}, \ldots, v_k)}$ für alle $(v_{k_1+1}, \ldots, v_k) \in V^{k_2}$.
Mit $\beta(w_1, w_2) := (\tilde{\beta}(w_2))(w_1)$ haben wir dann die gewünschte bilineare Abbildung konstruiert. $\square$

Definition und Satz 16.23. *Sei V ein K-Vektorraum der (endlichen) Dimension n.*
Der 2^n-dimensionale K-Vektorraum

$$\bigwedge V := \bigoplus_{k=0}^{n} \bigwedge^{k} V$$

(mit $\bigwedge^0 V := K$) wird durch die Verknüpfung

$$(v_0 + \cdots + v_n) \wedge (w_0 + \cdots + w_n) := \sum_{k=0}^{n} \sum_{k_1+k_2=k} v_{k_1} \wedge w_{k_2}$$

zu einer assoziativen (nicht kommutativen) K-Algebra. Diese heißt die äußere Algebra *oder* Grassmann-Algebra*; sie ist eine* graduierte Algebra*, d.h., man hat $\bigwedge^{k_1} V \wedge \bigwedge^{k_2} V \subseteq \bigwedge^{k_1+k_2} V$ (wenn man formal $\bigwedge^k V = \{\mathbf{0}\}$ für $k > n$ setzt).*

Beweis. Die Assoziativität des $\wedge$-Produkts rechnet man leicht nach, das Distributivgesetz ist äquivalent zur bereits gezeigten Bilinearität. □

Bemerkung. a) Die äußere Algebra kann genauso für einen Modul V über einem beliebigen kommutativen Ring R definiert werden, wenn man $\bigwedge V := \bigoplus_{k=0}^{\infty} \bigwedge^k V$ statt $\bigwedge V := \bigoplus_{k=0}^{n} \bigwedge^k V$ schreibt.
b) In analoger Weise kann man auch die symmetrische Algebra $\mathrm{Sym}(V)$ konstruieren.

17. Affine und projektive Geometrie

Definition 17.1. *Sei K ein Körper, V ein K-Vektorraum. Eine Menge X mit einer Abbildung*

$$\begin{array}{rccl} \tau: & X \times X & \longrightarrow & V \\ & (P,Q) & \longmapsto & \overrightarrow{PQ} = \tau(P,Q) = v \in V \end{array}$$

heißt affiner Raum über K mit Translationsraum $V = T(X)$, wenn gilt

a) *Zu $P \in X$, $v \in V$ gibt es genau ein $Q \in X$ mit $v = \tau(P,Q) = \overrightarrow{PQ}$.*
b) *Sind $P, Q, R \in X$, so ist $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$.*

Die Dimension von X ist die Dimension des Translationsraums V. Ist $\dim X = 1$, so heißt X eine Gerade.

Eine Teilmenge $Y \subseteq X$ heißt affiner Unterraum von X, wenn es einen Untervektorraum U von V gibt, so dass $U = \tau(Y \times Y)$ gilt und a) von oben erfüllt ist.
Ist $\dim U = 1$, so heißt Y eine (affine) Gerade in X.

Beispiel:

a) $V = X$, $\tau(P,Q) = Q - P$.
b) Sei $X \subseteq \mathbb{R}^3$ eine Ebene, die nicht durch den Ursprung geht und U die Ebene durch den Ursprung parallel zu X. Für jedes $\mathbf{x}_0 \in X$ ist also $X = \mathbf{x}_0 + U$ die Nebenklasse von $\mathbf{x}_0$ im Faktorraum $\mathbb{R}^3/U$; ist $(\mathbf{u}_1, \mathbf{u}_2)$ eine Basis von U, so hat X die Parameterdarstellung $X = \{\mathbf{x}_0 + \lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 \mid \lambda_1, \lambda_2 \in \mathbb{R}\}$.

   Für $P = \mathbf{x}_0 + \mathbf{u}, Q = \mathbf{x}_0 + \mathbf{u}' \in X$ mit $\mathbf{u}, \mathbf{u}' \in U$ setzen wir $\overrightarrow{PQ} := \tau(P,Q) = \mathbf{u}' - \mathbf{u} = Q - P \in U$, damit ist (X, V, τ) ein affiner Raum über $\mathbb{R}$ mit Translationsraum U.

   Die Summe von zwei Elementen von X können wir wegen $X \subseteq \mathbb{R}^3$ zwar in $\mathbb{R}^3$ bilden, sie ist aber weder in X noch im Translationsraum U. Auch die Differenz von zwei Elementen P, Q von X können wir wegen $X \subseteq \mathbb{R}^3$ in $\mathbb{R}^3$ bilden, sie ist gleich $\overrightarrow{PQ} := \tau(P,Q)$ und liegt in U.

   Analog könnnen wir allgemeiner für jeden Unterraum V eines K-Vektorraums W jede Nebenklasse $w + V$ von V im Faktorraum W/V als affinen Raum über K mit Translationsraum V auffassen.

   Der Begriff des affinen Raums über K liefert also einen Formalismus, um mit Hilfe der linearen Algebra im Translationsraum V analytische Geometrie in der Parallelverschiebung von V um einen nicht zu V gehörigen Vektor von W zu treiben.

Lemma 17.2. *Ist Y affiner Unterraum von X, so ist $(Y, \tau|_{Y \times Y})$ ein affiner Raum mit Translationsraum U.*

Beweis. Klar. □

Definition und Lemma 17.3. *Sei X ein affiner Raum, seien $P, Q \in X$, $P \neq Q$. Die Gerade PQ durch P und Q ist*

$$\{Q' \in X \mid \overrightarrow{PQ'} = \lambda \overrightarrow{PQ} \text{ für ein } \lambda \in K\}.$$

Sie ist eine Gerade im Sinne von Definition 17.1

Beweis. Klar. □

Lemma 17.4. *Sei X affiner Raum über K mit Translationsraum V, seien $P \in X$ und g eine Gerade in X. Dann gibt es genau eine Gerade g' in X mit $P \in g'$, die den gleichen Translationsraum hat wie g (parallel zu g ist).*

Beweis. Übung. □

Definition 17.5. *Affine Teilräume Y_1, Y_2 des affinen Raums X heißen parallel, wenn*

$$T(Y_1) \subseteq T(Y_2) \text{ oder } T(Y_2) \subseteq T(Y_1)$$

gilt.

Bemerkung. Parallelität ist keine Äquivalenzrelation. Eine Äquivalenzrelation erhält man, wenn man sich auf Parallelität affiner Teilräume einer festen Dimension k beschränkt.

Bemerkung. Allgemeiner wird eine affine Geometrie durch folgende Liste von Axiomen definiert:
Eine affine Geometrie ist ein 4-Tupel $(\mathcal{P}, \mathcal{G}, I, p)$ aus Mengen $\mathcal{P}$ (von Punkten), $\mathcal{G}$ (von Geraden), einer Relation $I \subseteq \mathcal{P} \times \mathcal{G}$ (ist $(P, g) \in I$, so sagt man, P liegt auf g oder g geht durch P), und einer Relation $P \subseteq \mathcal{G} \times \mathcal{G}$ (ist $(g, g') \in p$, so heißen g und g' parallele Geraden), für das gilt:

(AG 1) Zu $P, Q \in \mathcal{P}$, $P \neq Q$ gibt es genau eine Gerade $g = PQ$ durch P und Q.
(AG 2) Auf jeder Geraden liegen mindestens 2 Punkte.
(AG 3) p ist eine Äquivalenzrelation.
(AG 4) Sind $P \in \mathcal{P}$, $g \in \mathcal{G}$, so gibt es genau eine Gerade $g' \in \mathcal{G}$ durch P, die zu g parallel ist.
(AG 5) Sind P, Q, R drei (verschiedene) Punkte, die nicht auf einer Geraden liegen, und $P' \neq Q'$ Punkte, für die $P'Q'$ parallel zu PQ ist, so gibt es $R' \in \mathcal{P}$, für das $Q'R'$ parallel zu QR und $P'R'$ parallel zu PR ist.

Als Übung zeige man, dass durch Definition 17.1 und 17.5 eine affine Geometrie gegeben wird.

Bemerkung. a) Ist X ein affiner Raum mit Translationsraum V so kann man wegen 17.1 a) eine Abbildung

$$T : V \times X \longrightarrow X$$

durch $T(v,R) = Q$ mit $v = \overrightarrow{PQ}$ definieren. Man schreibt auch $T(v,P) = P + v$.
Wegen b) gilt:

$$P + \mathbf{0} = P \text{ für alle } P \in X$$
$$(P + v_1) + v_2 = P + (v_1 + v_2) \quad (v_1, v_2 \in V,\ P \in X).$$

Die Abbildung T definiert also eine Operation der Gruppe $(V,+)$ auf der Menge X (siehe Definition 7.4). Man nennt $P + v$ auch die Translation von P um v.
Ist Y ein affiner Unterraum von X mit Translationsraum U und $P \in Y$, so ist

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

b) Wählt man einen Ursprung $o \in X$ aus, so werden durch

$$\begin{array}{lcl} V \ni v & \longmapsto & o + v \in X \quad \text{und} \\ X \ni P & \longmapsto & \overrightarrow{oP} \in V \end{array}$$

zueinander inverse Bijektionen definiert. Unter dieser Bijektion entspricht dem affinen Unterraum $Y = P + U$ die Nebenklasse $\overrightarrow{oP} + U \subseteq V$ des Translationsraums U. Fasst man V wie im Beispiel nach Definition 17.1 als affinen Raum auf, so sind also die Elemente des Faktorraums V/U genau die affinen Unterräume mit Translationsraum U.

Satz 17.6. *Ist $(Y_i)_{i\in I}$ eine Familie affiner Teilräume des affinen Raums X, so ist $\bigcap_{i\in I} Y_i$ ein affiner Teilraum (dabei wird $\emptyset$ als affiner Raum der Dimension -1 angesehen).*

Beweis. Klar. □

Definition 17.7. *Für jede Teilmenge $A \subseteq X$ ist*

$$\mathrm{Aff}(A) := \bigcap_{\substack{Y \supseteq A \\ Y \text{ Teilraum}}} Y$$

der von A erzeugte Teilraum von X (die affine Hülle von A).

Sind speziell Y_1, Y_2 Teilräume des affinen Raums X, so heißt

$$Y_1 \vee Y_2 := \mathrm{Aff}(Y_1 \cup Y_2)$$

die Verbindung (join) von Y_1 und Y_2.

Beispiel: Für Punkte $P \neq Q$ von X ist die Gerade PQ die affine Hülle von $\{P, Q\}$ und die Verbindung $\{P\} \vee \{Q\}$.

Lemma 17.8. *Sind $Y_1 = P_1 + U_1$, $Y_2 = P_2 + U_2$ affine Unterräume von X und $U := K \cdot \overrightarrow{P_1P_2} + U_1 + U_2$, so ist $Y_1 \vee Y_2 = P_1 + U = P_2 + U$.*

Beweis. Übung. □

Bemerkung. Die Bildung der affinen Hülle ist ein sogenannter Hüllenoperator auf der Potenzmenge $\mathfrak{P}(X)$, d.h., es gilt:

a) $A \subseteq \mathrm{Aff}(A)$
b) $A \subseteq B \Rightarrow \mathrm{Aff}(A) \subseteq \mathrm{Aff}(B)$
c) $\mathrm{Aff}(\mathrm{Aff}(A)) = \mathrm{Aff}(A)$.

Definition und Lemma 17.9. *Seien $Q, P_0, \ldots, P_n$ Punkte des affinen Raums $X, \lambda_0, \ldots, \lambda_n \in K$ mit $\lambda_0 + \cdots + \lambda_n = 1$. Dann hängt*

$$\lambda_0 P_0 + \cdots + \lambda_n P_n := Q + \sum_{i=0}^{n} \lambda_i \overrightarrow{QP_i}$$

nicht von Q ab und heißt eine Affinkombination (affine Linearkombination) der Punkte $P_0, \ldots, P_n$.

Satz 17.10. *Sei $A \subseteq X$. Dann gilt*

$$\mathrm{Aff}(A) = \{\sum_{i=0}^{n} \lambda_i a_i \mid n \in \mathbb{N},\, a_0, \ldots, a_n \in A,\, \sum_{i=0}^{n} \lambda_i = 1\}.$$

Insbesondere ist $Y \subseteq X$ genau dann ein affiner Teilraum, wenn Y unter Bildung von Affinkombinationen abgeschlossen ist.

Beweis. Übung. □

Definition und Satz 17.11. *Sei X affiner Raum. Punkte $P_0, \ldots, P_n \in X$ heißen affin unabhängig, wenn die folgenden äquivalenten Bedingungen erfüllt sind:*

a) *$P_i \notin \mathrm{Aff}(\{P_0, \ldots, P_n\} \setminus \{P_i\})$ für $0 \leq i \leq n$*
b) *$P_0P_1, \ldots, \overrightarrow{P_0P_n}$ sind linear unabhängig.*
c) *Sind $\lambda_0, \ldots, \lambda_n,\ \mu_0, \ldots, \mu_n \in K$ mit $\sum_{i=0}^{n} \lambda_i = 1 = \sum_{i=0}^{n} \mu_i$ so, dass $\sum_{i=0}^{n} \lambda_i P_i = \sum_{i=0}^{n} \mu_i P_i$ gilt, so ist $\lambda_i = \mu_i$ für alle i.*

Eine Teilmenge $A \in X$ heißt affin unbhängig, wenn jede endliche Teilmenge affin unabhängig ist.

Beweis. Übung. □

Definition 17.12. *Eine Teilmenge $A \subseteq X$ des affinen Raums X heißt Erzeugendensystem von X, wenn $X = \mathrm{Aff}(A)$ gilt.*

A heißt affine Basis von X, wenn A affin unabhängig und Erzeugendensystem von X ist. Ist $A = \{P_0, \ldots, P_n\}$ affine Basis von X (mit verschiedenen P_i), so heißen für $Q \in X$ die eindeutig bestimmten $\lambda_0, \ldots, \lambda_n \in K$ mit $\sum_{i=0}^{n} \lambda_i P_i = Q$ die baryzentrischen Koordinaten von Q bezüglich der (geordneten) Basis $(P_0, \ldots, P_n)$.

Satz 17.13. *(Dimensionsformel)*
Y_1, Y_2 seien Teilräume des affinen Raums X. Dann gilt:

$$\dim Y_1 + \dim Y_2 = \begin{cases} \dim(Y_1 \vee Y_2) & + \quad \dim(Y_1 \cap Y_2) \\ & \quad \textit{falls } Y_1 \cap Y_2 \neq \emptyset \\ \dim(Y_1 \vee Y_2) & + \quad \dim(T(Y_1) \cap T(Y_2)) - 1 \\ & \quad \textit{falls } Y_1 \cap Y_2 = \emptyset. \end{cases}$$

Beweis. Übung □

Beispiel: Sei $\dim X = 3$, Y_1 und Y_2 seien verschiedene Geraden in X.

Ist $Y_1 \cap Y_2 \neq \emptyset$, so hat $Y_1 \cap Y_2$ Dimension 0, und $Y_1 \vee Y_2$ ist die von Y_1 und Y_2 aufgespannte Ebene.

Ist $Y_1 \cap Y_2 = \emptyset$ und Y_1 und Y_2 parallel zueinander, so ist $\dim(T(Y_1) \cap T(Y_2)) = 1$, also wiederum $\dim(Y_1 \vee Y_2) = 2$, $Y_1 \vee Y_2$ ist eine Ebene, in der beide Geraden liegen.

Ist $Y_1 \cap Y_2 = \emptyset$ und Y_1, Y_2 nicht parallel, so heißen die Geraden windschief. In diesem Fall ist $Y_1 \vee Y_2$ der ganze Raum X!

Bemerkung. Seien $P_0, P_1 \in X$, $Q \in P_0P_1$, $Q = \lambda_0 P_0 + \lambda_1 P_1$ mit $\lambda_0 + \lambda_1 = 1$. Man nennt $TV(P_0, P_1; Q) := \lambda_1$ das Teilverhältnis von Q bezüglich der Basis (P_0, P_1) der Geraden P_0P_1; man hat also $TV(P_0, P_1; P_1) = 1$. Ist etwa $K = \mathbb{R}$ und Q zwischen P_0 und P_1 ($\lambda_0, \lambda_1 \geq 0$), so ist $TV(P_0, P_1; Q) = \frac{d(P_0,Q)}{d(P_0,P_1)}$ (wenn d eine Metrik auf der Geraden bezeichnet).

Definition 17.14. *Sei $K = \mathbb{R}$.*

a) *Zu $P, Q \in X$ mit $v = \overrightarrow{PQ}$ ist*

$$\begin{aligned} \overline{PQ} &:= \{P + \lambda v \mid 0 \leq \lambda \leq 1\} \\ &= \{R \in X \mid \overrightarrow{PR} = \lambda \overrightarrow{PQ} \textit{ mit } 0 \leq \lambda \leq 1\} \end{aligned}$$

die Strecke von P nach Q.

b) *Eine Menge $K \subseteq X$ heißt konvex, wenn $\overline{PQ} \subseteq K$ für alle $P, Q \in K$ gilt.*

c) *Ist $A \subseteq X$, so ist die konvexe Hülle $C(A)$ definiert als Durchschnitt aller konvexer Mengen $K \supseteq A$.*

Satz 17.15. a) *Die konvexe Hülle $C(A)$ ist konvex.*

b) *Ist $A = \{P_j \mid j \in J\}$ so ist*

$$C(A) = \sum_{j \in J} \lambda_j P_j \textit{ mit } \lambda_j \in \mathbb{R}, \lambda_j \geq 0, \sum \lambda_j = 1\}.$$

Beweis. a) ist klar. Für b) bezeichnen wir die rechte Seite der Gleichung als C'. Man rechnet nach: Sind $\sigma, \tau \geq 0, \sigma + \tau = 1, \sum \lambda_j P_j, \sum \kappa_j P_j \in C'$ mit $\lambda_j, \kappa_j \geq 0$, $\sum \lambda_j = \sum \kappa_j = 1$, so ist

$$\sigma \sum \lambda_j P_j + \tau \sum \kappa_j P_j = \sum (\sigma \lambda_j + \tau \kappa_j) P_j$$

mit $\sigma\lambda_j \geq 0 \leq \tau\kappa_j$ und

$$\sum_j \sigma\lambda_j + \sum_j \tau\kappa_j = 1.$$

Also ist C' konvex.

Noch zu zeigen ist: Ist K konvex mit $P_j \in K$ für alle j, so ist $C' \subseteq K$. Dafür zeigen wir durch Induktion nach $r = \#\{j \in J \mid \lambda_j \neq 0\}$, dass gilt: Ist K konvex mit $P_j \in K$ für alle j, so ist $\sum \lambda_j P_j \in K$ für alle $(\lambda_j)_{j\in J}$ mit $\lambda_j \geq 0, \sum \lambda_j = 1$.
Klar für $r = 1$.
Sei $r > 1$, Behauptung gezeigt für $r' < r$, o.E.: $\lambda_j > 0$ für $1 \leq j \leq r$, $\lambda_j = 0$ für $j \notin \{1, \ldots, r\}$.
Dann:

$$\sum_{j=1}^{r} \lambda_j P_j = \lambda_1 P_1 + (1-\lambda_1)(\sum_{j=2}^{r} \lambda_j (1-\lambda_1)^{-1} P_j)$$

mit $\sum_{j=2}^{r} \lambda_j = 1 - \lambda_1$, also $\sum_{j=2}^{r} \lambda_j (1-\lambda_1)^{-1} = 1$.

Nach Induktionsannahme liegen P_1 und $Q := \sum_{j=2}^{r} \lambda_j (1-\lambda_1)^{-1} P_1$ in K, also auch $\lambda_1 P_1 + (1-\lambda_1)Q$ weil K konvex ist. □

Definition 17.16. *X_1, X_2 seien affine Räume über K mit Translationsräumen V_1, V_2. $\varphi : X_1 \longrightarrow X_2$ heißt affine Abbildung, wenn es $f \in \mathrm{Hom}(V_1, V_2)$ gibt mit*

$$\overrightarrow{\varphi(P)\varphi(Q)} = f(\overrightarrow{PQ}) \text{ für alle } P, Q \in X_1.$$

f heißt dann die Ableitung von φ.

Ist φ bijektiv, so heißt φ ein affiner Isomorphismus, ist zusätzlich $X_1 = X_2$, so heißt φ affiner Automorphismus.

Bemerkung. Äquivalent ist

$$\begin{array}{ccc} V_1 \times X_1 & \xrightarrow{+} & X_1 \\ (f,\varphi)\downarrow & & \downarrow\varphi \\ V_2 \times X_2 & \xrightarrow{+} & X_2 \end{array}$$

kommutiert ($f(v) + \varphi(p) = \varphi(v + P)$).

Definition und Satz 17.17. *Eine Translation von X ist eine Abbildung $T = T_v : X \longrightarrow X$ mit $T_v(P) = P + v$ für alle $P \in X$.*
Es gilt:

a) *Translationen sind affin mit Ableitung* Id_V
b) *Die Translationen bilden eine Untergruppe T der Gruppe* $\mathrm{Aut}\,\mathrm{Aff}(X)$ *der affinen Automorphismen von X..*
c) *Durch $v \longmapsto \tau_v$ wird ein Isomorphismus $V \longrightarrow T$ gegeben.*

Beweis. Übung □

Satz 17.18. *Sei $P_0 \in X_1$ fixiert. Für $f \in \mathrm{Hom}(V_1, V_2)$ und $Q \in X_2$ ist $\varphi_{Q,f}$ gegeben durch*

$$\varphi_{Q,f}(P) = f(\overrightarrow{P_0P}) + Q$$

eine affine Abbildung mit Ableitung f, und jede affine Abbildung $X_1 \longrightarrow X_2$ lässt sich eindeutig auf diese Weise darstellen.

Beweis. Man hat für $P, P' \in X_1$

$$\begin{aligned} \varphi_{Q,f}(P') &= f(\overrightarrow{P_0P'}) + Q \\ &= f(\overrightarrow{P_0P}) + f(\overrightarrow{PP'}) + Q, \end{aligned}$$

also

$$\overrightarrow{\varphi_{Q,f}(P)\varphi_{Q,f}(P')} = f(\overrightarrow{PP'}).$$

Umgekehrt sei φ affin mit Ableitung f, $Q := \varphi(P_0)$. Dann ist für $P \in X_1$:

$$\overrightarrow{\varphi(P_0)\varphi(P)} = \overrightarrow{f(P_0P)},$$

also

$$\begin{aligned} \varphi(P) &= \overrightarrow{\varphi(P_0)} + f(\overrightarrow{P_0P}) \\ &= Q + f(\overrightarrow{P_0P}), \end{aligned}$$

□

Bemerkung. Sei $X_1 = V_1$ mit $P_{0,1} = \mathbf{0}$, $X_2 = V_2$ mit $P_{0,2} = \mathbf{0}$, $X_3 = V_3$, seien $f_1 : V_1 \longrightarrow V_2$, $f_2 : V_2 \longrightarrow V_3$ linear, $Q_2 \in X_2$, $Q_3 \in X_3$. Dann ist $\varphi_{Q_2,f_1}(P) = f_1(P) + Q_2$,

$$\begin{aligned} \varphi_{Q_3,f_2}(\varphi_{Q_2,f_1}(P)) &= Q_3 + f_2(f_1(P)) + f_2(Q_2)) \\ &= Q_3 + f_2(Q_2) + (f_2 \circ f_1)(P), \end{aligned}$$

also

$$\varphi_{Q_3,f_2} \circ \varphi_{Q_2,f_1} = \varphi_{Q_3+f_2(Q_2),f_2\circ f_1}.$$

Speziell in der Gruppe $\mathrm{AGL}(V)$ der affinen Automorphismen von V gilt also mit $\varphi_{Q,f} =: (Q, f)$

$$(Q_2, f_2) \circ (Q_1, f_1) = (Q_1 + f_2(Q_2), f_1 \circ f_2).$$

Man sagt, $\mathrm{AGL}(V)$ sei das semidirekte Produkt $V \rtimes \mathrm{GL}(V)$ der (additiven) Gruppe V und der Gruppe $\mathrm{GL}(V)$ der linearen Automorphismen von V.

Satz 17.19. *Seien $X_1, X_2, X_3, V_1, V_2, V_3$ wie oben.*

a) *Mit affinen $\varphi_1 : X_1 \longrightarrow X_2$, $\varphi_2 : X_2 \longrightarrow X_3$ mit Ableitungen f_1, f_2 ist $\varphi_2 \circ \varphi_1$ affin mit Ableitung $f_2 \circ f_1$.*
b) *φ_1 ist genau dann injektiv (surjektiv), wenn f_1 diese Eigenschaft hat.*
c) *Ist φ_1 bijektiv, so ist φ_1^{-1} affin mit Ableitung f_1^{-1}.*

Beweis. a) folgt aus der Definition.
b), c) sieht man mit Hilfe des vorigen Satzes. □

Satz 17.20. *Sei $(P_0, \ldots, P_n)$ eine affine Basis von X_1, $Q_0, \ldots, Q_n \in X_2$. Dann gibt es genau eine affine Abbildung*

$$\varphi : X_1 \longrightarrow X_2 \text{ mit } \varphi(P_i) = Q_i \quad (0 \leq i \leq n),$$

und für diese gilt

$$\varphi(\sum_{i=0}^{n} \lambda_i P_i) = \sum_{i=0}^{n} \lambda_i Q_i \quad (\lambda_0 + \cdots + \lambda_n = 1)$$

Beweis. $\overrightarrow{P_0P_n}, \ldots, \overrightarrow{P_0P_1}$ sind eine Basis von V_1, also gibt es genau eine lineare Abbildung

$$f : V_1 \longrightarrow V_2 \text{ mit } f(\overrightarrow{P_0P_i}) = \overrightarrow{Q_0Q_i} \quad (1 \leq i \leq n).$$

Dann ist

$$\varphi_{Q_0,f}(P_i) = Q_0 + f(\overrightarrow{P_0P_0i} = Q_i,$$

und für

$$\begin{aligned} \varphi_{Q_0,f}(\textstyle\sum_{i=0}^{n} \lambda_i P_i) &= Q_0 + f(\overrightarrow{P_0(\sum \lambda_i P_i)}) \\ &= Q_0 + \textstyle\sum_i \lambda_i f(\overrightarrow{P_0P_i}) \\ &= \textstyle\sum_{i=0}^{n} \lambda_i Q_0 + \sum_{i=0}^{n} \lambda_i \overrightarrow{Q_0Q_i} \\ &= \textstyle\sum_{i=0}^{n} \lambda_i Q_i. \end{aligned}$$

□

Satz 17.21. *Unter den Voraussetzungen des vorigen Satzes gilt:*

a) *φ ist genau dann injektiv, wenn $(Q_0, \ldots, Q_n)$ linear unabhängig ist.*
b) *φ ist genau dann surjektiv, wenn $Q_0, \ldots, Q_n$ Erzeugendensystem von X_2 ist.*
c) *φ ist genau dann bijektiv, wenn $Q_0, \ldots, Q_n$ affine Basis von X_2 ist.*

Beweis. Übung □

Satz 17.22. *Ist $\varphi : X_1 \longrightarrow X_2$ affine Abbildung, so sind Bilder und Urbilder unter φ von affinen Teilräumen wieder affine Teilräume.*

Beweis. Übung. □

Beispiel: Sei X ein affiner Raum über K mit Translationsraum V, seien Y_1, Y_2 affine Unterräume mit Translationsräumen U_1, U_2. Sei $V = W \oplus U_1 = W \oplus U_2$. Für $P \in X$ sei

$$W(P) := \{Q \in X \mid \overrightarrow{PQ} \in W\} = P + W,$$

$W(P)$ ist die Nebenklasse $P + W$ und die Bahn von P unter der Operation von W auf X durch Translationen.

Sind $Q, Q' \in W(P) \cap Y_1$, so ist $Q = P + w$, $Q' = P + w'$, $Q' \in (P + W) \cap (Q + U_1)$, also $Q' = P + w' = P + w + u$ $(u \in U_1)$, also

$w' - w = u \in U_1 \cap W = \{\mathbf{0}\}$, also $Q = Q'$.

Andererseits ist $W(P) \cap Y_1 \neq \emptyset$, denn ist $Y_1 = R + U_1$ und $\overrightarrow{PR} = w - u$ mit $w \in W$, $u \in U$ (geht, da $U + W = V$), so ist

$$R + u = P + w \in W(P) \cap Y_1.$$

Also: Für jedes $P \in X$ existiert genau ein $y \in X$ mit

$$y \in W(P) \cap Y_1.$$

Definiere eine Abbildung

$$\begin{array}{rccl} \pi_W : & X & \longrightarrow & Y_1 \\ & x & \longmapsto & y \in W(P) \cap Y_1. \end{array}$$

π_W heißt Parallelprojektion längs W (anschaulich: W eine Gerade im dreidimensionalen Raum).

π_W ist affin und surjektiv.

Affin: Sei $p_W : V \longrightarrow U_1$ die Projektion auf U_1 (längs W) mit $w + u \longmapsto u$.
Für$P, P' \in X$ sei $\overrightarrow{PP'} = w_1 + u_1$

$$\pi_W(P) = P + w \in Y_1, \quad \pi_W(P') = P' + w' \in Y_1 = P + w_1 + u_1 + w'.$$

Dann ist $U_1 \ni \overrightarrow{\pi_W(P)\pi_W(P')} = w_1 + u_1 + w' - w$, also $\overrightarrow{\pi_W(P)\pi_W(P)} = u_1 = p_W(\overrightarrow{PP'})$.
Also: π_W ist affin mit Ableitung p_W.

π_W ist surjektiv, weil p_W surjektiv ist.

Ist Y_2 wie oben, so ist $\pi_W|_{Y_2} : Y_2 \longrightarrow Y_1$ bijektiv.

Um neben der Parallelprojektion auch Zentralprojektion zu beschreiben, führen wir projektive Räume ein.

Definition 17.23. *Sei V ein K-Vektorraum. Der projektive Raum $\mathbb{P}(V)$ zu V ist*

$$\begin{aligned} \mathbb{P}(V) &= \{U \subseteq V \mid U \text{ ist Unterraum, } \dim U = 1\} \\ &= \{\langle x \rangle = \mathrm{Lin}(x) \mid \mathbf{0} \neq x \in V\}. \end{aligned}$$

Ist $\dim V = n < \infty$, so ist $\dim \mathbb{P}(V) = n - 1$. (Insbesondere ist $\mathbb{P}(\{\mathbf{0}\}) = \emptyset$ mit Dimension -1).

Ist $V = K^{n+1}$, so schreibt man $\mathbb{P}(V) = \mathbb{P}_n(K) = \mathbb{P}^n(K)$, ist $P = \langle (x_0, \ldots, x_n) \rangle$, so schreibt man $P = (x_0 : \ldots : x_n)$ und nennt $x_0, \ldots, x_n$ homogene Koordinaten von P

Lemma 17.24. *Für $\mathbf{0} \neq (x_0, \ldots, x_n), (y_0, \ldots, y_n) \in K^{n+1}$ ist $(x_0 : \ldots : x_n) = (y_0 : \ldots : y_n)$ genau dann, wenn es $\lambda \in K^\times$ gibt mit*

$$y_i = \lambda x_i \quad (0 \leq i \leq n).$$

Bemerkung. Man schreibt auch $\mathbb{P}^n(K) = K^{n+1}/K^\times$.

Definition 17.25. *Ist $W \subseteq V$ ein Unterraum, so heißt $\{U \in \mathbb{P}(V) \mid U \subseteq W\}$ projektiver Unterraum $\mathbb{P}(W)$ von $\mathbb{P}(V)$.*

$Z \subseteq \mathbb{P}(V)$ heißt projektiver Unterraum von $\mathbb{P}(V)$, wenn $Z = \mathbb{P}(W)$ für einen Untervektorraum W von V ist. Ist $\dim W = 2$, so heißt Z eine (projektive) Gerade in $\mathbb{P}(V)$, ist $\dim W = 3$ eine (projektive) Ebene, ist $\dim W = \dim V - 1$ eine (projektive) Hyperebene.

Definition und Satz 17.26. *Sei V ein K-Vektorraum, $\mathbb{P}(W_i) = Z_i$ $(i \in I)$ projektive Unterräume. Dann ist*

$$\bigcap_{i\in I} Z_i = \mathbb{P}(\bigcap_{i\in I} W_i)$$

ein projektiver Unterraum.

$$\bigvee_{i\in I} Z_i := \mathbb{P}(\sum_{i\in I} W_i)$$

ist der Durchschnitt aller projektiven Unterräume, die $\bigcup_{i\in I} Z_i$ enthalten, und heißt die projektive Hülle der Z_i.

Beweis. Klar. Zur Definition von $\sum_{i\in I} W_i$ im Fall, dass I unendlich ist:

$$\sum_{i\in I} W_i = \{m_{i_1} + \cdots + m_{i_n} \mid n \in \mathbb{N},\ m_{i_k} \in W_{i_k}\ (1 \le k \le n)\}.$$

Als Übung zeige man: $\sum_{i\in I} W_i$ ist die lineare Hülle von $\bigcup_{i\in I} W_i$. □

Satz 17.27. *Ist V endlichdimensional und sind Z_1, Z_2 projektive Unterräume von $\mathbb{P}(V)$, so ist*

$$\dim(Z_1 \vee Z_2) = \dim Z_1 + \dim Z_2 - \dim(Z_1 \cap Z_2).$$

Insbesondere: Ist

$$\dim Z_1 + \dim Z_2 \ge \dim \mathbb{P}(V)$$

so ist

$$Z_1 \cap Z_2 \ne \emptyset.$$

Beweis. Folgt aus der Dimensionsformel für Untervektorräume □

Korollar 17.28. *In $\mathbb{P}^2(K)$ schneiden sich je zwei verschiedene Geraden in genau einem Punkt.*

Bemerkung. In der projektiven Geometrie gilt das Parallenaxiom nicht.

Satz 17.29. *Sei $n \in \mathbb{N}$, $U_0 := \{(1 : x_1 : \ldots : x_n) \in \mathbb{P}^n(K)\}$. Dann wird U_0 durch $\alpha_0 : (1 : x_1 : \ldots : x_n) \longmapsto (x_1, \ldots, x_n)$ bijektiv auf K^n abgebildet.*

Ist Z ein r-dimensionaler Unterraum von $\mathbb{P}^n(K)$ mit $Z \cap U_0 \ne \emptyset$, so ist das Bild von $Z \cap U_0$ ein affiner Unterraum der Dimension r in K^n (und umgekehrt).

Beweis. Die Bijektivität der Abbildung ist klar.
Sei $Z = \mathbb{P}(W)$ mit $\dim W = r+1$, sei $W_0 = \{(x_1, \ldots, x_n) \in K^n \mid (0, x_1, \ldots, x_n) \in W\}$; wegen $Z \cap U_0 \neq \emptyset$ ist $\dim(W_0) = \dim(W) - 1 = r$.
Sei

$$w^{(0)} = (1, x_1, \ldots, x_n) \in W.$$

Für $z = (1 : z_1 : \ldots : z_n) \in U_0$ ist dann

$$\begin{aligned} & z \in Z \cap U_0 = \mathbb{P}(W) \cap U_0 \\ \Leftrightarrow\ & (z_1 - x_1, \ldots, z_n - x_n) \in W_0 \\ \Leftrightarrow\ & \alpha_0(z) \in \alpha_0(\langle w^{(0)} \rangle) + W_0, \end{aligned}$$

wir haben also $\alpha_0(Z \cap U_0) = \alpha_0(w^{(0)}) + W_0$, und das ist ein r dimensionaler affiner Unterraum von K^n. □

Bemerkung. Der Satz gilt natürlich analog für

$$U_i := \{(x_0 : \ldots : x_n) \in \mathbb{P}^n(K) \mid x_i \neq 0\}.$$

Korollar 17.30. a) *$\mathbb{P}^n(K)$ hat eine Überdeckung $\mathbb{P}(K) = U_0 \cup \cdots \cup U_n$ durch $n+1$ affine Räume der Dimension n.*

b) *Bezeichnen wir den als n-dimensionaler affiner Raum über K aufgefassten Vektorraum K^n mit $\mathbb{A}^n = \mathbb{A}^n(K)$ und identifizieren oben U_0 mittels α_0 mit $\mathbb{A}^n$, so haben wir die Zerlegung $\mathbb{P}^n(K) = U_0 \dot\cup \mathbb{P}^{n-1}(K) = \mathbb{A}^n \dot\cup \mathbb{A}^{n-1} \dot\cup \cdots \dot\cup \mathbb{A}^0$.*
Insbesondere entsteht die projektive Gerade aus der affinen Geraden $\mathbb{A}^1$ durch Hinzunahme eines Punktes.

Bemerkung. a) Man sagt, die U_i seien die Karten eines (affinen) Atlas für $\mathbb{P}^n(K)$. Wir haben die Bijektionen

$$\begin{aligned} \alpha_i :\ U_i &\longrightarrow K^n \\ (x_0 : \ldots : x_n) &\longmapsto \left(\tfrac{x_0}{x_i}, \ldots, \tfrac{x_{i-1}}{x_i}, \tfrac{x_{i+1}}{x_i}, \ldots, \tfrac{x_1}{x_i}\right), \end{aligned}$$

für $z = (x_0 : \ldots : x_n) \in U_0 \cap U_1$ etwa ist

$$\begin{aligned} (\alpha_1^{-1} \circ \alpha_0)(z_0) &= \alpha_1^{-1}\left(\tfrac{x_1}{x_0}, \ldots, \tfrac{x_n}{x_0}\right) \\ &= \left(\tfrac{x_1}{x_0} : 1 : \tfrac{x_2}{x_0} : \ldots : \tfrac{x_n}{x_0}\right) \\ &= (x_1 : x_0 : x_2 : \ldots : x_n) \end{aligned}$$

Beim perspektivischen Zeichnen etwa wird der Punkt $(1, x_1, x_2)$ in der Ebene $x_0 = 1$ auf den Punkt $(\frac{1}{x_1}, 1, \frac{x_2}{x_1})$ in der Ebene $x_1 = 1$ abgebildet (falls $x_1 \neq 0$ war).
Die Gerade $\{x_2 = \lambda x_1 + b\}$ geht dabei auf $\{(\frac{1}{x_1}, 1, \lambda + \frac{b}{x_1}\}$.
Für $x_1 \longrightarrow \infty$ streben diese Punkte gegen $(0, 1, \lambda)$: Die parallelen Geraden schneiden sich im Horizontpunkt $(0, 1, \lambda)$.

b) Man sagt auch in der Situation des Korollars:
Mit $W_0 = \{(0, x_1, \ldots, x_n) \in K^n\}$ ist $H_0 = \mathbb{P}(W_0)$ die uneigentliche Hyperebene, die den affinen Raum $\mathbb{P}(V) \setminus H_0 = U_0$ zum

projektiven Raum ergänzt, die Punke von H_0 sind die uneigentlichen Punkte des affinen Raums U_0.

Natürlich ist hier die Auszeichnung einer Koordinatenhyperebene willkürlich: Ist $H_0 = \mathbb{P}(W)$ irgendeine Hyperebene in $\mathbb{P}(V)$, so ist $H := \mathbb{P}(V) \setminus H_0$ ein affiner $\dim \mathbb{P}(V)$-dimensionaler Raum mit Translationsraum W. Es gilt dabei:

(i) Ist $Z = \mathbb{P}(W_1)$ ein projektiver Teilraum von $\mathbb{P}(W)$, ist $Z \cap A$ ein affiner Teilraum von A (möglicherweise leer).

ii) Ist $\emptyset \neq B \subseteq A$ ein affiner Teilraum, so gibt es genau einen projektiven Teilraum $Z_B = \mathbb{P}(W_1)$ von $\mathbb{P}(W)$ mit $B = Z_B \cap A$, dabei ist $\dim Z_B = \dim B$ und $W_1 \cap W$ der Translationsraum von B.

iii) Ist $H \subseteq A$ eine affine Hyperebene, $B \neq \emptyset$ affiner Unterraum, Z_H, Z_B wie oben, so gilt:

$$\Lambda \| H \Leftrightarrow Z_\Lambda \cap Z_H \subseteq H_0$$

(d.h., Z_Λ und Z_H schneiden sich "im Unendlichen")

Wir können also die projektive Geometrie als Erweiterung der affinen Geometrie auffassen, bei der sich parallele affine Teilräume in der hinzugenommenen "uneigentlichen Hyperebene" (im Unendlichen) schneiden.

Definition und Satz 17.31. *Ist $Z = \mathbb{P}(V)$ ein projektiver Raum über K, so heißen $P_0, \ldots, P_r \in Z$ projektiv unabhängig, wenn die folgenden äquivalenten Bedingungen erfüllt sind:*

a) *Ist $P_j = \langle v_j \rangle = \mathrm{Lin}(v_j)$ mit $v_j \in V$, so sind die v_j linear unabhängig.*
b) *Die $\mathrm{Lin}(v_j)$ bilden eine direkte Summe.*
c) *Keines der P_i ist in der projektiven Hülle der anderen enthalten.*
d) $\dim(P_0 \vee \cdots \vee P_r) = r$.

Beweis. Übung. □

Lemma 17.32. *Sei V ein $(n+1)$-dimensionaler K-Vektorraum, $Z = \mathbb{P}(V)$, seien $n+2$ Punkte $P_0, P_1, \ldots, P_n, E \in Z$ gegeben, so dass je $n+1$ dieser Punkte projektiv unabhängig sind. Dann gilt:*

a) *Ist $E = \mathrm{Lin}(e)$, so gibt es eindeutig bestimmte $v_j \in V$ mit $P_j = \mathrm{Lin}(v_j)$ für $0 \leq j \leq n$, so dass $e = \sum_{j=0}^n v_j$ gilt.*
b) *Ist $E = \mathrm{Lin}(e) = \mathrm{Lin}(e')$ und sind v_j, v_j' wie in a) gegeben mit $e = \sum_{j=0}^n v_j, e' = \sum_{j=0}^n v_j'$, so gibt es $\lambda \in K \setminus \{0\} = K^\times$ mit $e' = \lambda e, v_j' = \lambda v_j$ für $0 \leq j \leq n$.*
c) *Ist $P = \mathrm{Lin}(v) = \mathrm{Lin}(v') \in Z$ und sind e, e', v_j, v_j' wie in a), b), so gibt es $c_j, c_j', \lambda \in K$ mit $v = \sum_{j=0}^n c_j v_j, v' = \sum_{j=0}^n c_j' v_j'$ und $c_j' = \lambda c_j$ für $0 \leq j \leq n$. Durch $P \mapsto c(P) := (c_0 : c_1 : \cdots : c_n)$ wird eine bijektive Abbildung $c : Z \to \mathbb{P}^n(K)$ gegeben.*

Beweis. Da $P_0, \ldots, P_n$ projektiv unabhängig sind, ist $V = \bigoplus_{j=0}^n P_j$, also gibt es eindeutig bestimmte $v_j \in P_j$ mit $e = \sum_{j=0}^n v_j$. Da E zusammen mit je n beliebigen der P_i projektiv unabhängig ist, ist hierbei keines der v_j gleich $\mathbf{0}$, wir haben also $P_j = \mathrm{Lin}(v_j)$ wie in a) gefordert. Ist jetzt zusätzlich e' mit $E = \mathrm{Lin}(e) = \mathrm{Lin}(e')$, so ist $e' = \lambda e$ mit $\lambda \in K^\times$, also $e' = \sum_{j=0}^n (\lambda v_j)$, und wir haben b) gezeigt.
c) schließlich folgt direkt aus a) und b) und der Tatsache, dass wegen der projektiven Unabhängigkeit der P_j sowohl die v_j als auch die v'_j eine Basis von V bilden. □

Definition 17.33. *Sei V ein $(n+1)$-dimensionaler K-Vektorraum, $Z = \mathbb{P}(V)$, seien $n+2$ Punkte $P_0, P_1, \ldots, P_n, E \in Z$ wie im vorigen Lemma gegeben. Dann heißt $(P_0, \ldots, P_n; E)$ ein projektives Koordinatensystem für Z mit Grundpunkten $P_0, \ldots, P_n$ und Einheitspunkt E.*
Für $P \in Z$ heißen die Repräsentanten in K^{n+1} des nach c) des Lemmas zugeordneten Punktes $c(P) = (c_0 : c_1 : \cdots : c_n) \in \mathbb{P}^n(K)$ homogene oder projektive Koordinaten von P bezüglich des Koordinatensystems $(P_0, \ldots, P_n; E)$.

Beispiel: die projektiven Koordinaten von $P_0, \ldots, P_n, E$ sind $(1 : 0 : \cdots : 0), \ldots, (0 : \cdots : 0 : 1), (1 : 1 : \cdots : 1)$.

Definition und Satz 17.34. *Seien V, W K-Vektorräume, $Y = \mathbb{P}(V)$, $Z = \mathbb{P}(W)$ die projektiven Räume zu V, W. Eine Abbildung $\Phi : \mathbb{P}(V) \to \mathbb{P}(W)$ heißt eine projektive Abbildung, wenn es eine injektive lineare Abbildung $\varphi : V \to W$ gibt, so dass $\Phi(\mathrm{Lin}(v)) = \mathrm{Lin}(\varphi(v))$ für alle $v \in V \backslash \{\mathbf{0}\}$ gilt; man sagt dann, dass $\Phi : \mathbb{P}(V) \to \mathbb{P}(W)$ von $\varphi : V \to W$ induziert wird.*
Die projektive Abbildung Φ ist genau dann bijektiv, wenn die induzierende Abbildung φ bijektiv ist, in diesem Fall heißt Φ eine Projektivität. Ist speziell $V = W$, so wird die Menge der Projektivitäten von $\mathbb{P}(V)$ in sich mit $\mathrm{PGL}(V)$ bezeichnet und heißt die projektive lineare Gruppe von V.

Satz 17.35. $\mathrm{PGL}(V) \cong \mathrm{GL}(V)/K^\times \cdot \mathrm{Id}$.

Beweis. Die Abbildung $\varphi \mapsto \Phi$ ist offenbar ein surjektiver Gruppenhomomorphismus von $GL(V) = \mathrm{Aut}(V)$ in $PGL(V)$. Dabei ist $\Phi = \mathrm{Id}_Z$ genau dann, wenn $\mathrm{Lin}(\varphi(v)) = \mathrm{Lin}(v)$ für alle $v \in V$ gilt, also genau dann, wenn $\varphi(v) = \lambda_v \cdot v$ mit $\lambda_v \in K$ für alle $v \in V$ gilt.
Gäbe es $v_1, v_2 \in V \setminus \{\mathbf{0}\}$ mit $\lambda_1 := \lambda_{v_1} \neq \lambda_{v_2} =: \lambda_2$, so wären v_1, v_2 als Eigenvektoren von φ zu verschiedenen Eigenwerten linear unabhängig, und mit $\lambda_{12} := \lambda_{v_1+v_2}$ hätten wir

$$\begin{aligned} \lambda_{12}(v_1 + v_2) &= \varphi(v_1 + v_2) \\ &= \varphi(v_1) + \varphi(v_2) \\ &= \lambda_1 v_1 + \lambda_2 v_2. \end{aligned}$$

Wegen der linearen Unabhängigkeit von v_1, v_2 wäre also $\lambda_1 = \lambda_{12}$, $\lambda_2 = \lambda_{12}$, Widerspruch.
Also hat der Gruppenhomomorphismus $\varphi \longmapsto \Phi$ Kern $\{\lambda \cdot \mathrm{Id}_V\}$, und aus dem Homomorphiesatz der Gruppentheorie folgt die Behauptung. $\square$

Definition 17.36. *V, W seien K-Vektorräume und σ ein Automorphismus von K (also $\sigma : K \longrightarrow K$ bijektiv mit $\sigma(a+b) = \sigma(a) + \sigma(b)$, $\sigma(ab) = \sigma(a)\sigma(b)$ für alle $a, b \in K$).*

$\varphi : V \longrightarrow W$ heißt semilinear bzgl. σ, wenn gilt

a) *$\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$ für alle $v_1, v_2 \in V$,*
b) *$\varphi(av) = \sigma(a)\varphi(v)$ für alle $a \in K$, $v \in V$.*

$\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$ heißt semiprojektiv, wenn gilt: Es gibt eine injektive semilineare Abbildung $\varphi : V \longrightarrow W$ mit

$$\Phi(\langle v \rangle) = \langle \varphi(v) \rangle \text{ für alle } v \in V \setminus \{\mathbf{0}\}.$$

Ist Φ zusätzlich bijektiv, so heißt Φ eine Semiprojektivität.

Beispiel: $\varphi : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ mit $\varphi \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} \overline{z}_1 \\ \vdots \\ \overline{z}_n \end{pmatrix}$ ist semilinear bzgl. der komplexen Konjugation.

Lemma 17.37. *Ist $\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$ semiprojektiv, $Z \subseteq \mathbb{P}(V)$ ein projektiver Unterraum, so ist $\Phi(Z) \subseteq \mathbb{P}(W)$ ein projektiver Unterraum der Dimension* $\dim Z$.

Beweis. Analog zum projektiven Fall (Übung)! $\square$

Definition und Satz 17.38. *Eine Kollineation ist eine Abbildung $\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$, die Geraden in Geraden überführt.*

Es gilt $\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$ ist genau dann Kollineation, wenn Φ semiprojektiv ist.

Beweis. Siehe Fischer, Analytische Geometrie. $\square$

Definition 17.39. *Sei V ein K-Vektorraum, $\mathcal{P}$ die Menge aller projektiven Teilräume von $\mathbb{P}(V)$. $\kappa : \mathcal{P} \longrightarrow \mathcal{P}$ bijektiv heißt eine Korrelation in $\mathbb{P}(V)$, wenn gilt: $Z' \subseteq Z \Leftrightarrow \kappa(Z) \subseteq \kappa(Z')$.*

Beispiel: V sei endlichdimensional, $\delta : V \longrightarrow V^*$ ein Isomorphismus (etwa: $e_i \longmapsto e_i^*$ nach Wahl einer Basis).

Für $U \subseteq V$ ist $\mathrm{Ann}(U) = \{\varphi \in V^* \mid \varphi|_U \equiv 0\}$ ein Teilraum von V^* und man hat

$$U \subseteq U' \Leftrightarrow \mathrm{Ann}(U) \supseteq \mathrm{Ann}(U').$$

Die Abbildung $\kappa : \mathcal{P} \longrightarrow \mathcal{P}$ mit

$$\kappa(U) = \delta^{-1}(\mathrm{Ann}(U))$$

ist dann eine Korrelation.

Bemerkung. *Für eine Korrelation ist die Umkehrabbildung ebenfalls eine Korrelation.*

Lemma 17.40. *Sei V endlichdimensional, $\kappa : P \longrightarrow P$ eine Korrelation. Dann gilt:*

a) $\dim \kappa(Z) = \dim \mathbb{P}(V) - (\dim Z + 1)$
b) $\kappa(Z \cap Z') = \kappa(Z) \vee \kappa(Z')$
c) $\kappa(Z \vee Z') = \kappa(Z) \cap \kappa(Z')$.

Beweis. a) Wir können Z in eine Fahne projektiver Unterräume einspannen.

$$\emptyset = Z_{-1} \subsetneqq Z_0 \subsetneqq \cdots \subsetneqq Z_n = \mathbb{P}(V)$$

mit $\dim(Z_i) = i$ und $Z_k = Z$ für $k = \dim Z$.

Anwenden von κ liefert die Fahne

$$Z'_n = \kappa(Z_{-1}) \supsetneqq \kappa(Z_0) \supsetneqq \cdots \supsetneqq \kappa(Z_n) = Z'_{-1}$$

der Länge $n+1$, so dass zwangsläufig $\kappa(Z_1) = \mathbb{P}(V)$, $\kappa(Z_n) = \emptyset$ ist und $\kappa(Z_k) = Z'_{n-k-1}$, also $\dim \kappa(Z_k) = n - k - 1$ wie behauptet.

b) Wir haben $\kappa(Z \cap Z') \supseteq \kappa(Z) \cup \kappa(Z')$, also $\supseteq \kappa(Z) \vee \kappa(Z')$.

Anwenden von κ^{-1} auf $\kappa(Z) \subseteq \kappa(Z) \vee \kappa(Z')$ gibt

$$\begin{aligned} \kappa^{-1}(\kappa(Z) \vee \kappa(Z') &\subseteq Z \\ \kappa^{-1}(\kappa(Z) \vee \kappa(Z') &\subseteq Z' \end{aligned}$$

also

$$\begin{aligned} \Rightarrow \quad Z \cap Z' &\supseteq \kappa^{-1}(\kappa(Z) \vee \kappa(Z')) \\ \Rightarrow \quad \kappa(Z \cap Z') &\subseteq \kappa(Z) \vee \kappa(Z'). \end{aligned}$$

c) Analog.

□

Satz 17.41. *(Dualitätsprinzip der projektiven Geometrie)*
Gegeben sei eine Aussage über projektive Unterräume des projektiven Raums $\mathbb{P}(V)$, die sich mit Hilfe von $\subseteq, \cap, \vee$ und dim *ausdrücken lässt. Dann ist auch die dazu duale Aussage richtig, bei der Unterräume der Dimension k durch solche der Dimension $n - k - 1$, $\subseteq$ durch $\supseteq$, $\vee$ durch $\cap$, $\cap$ durch $\vee$ ersetzt wird.*

Beispiel: In der projektiven Ebene $\mathbb{P}^2(K)$ gilt der Satz von Desargues: Gegeben seien Punkte $P_1, P_2, P_3, P'_1, P'_2, P'_3$ paarweise verschieden und so, dass sich $P_1P'_1$, $P_2P'_2$, $P_3P'_3$ in einem Punkt schneiden. Dann liegen die Schnittpunkte

$$\begin{aligned} P_1P_2 &\cap P'_1P'_2 \\ P_2P_3 &\cap P'_2P'_3 \\ P_1P_3 &\cap P'_1P'_3 \end{aligned}$$

auf einer gemeinsamen Geraden.

Die Dualisierung ergibt genau die Umkehrung des Satzes.

Definition 17.42. *Sei V ein K-Vektorraum, Q_0, Q_1, Q_2 seien paarweise verschiedene kollineare Punkte in $\mathbb{P}(V)$, $Q_3 \in \mathbb{P}(V)$, seien $(z_0 : z_1)$ die homogenen Koordinaten von Q_3 bzgl. des projektiven Koordinatensystems $(Q_0, Q_1; Q_2)$.*

Dann ist das Doppelverhältnis von Q_0, Q_1, Q_2, Q_3 definiert durch

$$DV(Q_0, Q_1, Q_2, Q_3) := \begin{cases} \frac{z_1}{z_0} & \text{falls } z_0 \neq 0 \\ \infty & \text{falls } z_0 = 0. \end{cases}$$

Bemerkung. a) $z_0 = 0 \Leftrightarrow Q_3 = Q_1$.

b) *Sei $V = K^2$, also $\mathbb{P}(V) = \mathbb{P}^1(K)$, seien $Q_0 = (1 : x_0)$, $Q_1 = (1 : x_1)$, $Q_2 = (1 : x_2)$, $Q_3 = (1 : x_3)$ Punkte im "affinen Teil" U_0 von $\mathbb{P}^1(K)$.*

Dann ist $(\lambda, \lambda x_0) + (\mu, \mu x_1) = (1, x_2)$ mit $\lambda = \frac{x_1 - x_2}{x_1 - x_0}$, $\mu = \frac{x_2 - x_0}{x_1 - x_0}$.

Damit ist $(1, x_3) = z_1(\lambda, \lambda x_0) + z_2(\mu, \mu x_1)$ mit $\frac{z_1}{z_0} = \frac{x_3 - x_0}{x_3 - x_1} : \frac{x_2 - x_0}{x_2 - x_1}$. Das Doppelverhältnis ist also das Verhältnis aus den Teilverhältnissen $(Q_0, Q_1; Q_3)$ und $(Q_0, Q_1; Q_2)$. (Daher der Name)

Lemma 17.43. *Seien $Q_0, \ldots, Q_3$ kollineare Punkte in $\mathbb{P}^n(V)$ wie oben mit homogenen Koordinatenvektoren*

$$\mathbf{x}^{(0)} = \begin{pmatrix} x_{00} \\ \vdots \\ x_{n0} \end{pmatrix}, \ldots, \begin{pmatrix} x_{03} \\ \vdots \\ x_{n3} \end{pmatrix} = \mathbf{x}^{(3)}.$$

Dann gibt es $0 \leq i < j \leq 3$, so dass $\det \begin{pmatrix} z_{i0} & z_{i1} \\ z_{j0} & z_{j1} \end{pmatrix} \neq 0$ ist.

Für jedes solche Paar (i, j) ist

$$DV(Q_0, Q_1, Q_2; Q_3) = \frac{\det \begin{pmatrix} x_{i0} & x_{i3} \\ x_{j0} & x_{j3} \end{pmatrix}}{\det \begin{pmatrix} x_{i3} & x_{i1} \\ x_{j3} & x_{j1} \end{pmatrix}} \cdot \frac{\det \begin{pmatrix} x_{i2} & x_{i1} \\ x_{j2} & x_{j1} \end{pmatrix}}{\det \begin{pmatrix} x_{i0} & x_{i2} \\ x_{j0} & x_{j2} \end{pmatrix}}.$$

Beweis. Die Existenz des Paares i, j folgt daraus, dass die Matrix aus der ersten und zweiten Spalte Rang 2 hat. Nach Definition des Doppelverhältnisses ist dann (Rechnung wie im Beispiel) zunächst $\mathbf{x}^{(2)} = \lambda \mathbf{x}^{(0)} + \lambda' \mathbf{x}^{(1)}$ und $\mathbf{x}^{(3)} = z_0 \lambda \mathbf{x}^{(0)} + z_1 \lambda' \mathbf{x}^{(1)}$.

Durch Lösen der hier gegebenen linearen Gleichungssysteme für λ, λ' einerseits, $\lambda z_0, \lambda' z_1$ andererseits mit Hilfe der Kramerschen Regel erhalten wir Formeln für $\frac{\lambda}{\lambda'}$, $\frac{\lambda'}{\lambda} \cdot \frac{z_1}{z_0}$ und damit für $\frac{z_1}{z_0}$ wie angegeben. □

Lemma 17.44. *Sei $\lambda = DV(Q_0, Q_1, Q_2, Q_3)$. Dann ist $DV(Q_{\sigma(0)}, Q_{\sigma(1)}, Q_{\sigma(2)}, Q_{\sigma(3)})$ für jedes $\sigma \in S_4$ eines von*

$$\lambda, \frac{1}{\lambda}, 1 - \lambda, 1 - \frac{1}{\lambda}, \frac{1}{1 - \lambda}, 1 - \frac{1}{1 - \lambda}.$$

Dabei wird das Doppelverhältnis von der Vierergruppe

$$\{\mathrm{Id}, (01)(12), (02)(13), (03)(12)\}$$

konstant gelassen

Beweis. Nachrechnen! □

Satz 17.45. *Seien Q_0, Q_1, Q_2, Q_3 kollineare Punkte in $\mathbb{P}(V)$, $Q_0, \ldots, Q_2$ paarweise verschieden, $\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$ eine Projektivität. Dann sind auch $\Phi(Q_0), \ldots, \Phi(Q_3)$ kollinear und haben das gleiche Doppelverhältnis wie $Q_0, \ldots, Q_3$.*
Umgekehrt gilt: Sind zusätzlich Q'_0, Q'_1, Q'_2, Q'_3 kollineare Punkte in $\mathbb{P}(W)$, $Q'_0, \ldots, Q'_2$ paarweise verschieden und hben Q'_0, Q'_1, Q'_2, Q'_3 das gleiche Doppelverhältnis wie Q_0, Q_1, Q_2, Q_3, so gibt es eine projektive Abbildung $\Phi : \mathbb{P}(V) \longrightarrow \mathbb{P}(W)$, die Q_i in Q'_i abbildet ($0 \leq i \leq 3$).

Beweis. Sei $Z \subseteq (V)$ die Gerade, auf der $Q_0, \ldots, Q_3$ liegen, und $\varphi(Z) =: Z' \subseteq \mathbb{P}(W)$; das ist eine Gerade, auf der die $\varphi(Q_i)$ liegen.

Wir haben Koordinatenabbildungen $c : Z \longrightarrow \mathbb{P}^1(K)$ bzgl. $(Q_0, Q_1; Q_2)$ $c' : Z' \longrightarrow \mathbb{P}^1(\kappa)$ bzgl. $(\Phi(Q_0), \Phi(Q_1); \Phi(Q_2))$ mit Umkehrabbildungen κ, κ'.

$$\begin{array}{ccc} & \nearrow_{\kappa} & Z \\ \mathbb{P}^1(K) & & \downarrow \Phi \\ & \searrow_{\kappa} & Z' \end{array} \qquad \text{ist kommutativ.}$$

Dann ist

$$\begin{aligned} DV(Q_0, Q_1, Q_2, Q_3) &= c(Q_3) \\ &= c'(\Phi(Q_3)) \\ &= DV(\Phi(Q_0), \Phi(Q_1), \Phi(Q_2), \Phi(Q_3)). \end{aligned}$$

□

18. Unendlich dimensionale Vektorräume und Zornsches Lemma

Definition 18.1. *Sei X eine Menge mit einer Relation $\leq$ auf X. Die Relation $\leq$ heißt eine partielle Ordnung, wenn gilt:*

a) *$x \leq y$ für alle $x \in X$ (reflexiv).*
b) *Aus $x \leq y$ und $y \leq x$ folgt $y = x$ (antisymmetrisch).*
c) *Aus $x \leq y$ und $y \leq z$ folgt $x \leq z$ (transitiv).*

Wir schreiben $x < y$ für ($x \leq y$ und $x \neq y$) sowie $x \geq y$ für $y \leq x$.
Die Ordnung heißt total, wenn für $x, y \in X$ stets $x \leq y$ oder $y \leq x$ gilt.
Eine Kette in X ist eine bzgl. $\leq$ total geordnete Teilmenge K von X.

Ein Element $x_0 \in X$ heißt maximal, wenn gilt: Ist $y \in x$ mit $y \geq x_0$, so ist $y = x_0$.
Die Ordnung $\leq$ heißt eine Wohlordnung, wenn gilt: Jede nichtleere Teilmenge $Y \subseteq X$ hat (genau) ein bzgl. $\leq$ kleinstes Element.

Wir wollen jetzt die drei folgenden Aussagen betrachten:

(A) Auswahlaxiom: $I \neq \emptyset$ sei eine Menge, $(X_i)_{i \in I}$ eine Familie von Teilmengen einer Menge X. Dann gibt es eine Abbildung $f : I \longrightarrow X$ mit $f(i) \in X_i$ für alle $i \in I$. Eine solche Abbildung nennen wir eine Auswahlfunktion.

(B) Zornsches Lemma: Sei X eine teilweise geordnete Menge, in der jede Kette eine obere Schranke hat. Dann gibt es in X ein maximales Element.

(C) Wohlordnungssatz: Jede Menge X besitzt eine totale Ordnung, bezüglich der sie wohlgeordnet ist.

Satz 18.2. *Das Zornsche Lemma folgt aus dem Auswahlaxiom.*

Beweis. (siehe Halmos: Naive Mengenlehre) Für $x \in X$ sei

$$S_x := s(x) := \{y \in X \mid y \leq x\}.$$

Es gilt: $x_1 \leq x_2 \Leftrightarrow S_{x_1} \subseteq S_{x_2}$, sei

$$\mathcal{S} := \{S_x \mid x \in X\} \subseteq \mathfrak{P}(X),$$

$\mathcal{S}$ ist durch Inklusion partiell geordnet.

Ist K eine Kette in X, so ist $\{S_x \mid x \in K\}$ eine Kette in $\mathcal{S}$. Sei $\mathcal{K} \subseteq \mathfrak{P}(X)$ die Menge aller Ketten $K \subseteq X$. $\mathcal{K}$ ist wieder durch Inklusion partiell geordnet, und es gilt: Ist $M \in \mathcal{K}$, $M' \subseteq M$, so ist $M' \in \mathcal{K}$. ist $\mathcal{C} \subseteq \mathcal{K}$ eine Kette in $\mathcal{K}$, so sind die Elemente von $\mathcal{C}$ total geordnete Teilmengen von X, und $\bigcup_{A \in \mathcal{C}} A$ ist ebenfalls eine total geordnete Teilmenge von X, denn: Sind $a, a' \in \bigcup_{A \in \mathcal{C}} A$, $a \in A$, $a' \in A'$, so ist o.E. $A' \subseteq A$ (weil $\mathcal{C}$ Kette bzgl. der Inklusion ist), also $a, a' \in A$, also $a \leq a'$ oder $a' \leq a$, weil A total geordnet ist.
Also: In $\mathcal{K}$ hat jede Kette eine obere Schranke, und $\bigcup_{A \in \mathcal{C}} A$ ist in jeder oberen Schranke von $\mathcal{C}$ enthalten, also obere Grenze.

Ferner gilt: Ist $K \in \mathcal{K}$, so hat K eine obere Schranke $x \in X$, also ist $K \subseteq S_x$ für ein geeignetes $x \in X$.

Ist K maximales Element von $\mathcal{K}$, so muss $x \in K$ für dieses x gelten, denn sonst wäre $K \cup \{x\}$ eine echt größere Kette.

Ist dann $y \in x$ mit $x \leq y$, so ist $K \cup \{y\}$ eine Kette, es gilt also $y \in K$, also $y \leq x$.

Also: Falls $\mathcal{K}$ ein maximales Element K hat, so ist jede obere Schranke von K ein maximales Element von X.

Wir haben also die Behauptung gezeigt, wenn wir zeigen: In $\mathcal{K}$ gibt es ein maximales Element. Damit haben wir die Behauptung auf den Spezialfall reduziert, in dem gilt:

a) Die Menge X ist eine Teilmenge von $\mathfrak{P}(Y)$ für eine gewisse Menge Y, geordnet durch Inklusion.
b) In X hat jede Kette K eine obere Grenze, nämlich $\bigcup_{A \in K} A$.
c) Ist $M \in X$ und $M' \subseteq M$, so ist auch $M' \in X$

Sei $f : \mathfrak{P}(Y) \setminus \{\emptyset\} \longrightarrow Y$ eine Abbildung mit $f(A) \in A$ für alle $\emptyset \neq A \subseteq Y$, ein solches f nennt man eine Auswahlfunktion, Auswahlfunktionen existieren nach dem Auswahlaxiom.

Für $A \in X$ sei

$$\hat{A} := \{y \in Y \mid A \cup \{y\} \in X\}. \quad (\supseteq A).$$

Dann definieren wir $g : X \longrightarrow X$ so:

$$g(A) = \begin{cases} A \cup f(\hat{A} \setminus A)\,(\in X), & \text{falls } \hat{A} \setminus A \neq \emptyset \\ A\,(\in X) & \text{sonst.} \end{cases}$$

Also: $g(A) = A \Leftrightarrow A$ ist maximales Element von X.

Zu zeigen: g hat einen solchen Fixpunkt. Wir nennen $Z \subseteq X$ einen Turm (zulässig), wenn gilt:

a) $\emptyset \in Z$
b) $A \in Z \Rightarrow g(A) \in Z$
c) Ist $\mathcal{C}$ eine Kette in Z, so ist $\bigcup_{A \in \mathcal{C}} A$ in Z.

Türme existieren (z.B. X selbst), und der Durchschnitt von Türmen ist ein Turm. Also: Sei Z_0 der Durchschnitt aller Türme in X (der kleinste Turm in X). Wir sind fertig, wenn wir zeigen können:
Z_0 ist Kette, dann ist $B := \bigcup_{A \in Z_0} A \in Z_0$, also $Z \ni g(B) \subseteq B$, also $g(B) = B$.

Nenne $C \in Z_0$ vergleichbar, wenn C mit allen $A \in Z_0$ vergleichbar ist, wenn also $A \subseteq C$ oder $C \subseteq A$ für alle $A \in Z_0$ gilt.

Zeige: Die vergleichbaren Elemente von Z_0 bilden einen Turm (da Z_0 minimal ist, sind dann alle Elemente von Z_0 vergleichbar, d.h., Z_0 ist eine Kette).
a) $\emptyset$ ist vergleichbar.

c) ist ebenfalls erfüllt (offensichtlich).
Kritisch ist b):
Sei $\mathcal{U} = \{A \in Z_0 \mid A \subseteq C \text{ oder } g(C) \subseteq A\}$. Für $A \in \mathcal{U}$ ist $A \subseteq C \subseteq g(C)$ oder $g(C) \subseteq A$.
Ist $\mathcal{C}$ eine Kette in $\mathcal{U}$, so ist $\bigcup_{A \in \mathcal{C}} A \in \mathcal{U}$, ferner ist $\emptyset \in \mathcal{U}$.

Zeige $g(A) \in \mathcal{U}$ für $A \in \mathcal{U}$. $A \subsetneqq C \Rightarrow g(A) \subseteq C$ (s.o.), also $g(A) \in \mathcal{U}$.

Sei $A \subsetneqq C$. Es gilt (weil C vergleichbar ist) $g(A) \subseteq C$ oder $C \subsetneqq g(A)$, also $A \subsetneqq C \subsetneqq g(A)$, Widerspruch, denn $g(A) \setminus A$ hat maximal ein Element. Also muss $g(A) \subseteq C$ gelten.
$A = C \Rightarrow g(A) = g(C) \supseteq g(C)$, also $g(A) \in \mathcal{U}$.
Ist $g(C) \subseteq A$, so ist $g(C) \subseteq g(A)$, also $g(A) \in \mathcal{U}$.
Also: $\mathcal{U}$ ist Turm, also $\mathcal{U} = Z_0$, also ist jedes $A \in Z_0$ mit $g(C)$ vergleichbar, also ist $g(C)$ vergleichbar.
Also: Die Menge der vergleichbaren Elemente von Z ist ein Turm, also ist ganz Z_0 vergleichbar, also Z_0 Kette, und wir sind fertig. □

Satz 18.3. *Aus dem Zornschen Lemma folgt der Wohlordnungssatz.*

Beweis. In der Menge X betrachten wir Teilmengen $A \neq \emptyset$ mit einer Wohlordnung $<_A$. Wir sagen, $(B, <_B)$ sei eine Fortsetzung von $(A, <_A)$, wenn $A \subseteq B$ gilt und $<_B |_{A \times A} = <_A$ sowie $\min(A) = \min(B)$ gilt, wir schreiben dann $(A, <_A) \prec (B, <_B)$.

Die Menge Z der Paare $(A, <_A)$ mit dieser Relation ist teilweise geordnet. Ist $(A_i, <_{A_i})_{i \in I}$ eine Kette solcher Paare, so wird auf $A := \bigcup_{i \in I} A_i$ durch $a <_A b \Leftrightarrow a <_{A_i} b$ für ein i mit $a, b \in A_i$ eine Wohlordnung definiert.

Also hat in Z jede Kette eine obere Schranke, nach dem Zornschen Lemma gibt es in Z ein maximales Element $(A, <_A)$.

Wäre $x_0 \in X$ mit $x_0 \notin A$, so könnte man durch $a < x_0$ für alle $a \in A$ die Wohlordnung von A auf $A \cup \{x_0\}$ fortsetzen, im Widerspruch zur Maximalität. Also ist $A = X$ und $<_A$ die gesuchte Wohlordnung von X. □

Satz 18.4. *Aus dem Wohlordnungssatz folgt das Auswahlaxiom.*

Beweis. Seien $X, I, (X_i)_{i \in I}$ wie im Auswahlaxiom gegeben. Auf I betrachten wir eine Wohlordnung $\leq$.
Für $j \in I$, sei $I_j := \{i \in I \mid i \leq j\}$.
Sei $J = \{j \in I \mid \exists f_j : I_j \longrightarrow X \text{ mit } f(i) \in X_i \text{ für alle } i \in I_j\}$.
Ist $j_2 \in J$ und $j_1 \in I$ mit $j_1 < j_2$, so ist $j_1 \in J$.
Ferner: Die Abbildung $f : J \longrightarrow X$, die durch $f(j) = f_j(j)$ gegeben ist, ist eine Auswahlfunktion (also $f(j) \in X_j$ für alle $j \in J$).
Annahme: $J \neq I$.
Dann sei i_0 das kleinste Element von $I \setminus J$. Es gilt $i_0 > j$ für alle $j \in I$,

also ist $S_{i_0} = J \cup \{i_0\}$ und wir können $g: J \cup \{i_0\} \longrightarrow X$ durch

$$g(i) = \begin{cases} f(j) & j \in J \\ x_{i_0} & j = i_0 \end{cases}$$

mit beliebigen $x_{i_0} \in X_{i_0}$ definieren.
Dann ist aber $i_0 \in J$, Widerspruch.
Also ist $J = I$ und f ist die gesuchte Auswahlfunktion. □

Satz 18.5. *V sei ein K-Vektorraum.*

a) *Ist $M \subseteq V$ eine Menge linear unabhängiger Vektoren, so gibt es eine Basis von V, die M enthält. Insbesondere hat V eine Basis.*
b) *Ist $\mathcal{B}$ eine Basis von V (als Menge von Vektoren betrachtet) und $M \neq \emptyset$ eine Menge linear unabhängiger Vektoren, so gibt es eine Menge $M' \subseteq \mathcal{B}$, so dass $(\mathcal{B} \setminus M') \cup M$ Basis von V ist.*

Beweis. a) Sei $X = \{S \supseteq M \mid S \text{ ist linear unabhängig}\}$. X ist durch Inklusion teilweise geordnet und nicht leer.
Sei K eine Kette in X. Dann ist $\tilde{S} := \bigcup_{S \in K} S$ linear unabhängig. Ist nämlich $\{v_1, \ldots, v_n\}$ eine endliche Teilmenge von (paarweise verschiedenen) $v_i \in \tilde{S}$, so ist $v_i \in S_i$ für ein $S_i \in K$. Da K total geordnet ist, gibt es ein maximales S_{i_0} unter den S_i, also $v_i \in S_{i_0}$ für alle i. Da S_{i_0} linear unabhängig ist, sind die v_i linear unabhängig.
Also hat jede Kette eine obere Schranke, es gibt also nach dem Zornschen Lemma ein maximales S in X.

Wir wissen aber: Jede maximale linear unabhängige Teilmenge ist eine Basis.

b) Sei jetzt X die Menge aller $S \subseteq M$, die man in eine Basis von V hinein tauschen kann. X ist nicht leer, denn die 1-elementigen Teilmengen kann man hinein tauschen.

X ist teilweise geordnet, und wie in a) zeigt man, dass jede Kette eine obere Schranke hat. Sei S_0 ein maximales Element von X, $\mathcal{B}' = (\mathcal{B} \setminus S_0') \cup S_0$ eine Basis von V. Wäre $S_0 \neq M$, so könnte man ein $v \in M$ in $\mathcal{B}'$ hinein tauschen, aber nicht gegen ein Element von $S_0 \subseteq M$, weil M linear unabhängig ist. Also könnte man $S_0 \cup \{v\}$ in $\mathcal{B}$ hineintauschen, Widerspruch.
Also ist $M = S_0$.

□

Satz 18.6. *Sei $(M_i)_{i \in I}$ eine Familie von Mengen. Dann ist*

$$\prod_{i \in I} M_i := \{f : I \longrightarrow \bigcup_{i \in I} M_i \mid f(i) \in M_i \ \forall i\}$$

nicht leer und hat die folgende universelle Eigenschaft:
Für $j \in I$ sei $p_j : \prod_{i \in I} M_i \longrightarrow M_j$ gegeben durch $p_j(f) = f(j) \in M_j$.
Ist dann X eine Menge mit Abbildungen $h_i \longrightarrow X \longrightarrow M_i$ für alle

$i \in I$, so gibt es genau eine Abbildung $h : X \longrightarrow \prod_{i\in I} M_i$ mit $p_j \circ h = h_j$ für alle $j \in I$.
Ist jedes $M_i = V_i$ ein K-Vektorraum und schreiben wir wie üblich $(v_i)_{i\in I}$ für $f : I \to \cup_{i\in I} V_i$ mit $f(i) = v_i \in V_i$, so wird $\prod_{i\in I} V_i$ durch

$$\begin{array}{rcl} (v_i)_{i\in I} + (w_i)_{i\in I} & = & (v_i + w_i)_{i\in I} \\ c(v_i)_{i\in I} & = & (cv_i)_{i\in I} \end{array}$$

zu einem K-Vektorraum, die p_j sind dann lineare Abbildungen, und sind oben X ein K-Vektorraum und die h_i lineare Abbildungen, so ist auch h linear.

Beweis. Klar. □

Definition und Satz 18.7. *Seien V_i ($i \in I$) wie oben. Die (externe) direkte Summe $\bigoplus_{i\in I} V_i$ der V_i ist*

$$\tilde{V} := \{(v_i)_{i\in I} \in \prod_{i\in I} V_i \mid v_i = \mathbf{0} \text{ für fast alle } i \in I\}.$$

$\bigoplus_{i\in I} V_i$ ist ein K-Vektorraum und hat folgende universelle Eigenschaft: Sei

$$g_i := V_i \longrightarrow \bigoplus_{i\in I} V_i$$

gegeben durch

$$g_i(v) = (v_j)_{j\in I} \text{ mit } v_j = \begin{cases} \mathbf{0} & i \neq j \\ v & i = j \end{cases}.$$

Sei X ein K-Vektorraum mit linearen Abbildungen

$$h_i : V_i \longrightarrow X.$$

Dann gibt es genau eine lineare Abbildung

$$h : \bigoplus_{i\in I} V_i \longrightarrow X \text{ mit } h \circ g_i = h_i \text{ für alle } i \in I.$$

Beweis. Man setze

$$h((v_i)_{i\in I}) = \sum_{i\in I} h_i(v_i).$$

Das ist wohldefiniert, weil auf der rechten Seite nur endlich viele Summanden $\neq \mathbf{0}$ stehen. Dass h die gewünschte Eigenschaft hat und eindeutig bestimmt ist, ist klar. □

Beispiel: V sei ein K-Vektorraum mit Basis $\{v_i \mid i \in I\}$. Sei

$$V_i = \langle v_i \rangle = \mathrm{Lin}(v_i).$$

Dann ist

$$V \cong \bigoplus_{i\in I} V_i.$$

Satz 18.8. *Sei I, V_i wie oben gegeben,*

$$V = \bigoplus_{i \in I} V_i.$$

Dann ist

$$V^* \cong \prod_{i \in I} V_i^*.$$

Beweis. Seien $\varphi_i : V_i \longrightarrow K$ aus $\prod_{i \in I} V_i^*$ gegeben. Dann kann man $\varphi : V \longrightarrow K$ definieren durch $\varphi((v_i)_{i \in I}) = \sum_{i \in I} \varphi_i(v_i)$, das ist wohldefiniert, weil fast alle Summanden $\mathbf{0}$ sind. Umgekehrt: Ist $\varphi : V \longrightarrow K$ linear gegeben, so setze man $\varphi_j(v_j) = \varphi((w_i)_{i \in I})$, wo

$$w_i = \begin{cases} v_j & i = j \\ \mathbf{0} & \text{sonst} \end{cases}$$

ist.
Die beiden Abbildungen

$$\begin{array}{rcl} (\varphi_i)_{i \in I} & \longmapsto & \varphi, \\ \varphi & \longmapsto & (\varphi_j)_{j \in I} \end{array}$$

sind zueinander inverse Bijektionen. □

Korollar 18.9. *Sei V ein unendlichdimensionaler K-Vektorraum. $\iota : V \longrightarrow V^{**}$ die kanonische Einbettung, die durch $\iota(v)(\varphi) = \varphi(v)$ gegeben ist. Dann ist ι nicht surjektiv.*

Beweis. Ist $\{v_i \mid i \in I\}$ eine Basis von V und $V_i = \langle v_i \rangle$, so ist

$$V^* \cong \prod_{i \in I} V_i^* \quad \text{mit } V_i^* \cong K$$

Seien $\varphi_i \in V^*$ die zu den v_i dualen Linearformen (also $\varphi_i(v_j) = \delta_{ij}$) und $\psi \in V^{**}$ gegeben durch $\psi(\varphi_i) = 1$ für alle $i \in I$.

Für $v = \sum_{i \in I_0} c_i v_i$ mit einer endlichen Menge $I_0 \subseteq I$ ist daher

$$\iota(v)(\varphi_j) = 0,$$

falls $j \notin I_0$ gilt.
Also ist $\iota(v) \neq \psi$ für alle $v \in V$. □