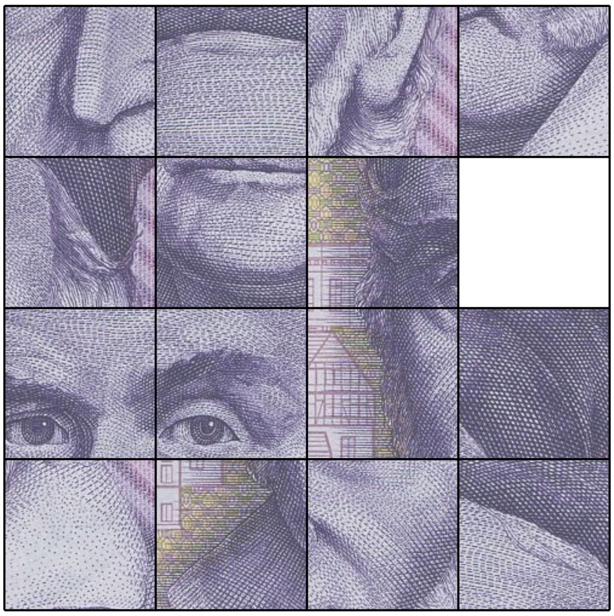
Lineare Algebra

Benjamin Sambale
Leibniz Universität Hannover

Version: 27. Februar 2025



Inhaltsverzeichnis

Einleitung

Vorwort

Dieses Skript ist eine Erweiterung meiner Vorlesung Lineare Algebra A&B im Wintersemester 2020/21 und Sommersemester 2021 an der Leibniz Universität Hannover. Während sich die Vorlesung hauptsächlich an Studierende der Informatik gewendet hat, ist das vorliegende Skript an Mathematiker gerichtet. Einige der präsentierten Sätze (u. a. von Fillmore, Mirsky, Mazur-Ulam, Schur-Horn und Frobenius) findet man nur schwer in der Standard-Literatur. Außerdem wird die Frobenius-Normalform in voller Allgemeinheit behandelt. Die Übungsaufgaben beschränken sich auf theoretische Aspekte und sollten durch praktische Beispiele ergänzt werden. Ich danke Gereon Koßmann für Fehlerhinweise (weitere Hinweise sind willkommen). Das folgende Buch deckt in etwa die Themen dieses Skripts ab:

Hoffman, Kunze, *Linear algebra*, 2nd edition, Prentice-Hall, New Jersey, 1971

Motivation

Sie haben zu verschiedenen Zeitpunkten $t_1 = 1, t_2 = 2, \ldots$ durch physikalische Experimente Messdaten $d_1 = -2, d_2 = 3, \ldots$ gewonnen. Aus theoretischen Überlegungen sei bekannt, dass diese Daten einem Gesetz folgen, das heißt, es gibt eine Funktion f mit $f(t_i) = d_i$ für $i = 1, 2, \ldots$. Dabei hängt f (linear) von unbekannten Parametern $x_1, x_2, \ldots$ ab, zum Beispiel $f(t) = t^2x_1 - tx_2 + x_3$. Die Bestimmung dieser Parametern auf Grundlage der Messdaten führt auf ein lineares Gleichungssystem:

$$\begin{aligned} x_1 - x_2 + x_3 &= -2 \\ 4x_1 - 2x_2 + x_3 &= 3 \\ &\vdots \end{aligned} \tag{S}$$

Wir beantworten unter anderem folgende Fragen:

- Wann ist das System (S) lösbar? (Satz 6.4)
- Wie viele Lösungen gibt es? (Bemerkung 6.7(b))
- Welche Struktur hat die Lösungsmenge? (Satz 6.6)
- Wie berechnet man alle Lösungen in der Praxis? (Satz 6.15)

Die entwickelten Methoden (Vektorräume, Matrizen und lineare Abbildungen) haben zahlreiche Anwendungen in anderen Gebieten:

- Bildverarbeitung: Wie wertet man verzerrte Blitzerfotos aus?
- Suchmaschinen: Nach welchen Kriterien bewertet Google Internetseiten?

- Codierungstheorie: Wie erkennt und korrigiert man Fehler bei der Übertragung digitaler Daten?
- Elektrotechnik: Wie berechnet man Widerstände in Schaltkreisen?
- Meteorologie: Wie sagt man das Wetter von Morgen voraus?
- Stochastik: Mit welcher Wahrscheinlichkeit gelangt man nach einer Irrfahrt zum Ziel?

Notation

d. h.	das heißt
ggf.	gegebenenfalls
o. B. d. A.	ohne Beschränkung der Allgemeinheit
vgl.	vergleiche
w, **f**	wahr, falsch
$\wedge, \vee, \neg, \Rightarrow, \Leftrightarrow, \exists, \forall$	logische Ausdrücke
$:=$, $:\Leftrightarrow$	linke Seite wird durch rechte Seite definiert
$\square$	Beweisende
$\varnothing$	leere Menge
$\mathcal{P}(M)$	Potenzmenge von M
$\cup$, $\dot\cup$, $\cap$, $\setminus$	Vereinigung (disjunkt), Durchschnitt, Differenz von Mengen
$\lvert A\rvert$	Mächtigkeit von A (Anzahl der Elemente)
(a, b), $(x_1, \dots, x_n)$	Paar, n-Tupel
$A_1 \times \ldots \times A_n$	kartesisches Produkt von Mengen $A_1, \dots, A_n$
$[a]$	Äquivalenzklasse von $a \in A$
$A \to B$, $a \mapsto b$	Abbildung von A nach B
$\mathrm{Abb}(A, B)$	Menge aller Abbildungen $A \to B$
$\mathbb{N}, \mathbb{N}_0, \mathbb{Z}$	natürliche Zahlen (ohne und mit 0), ganze Zahlen
$\mathbb{Q}, \mathbb{R}, \mathbb{R}_{>0}, \mathbb{C}$	rationale, reelle (positive) und komplexe Zahlen
$\mathbb{F}_2$	Körper mit zwei Elementen
$K^\times$	$= K \setminus \{0\}$ (multiplikative Gruppe)
$f_{\lvert A}$, f^{-1}, $f \circ g$	Einschränkung, Umkehrfunktion und Komposition von Funktionen
f^*	duale oder zu f adjungierte Abbildung
$f(A), f^{-1}(B), \mathrm{Ker}(f)$	Bild, Urbild, Kern von f
id, id_V	Identität (auf V)
0_K, 0_V, 1_G	Nullelement, Nullvektor, neutrales Element in G
δ_{ij}	Kronecker-Delta
$\sum_{i=1}^n \lambda_i v_i$	Linearkombination
$U \leq V$, $U < V$, V/U	Unterraum (echter), Faktorraum
$H \leq G$, $H < G$	Untergruppe (echte) von G
$U \cong V$	isomorphe Vektorräume
V^*, V^{**}	Dualraum, Bidualraum
U^0, U_0	duale Komplemente
$U \oplus W$, $U \times W$	direkte Summe/Produkt von U und W
$e_1, \dots, e_n$	Standardbasis von K^n
$b_1^*, \dots, b_n^*$	duale Basis
$\langle S\rangle, \langle s_1, \dots, s_n\rangle$	Spann von S
$\dim_K V = \dim V$	Dimension von V über K
${}_B[v]$	Koordinatendarstellung von v bzgl. der Basis B

$\mathrm{Hom}(V,W)$	Vektorraum aller linearen Abbildungen $V \to W$
$\mathrm{End}(V)$	$= \mathrm{Hom}(V,V)$
$K^{n\times m}$	Vektorraum der $n \times m$-Matrizen über K
$\mathrm{GL}(V)$, $\mathrm{GL}(n,K)$	allgemeine lineare Gruppe
$\mathrm{SL}(n,K)$	spezielle lineare Gruppe
$\mathrm{O}(V)$, $\mathrm{O}(n,K)$	orthogonale Gruppe
$\mathrm{SO}(n,K)$	spezielle orthogonale Gruppe
$\mathrm{U}(V)$, $\mathrm{U}(n,\mathbb{C})$	unitäre Gruppe
$\mathrm{SU}(n,\mathbb{C})$	spezielle unitäre Gruppe
$\mathrm{Aff}(V)$	affine Gruppe
$0_{n\times m}$, 0_n, 1_n	Nullmatrix, Einheitsmatrix
E_{st}	Standardmatrix mit 1 an Position (s,t)
$A \sim B$	A zeilen-äquivalent zu B
$A \approx B$	A ähnlich zu B
$(A\|b)$	erweiterte Koeffizientenmatrix
A^{t}, A^{-1}, $A^{-\mathrm{t}}$	Transponierte, Inverse, Transponiert-Inverse von A
$\widehat{A}$, $\widetilde{A}$	Zeilenstufenform, komplementäre Matrix von A
$\overline{A}$, A^*	komplex-konjugierte, adjungierte Matrix von A
$\mathrm{rk}(A)$, $\mathrm{tr}(A)$, $\det(A)$	Rang, Spur, Determinante von A
${}_C[f]_B$, $[f]$, ${}_C\Delta_B$	Darstellungsmatrix, Basiswechselmatrix
$E_\lambda(f)$, $H_\lambda(f)$	Eigenraum, Hauptraum zum Eigenwert λ von f
S_n	symmetrische Gruppe vom Grad n
A_n	alternierende Gruppe vom Grad n
$\mathrm{sgn}(\sigma), P_\sigma$	Signum, Permutationsmatrix von $\sigma \in S_n$
$K[X]$	Vektorraum der Polynome mit Koeffizienten in K
$K(X)$	Körper der rationalen Funktionen
α'	Ableitung von $\alpha \in K[X]$
$\deg(\alpha)$	Grad von $\alpha \in K[X]$
$\alpha \mid \beta$	α teilt β
$\alpha \equiv \beta \pmod{\delta}$	$\delta \mid \alpha - \beta$
χ_A, χ_f, μ_A, μ_f	charakteristisches Polynom, Minimalpolynom von A, f
μ_v	Minimalpolynom des zyklischen Unterraums
$[v,w]$, $\|v\|$	Skalarprodukt, Norm von v
$v \perp w$	v und w sind orthogonal, d. h. $[v,w]=0$
π	Länge des Halbkreisbogens mit Radius 1
$\cos\varphi$, $\sin\varphi$	Kosinus, Sinus von φ
$S^\perp$	orthogonales Komplement von $S \subseteq V$
$v \times w$	Kreuzprodukt von v und w
$D(\varphi)$, $S(\varphi)$	Drehung, Spiegelung in $\mathbb{R}^2$
S_v	Spiegelung an de Hyperebene $v^\perp$
$\mathrm{Re}(z)$, $\mathrm{Im}(z)$, $\overline{z}$	Realteil, Imaginärteil, komplexe Konjugation von z
i	imaginäre Einheit
$\mathrm{Bil}(V)$	Vektorraum der Bilinearformen auf V
${}_B[\beta]_B$	Gram-Matrix einer Bilinearform β
$\mathrm{ind}(\beta)$, $\mathrm{ind}(A)$	Index von $\beta \in \mathrm{Bil}(V)$, $A \in K^{n\times n}$
$J_n(\lambda)$	Jordanblock für $\lambda \in K$ der Größe $n \times n$
$J_k(\gamma)$	verallgemeinerter Jordanblock für $\gamma \in K[X]$
$B(\alpha)$	Begleitmatrix von $\alpha \in K[X]$
$\mathrm{C}(f)$, $\mathrm{C}(A)$	Zentralisator von $f \in \mathrm{End}(V)$, $A \in K^{n\times n}$

Konventionen

- K ist stets ein Körper, V ein endlich-dimensionaler K-Vektorraum, Unterräume heißen meist U, W, V_1 etc.
- Mengen und Matrizen werden mit lateinischen Großbuchstaben bezeichnet ($A, B, \ldots, M, \ldots$).
- Elemente von Mengen werden mit Kleinbuchstaben bezeichnet, Vektoren mit u, v, w, natürliche Zahlen mit n, m, k, l, Abbildungen mit f, g, h etc.
- Für Mengen von Mengen benutzt man oft „geschwungene“ Buchstaben ($\mathcal{M}$, $\mathcal{P}$)
- Für Polynome, Skalare (Körperelemente im Kontext von Vektorräumen) und Bilinearformen verwenden wir griechische Buchstaben. Die gebräuchlichsten sind:

α	β	γ, Γ	δ, Δ	ϵ, ε	ζ	η	$\theta, \vartheta, \Theta$	λ, Λ	μ
alpha	beta	gamma	delta	epsilon	zeta	eta	theta	lambda	my

ν	ξ	π, Π	ρ, ϱ	σ, Σ	τ	φ, ϕ, Φ	χ	ψ, Ψ	ω, Ω
ny	xi	pi	rho	sigma	tau	phi	chi	psi	omega

Lineare Algebra I

1 Aussagenlogik und Mengenlehre

1.1 Aussagen

Bemerkung 1.1. Die Sprache der Mathematik basiert auf logischen Prinzipien, die man letztlich als gegeben hinnehmen muss. Alle „höheren“ mathematischen Objekte lassen sich auf mengentheoretische Konstrukte zurückführen. Wir behandeln diese Themen hier nur soweit, wie sie zum Verständnis der linearen Algebra benötigt werden. Mehr Informationen findet man in meinem Skript zur Mengenlehre.

Definition 1.2.

- Eine *Aussage* A ist ein deutscher Satz, der entweder den *Wahrheitswert wahr* (**w**) oder *falsch* (**f**) annimmt. Man sagt dann A *gilt* bzw. A *gilt nicht*.
- Für Aussagen A und B sind auch $\neg A$ (*nicht* A), $A \wedge B$ (A *und* B), $A \vee B$ (A *oder* B), $A \Rightarrow B$ (A *impliziert* B) und $A \Leftrightarrow B$ (A *genau dann wenn* B) Aussagen mit folgenden Wahrheitswerten:

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$
w	**w**	**f**	**w**	**w**	**w**	**w**
w	**f**	**f**	**f**	**w**	**f**	**f**
f	**w**	**w**	**f**	**w**	**w**	**f**
f	**f**	**w**	**f**	**f**	**w**	**w**

- Zwei Aussagen A und B nennt man *äquivalent*, falls $A \Leftrightarrow B$ wahr ist, d. h. wenn A und B den gleichen Wahrheitswert haben.
- Ein *Prädikat* ist eine Eigenschaft $A = A(x)$, die erst durch Einsetzen einer Variablen x zu einer Aussage wird. Ggf. sind $\forall x : A(x)$ (*für alle* x gilt $A(x)$) und $\exists x : A(x)$ (*es existiert* ein x, sodass $A(x)$ gilt) Aussagen.

Beispiel 1.3. Folgende Sätze sind Aussagen (selbst wenn wir den Wahrheitswert nicht kennen):

- Alle blauen Katzen können fliegen (**w**).
- $1 + 1 = 3$ (**f**).
- Jede gerade Zahl größer als 2 ist die Summe zweier Primzahlen (?)[1]

Keine Aussagen dagegen sind:

- Sei $\epsilon > 0$ (*Annahme*).
- $a^2 + b^2 = c^2$ (*Gleichung*).
- Dieser Satz ist falsch (*Paradoxon*).

[1] Goldbachs Vermutung

Aus dem Prädikat $x > 0$ kann man die wahre Aussage $\forall x > 4 : x > 0$ bilden.

Bemerkung 1.4.

(a) Im Gegensatz zum alltäglichen Sprachgebrauch unterscheidet sich das mathematische *oder* vom *entweder oder*. Das heißt, die Aussage $\mathbf{w} \vee \mathbf{w}$ ist wahr. Wir führen kein eigenständiges Symbol für *entweder oder* ein.[2] Unterscheiden Sie außerdem die Formulierungen „Es gibt ein…“ und „Es gibt genau ein…“.

(b) Die Wahrheit der Aussage $\mathbf{f} \Rightarrow \mathbf{f}$ irritiert viele Anfänger (siehe Beispiel 1.3). Interpretation: Wenn die Voraussetzung nicht erfüllt ist, ist auch nichts zu zeigen. Man unterscheide außerdem die Aussage $A \Rightarrow B$ von ihrer *Umkehrung* $B \Rightarrow A$.[3]

(c) Für Aussagen $A_1, \ldots, A_n$ definiert man $A_1 \wedge \ldots \wedge A_n$ durch $\forall i : A_i$ und $A_1 \vee \ldots \vee A_n$ durch $\exists i : A_i$.

(d) Um den Wahrheitswert einer Aussage A zu bestimmen, führt man *Äquivalenzumformungen* durch, d. h. man ersetzt A durch eine äquivalente Aussage. Dafür sind folgende Schlussregeln nützlich.[4]

Lemma 1.5. *Seien A, B und C Aussagen. Dann gilt:*

(a) Die folgenden Aussagen sind äquivalent zu A:

$$\neg\neg A, \qquad A \wedge \mathbf{w}, \qquad A \vee \mathbf{f}, \qquad A \wedge A, \qquad A \vee A, \qquad \mathbf{w} \Rightarrow A$$

(b) $A \wedge B$ und $B \wedge A$ sind äquivalent sowie $A \vee B$ und $B \vee A$ (Kommutativgesetz).

(c) Es gilt $A \vee \neg A$ (Satz vom ausgeschlossenen Dritten) *und $\neg(A \wedge \neg A)$* (Satz vom Widerspruch).

(d) $A \wedge (B \vee C)$ und $(A \wedge B) \vee (A \wedge C)$ sind äquivalent sowie $A \vee (B \wedge C)$ und $(A \vee B) \wedge (A \vee C)$ (Distributivgesetz).

(e) $\neg(A \wedge B)$ und $\neg A \vee \neg B$ sind äquivalent sowie $\neg(A \vee B)$ und $\neg A \wedge \neg B$ (De Morgansche Regeln).

(f) $A \Rightarrow B$, $\neg A \vee B$ und $(\neg B) \Rightarrow (\neg A)$ sind äquivalent (Kontraposition).

(g) $(A \Rightarrow B) \wedge (B \Rightarrow C)$ impliziert $A \Rightarrow C$ (Transitivität).

(h) Aus $A \wedge (A \Rightarrow B)$ folgt B (Modus ponens).

(i) $A \Leftrightarrow B$ ist äquivalent zu $(A \Rightarrow B) \wedge (B \Rightarrow A)$.

Beweis. Alle Behauptungen lassen sich leicht durch Wahrheitstabellen verifizieren. Bei drei Variablen muss man dafür $2^3 = 8$ Fälle unterscheiden (wer einen schnelleren Weg findet, kann eine Million Dollar verdienen[5]). Alternativ kann man einige der Behauptungen aus bereits bewiesenen ableiten. So folgt die zweite De Morgansche Regel aus der ersten:

$$\neg(A \vee B) \overset{(a)}{\iff} \neg((\neg\neg A) \vee (\neg\neg B)) \iff \neg(\neg(\neg A \wedge \neg B)) \overset{(a)}{\iff} (\neg A \wedge \neg B). \qquad \square^6$$

[2] In der Informatik spricht man von `XOR`.
[3] Man könnte auch $A \Leftarrow B$ schreiben.
[4] Ein *Lemma* ist ein Hilfssatz mit wenig eigener Bedeutung.
[5] Das SAT-*Problem* der theoretischen Informatik ist NP-vollständig. Eines der sieben *Millenniumsprobleme* fragt, ob P = NP.
[6] Diese Box markiert das Ende eines Beweises.

Bemerkung 1.6.

(a) Die De Morganschen Regeln lassen sich allgemeiner in der Form $(\neg\forall x : A(x)) \Leftrightarrow (\exists x : (\neg A(x)))$ und $(\neg\exists x : A(x)) \Leftrightarrow (\forall x : (\neg A(x)))$ für Prädikate formulieren.

(b) Lemma 1.5 zeigt, dass man allein mit den Symbolen $\neg$ und $\wedge$ alle weiteren Terme ausdrücken kann. Zu Gunsten der Lesbarkeit sollte man jedoch alle Symbole sparsam einsetzen.

1.2 Mengen

Definition 1.7 (Cantor)**.** Eine *Menge* M ist eine Zusammenfassung von bestimmten wohlunterschiedenen Objekten x unserer Anschauung oder unseres Denkens zu einem Ganzen.[7] Man sagt dann: x ist ein *Element* von M und schreibt $x \in M$ sowie $M = \{x : x \in M\}$ (bzw. $x \notin M$ für $\neg(x \in M)$). Die Anzahl $|M|$ der Elemente von M heißt *Kardinalität* oder *Mächtigkeit* von M. Im Fall $|M| < \infty$ heißt M *endlich* und anderenfalls *unendlich*.

Bemerkung 1.8.

(a) Definition 1.7 ist ungenau, denn sie lässt Mengen zu, die zu logischen Widersprüchen führen. Sei beispielsweise

$$M := \{x : x \notin x\} \qquad \text{(Russell\textit{sche Antinomie})}^8$$

Die Aussage $M \in M$ kann dann weder wahr noch falsch sein. In der modernen Mathematik verhindert man solche Widersprüche durch Einführung eines *Axiomensystems*, d. h. man gibt den Wahrheitswert von möglichst wenigen „elementaren“ Aussagen (Axiomen) vor. Weit verbreitet ist das Zermelo-Fraenkel-System. Eines seiner Axiome besagt:

Mengen sind genau dann gleich, wenn sie die gleichen Elemente enthalten.

Dies impliziert, dass die Elemente einer Menge keine feste Reihenfolge haben. Es gilt also $\{2, 1, 1, 2, 2\} = \{1, 2\}$.

(b) In manchen Situationen benötigt man zusätzlich das sogenannte *Auswahlaxiom* (siehe Beispiel 2.3). Es wird von den meisten Mathematikern anerkannt, obwohl es die Konstruktion kontraintuitiver Mengen zulässt: Das *Banach-Tarski-Paradoxon* besagt beispielsweise, dass man eine Kugel vom Volumen 1 in fünf Teile zerlegen kann, die anders zusammengesetzt zwei Kugeln vom Volumen 1 ergeben.

(c) Nach Gödels zweiten *Unvollständigkeitssatz* ist es unmöglich zu beweisen, dass die Zermelo-Fraenkel-Axiome keine Widersprüche liefern. Ist dies tatsächlich der Fall (wovon die meisten Mathematiker ausgehen), so besagt Gödels erster Unvollständigkeitssatz, dass es Aussagen gibt, deren Wahrheitswert sich nicht bestimmen lässt.[9] Das bekannteste Beispiel hierfür ist die *Kontinuumshypothese* (siehe Bemerkung 2.11(f)).

[7] Cantors Wortlaut

[7] Das Symbol := besagt, dass die linke Seite durch die rechte Seite festgelegt wird.

[9] Die Beweisidee besteht darin die Aussage „Dieser Satz ist nicht beweisbar.“ zu formalisieren.

Definition 1.9.

(a) Für Mengen A und B sei

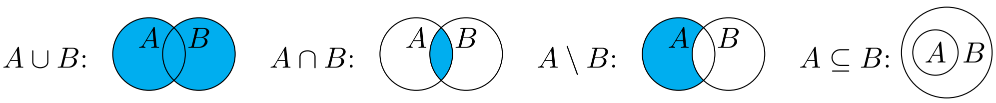
$$\begin{aligned} \varnothing &:= \{\} && (\textit{leere Menge}), \\ A \cup B &:= \{x : x \in A \vee x \in B\} && (\textit{Vereinigung}), \\ A \cap B &:= \{x : x \in A \wedge x \in B\} && (\textit{Durchschnitt}),^{10} \\ A \setminus B &:= \{x : x \in A \wedge x \notin B\} && (\textit{Differenz}).^{11} \end{aligned}$$

(b) Im Fall $A \cup B = B$ ist A eine *Teilmenge* von B. Man schreibt dann $A \subseteq B$ oder $A \subsetneq B$, falls zusätzlich $A \neq B$ (man spricht dann von einer *echten* Teilmenge[12]). Ist A keine Teilmenge von B, so schreibt man $A \nsubseteq B$.

(c) Man nennt A und B *disjunkt*, falls $A \cap B = \varnothing$. Ggf. nennt man $A \dot\cup B := A \cup B$ eine *disjunkte Vereinigung*.

Bemerkung 1.10.

(a) Beziehungen zwischen Mengen lassen sich durch VENN-Diagramme veranschaulichen:



Achtung: Sind mehr als drei Mengen im Spiel, so kann die allgemeine Situation nicht mehr durch Kreise dargestellt werden.[13]

(b) Vereinigung und Durchschnitt von beliebig vielen Mengen A_i (wobei i aus einer Indexmenge I stammt) lassen sich wie folgt definieren:

$$\bigcup_{i \in I} A_i := \{x : \exists i \in I : x \in A_i\}, \qquad \bigcap_{i \in I} A_i := \{x : \forall i \in I : x \in A_i\}.$$

Ist A die disjunkte Vereinigung von Mengen A_i, so spricht man von einer *Partition* von A.

(c) Um die Gleichheit von Mengen $A = B$ zu beweisen, ist es oft einfacher die äquivalente Aussage $(A \subseteq B) \wedge (B \subseteq A)$ zu zeigen.

Beispiel 1.11.

(a) Die Menge der *natürlichen Zahlen* $\mathbb{N} := \{1, 2, 3, \ldots\}$. Wir setzen $\mathbb{N}_0 := \{0, 1, 2 \ldots\} = \mathbb{N} \cup \{0\}$.[14] Achtung: Bei manchen Autoren ist $0 \in \mathbb{N}$.

(b) Die Menge der *ganzen Zahlen* $\mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Es gilt $\mathbb{N} = \{n \in \mathbb{Z} : n > 0\}$. Die ganzen Zahlen der Form $2n$ (bzw. $2n + 1$) mit $n \in \mathbb{Z}$ heißen *gerade* (bzw. *ungerade*).

(c) Die Menge der *rationalen Zahlen* $\mathbb{Q} := \{\frac{a}{b} : a, b \in \mathbb{Z},\ b \neq 0\}$.

[10] Man beachte die Ähnlichkeit der Symbole $\cup$ und $\vee$ sowie $\cap$ und $\wedge$.
[11] In manchen Bücher schreibt man $A - B$ anstatt $A \setminus B$.
[12] Das Symbol $\subset$ wird in der Literatur leider nicht einheitlich benutzt.
[13] siehe `https://de.wikipedia.org/wiki/Mengendiagramm`
[14] Streng axiomatisch definiert man $0 := \varnothing$, $1 := \{\varnothing\}$ und allgemein $n + 1 := n \cup \{n\}$.

(d) Die Menge der *reellen Zahlen* $\mathbb{R}$ besteht aus allen *Dezimalbrüchen* wie $2 = 2.0$, $\frac{1}{3} = 0.33\ldots$, $\sqrt{2} = 1.4142\ldots$ oder $\pi = 3.1415\ldots$ (die Dezimalbruchentwicklung kann abbrechend, periodisch oder unperiodisch sein). In der Analysis definiert man reelle Zahlen als Grenzwerte von rationalen CAUCHY-Folgen. Im Folgenden setzen wir die üblichen Regeln für die Grundrechenarten voraus. Auch diese lassen sich streng axiomatisch einführen.

(e) Es gilt $\mathbb{N} \subsetneq \mathbb{N}_0 \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$. Die Behauptung $\mathbb{Q} \neq \mathbb{R}$ zeigen wir indirekt. Annahme: $\mathbb{Q} = \mathbb{R}$. Dann ist $\sqrt{2} \in \mathbb{Q}$ und es existieren $a, b \in \mathbb{Z}$ mit $\sqrt{2} = \frac{a}{b}$ und $b \neq 0$. Ohne Beschränkung der Allgemeinheit (kurz: o. B. d. A.) können wir annehmen, dass a und b teilerfremd sind (anderenfalls kann man $\frac{a}{b}$ kürzen). Umstellen ergibt $2b^2 = a^2$. Insbesondere ist a^2 gerade. Da das Quadrat einer ungeraden Zahl ungerade ist ($(2n+1)^2 = 2(2n^2+2n)+1$), ist a gerade, sagen wir $a = 2c$. Es folgt $b^2 = 2c^2$. Mit dem gleichen Argument ist nun auch b gerade. Also ist 2 ein gemeinsamer Teiler von a und b. Dieser Widerspruch zeigt, dass die Annahme falsch war. Also ist $\mathbb{Q} \neq \mathbb{R}$.

(f) Die Elemente einer Menge können durchaus selbst Mengen sein. In solchen Fällen benutzt man oft geschwungene Buchstaben. Zum Beispiel besteht $\mathcal{M} := \{\{1,2\}, \{1,3\}, \{2,3\}\}$ aus allen 2-elementigen Teilmengen von $\{1,2,3\}$.

Lemma 1.12. *Für Mengen A, B und C gilt:*

(a) $A \cap B \subseteq A \subseteq A \cup B$.

(b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *und* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributivgesetz).

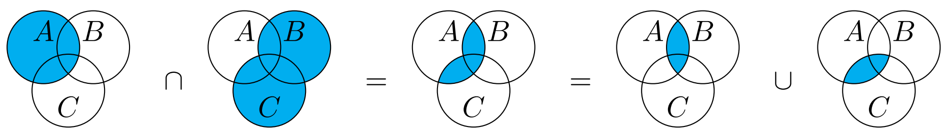
(c) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$ *und* $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$ (De Morgansche Regeln).

(d) $\boxed{|A \cup B| + |A \cap B| = |A| + |B|}$ *und* $|A \dot\cup B| = |A| + |B|$.

Beweis.

(a) Folgt direkt aus der Definition.

(b) Wir beweisen nur die erste Gleichheit (beweisen Sie die zweite selbst):



(c) Diesmal benutzen wir Lemma 1.5 (für die erste Gleichung):

$$x \in A \setminus (B \cup C) \iff (x \in A \wedge (x \notin B \cup C)) \iff (x \in A \wedge (x \notin B \wedge x \notin C))$$
$$\iff ((x \in A \wedge x \notin B) \wedge (x \in A \wedge x \notin C)) \iff x \in (A \setminus B) \cap (A \setminus C).$$

(d) Ist A oder B unendlich, so auch $A \cup B$ und die Behauptung gilt, wenn man $\infty + n = \infty$ für $n \in \mathbb{N}_0 \cup \{\infty\}$ interpretiert. Seien nun A und B endlich, sagen wir $A \cap B = \{x_1, \ldots, x_s\}$, $A = \{x_1, \ldots, x_s, a_1, \ldots, a_t\}$ und $B = \{x_1, \ldots, x_s, b_1, \ldots, b_u\}$. Dann gilt

$$|A \cup B| + |A \cap B| = s + t + u + s = |A| + |B|.$$

Sind A und B disjunkt, so gilt $|A \cap B| = |\varnothing| = 0$ und die zweite Behauptung folgt. $\square$

Definition 1.13. Die *Potenzmenge* $\mathcal{P}(M)$ einer Menge M ist die Menge aller Teilmengen von M, d. h.

$$\mathcal{P}(M) := \{N : N \subseteq M\}.$$

1.3 Vollständige Induktion

Satz 1.14 (Prinzip der vollständigen Induktion)**.** *Sei $A(n)$ ein Prädikat für $n \in \mathbb{N}$ mit den Eigenschaften:*

- *Induktionsanfang: $A(1)$ gilt.*
- *Induktionsschritt: $\forall n \in \mathbb{N} : (A(n) \implies A(n+1))$.*

Dann gilt $A(n)$ für alle $n \in \mathbb{N}$.

Beweis. Beweis durch Widerspruch: Gilt $A(n)$ nicht für alle $n \in \mathbb{N}$, so gibt es ein kleinstes n mit $\neg A(n)$. Nach dem Induktionsanfang ist $n \neq 1$. Nach Wahl von n gilt $A(n-1)$. Nach dem Induktionsschritt gilt $A(n-1) \Rightarrow A(n)$. Also gilt $A(n)$ nach Modus ponens. Widerspruch. $\square$

Bemerkung 1.15. Man verwendet oft Varianten der vollständigen Induktion. Zum Beispiel:

- Induktionsanfang: $A(1) \wedge A(2)$ gilt.
- Induktionsschritt: $\forall n \in \mathbb{N} : \big((A(n) \wedge A(n+1)) \Rightarrow A(n+2)\big)$.

Beispiel 1.16. Wir beweisen $(1+2+\ldots+n)^2 = 1^3 + 2^3 + \ldots + n^3$ für alle $n \in \mathbb{N}$.

Induktionsanfang: Für $n = 1$ gilt $1^2 = 1 = 1^3$.

Induktionsvoraussetzung: Es gelte bereits $(1+2+\ldots+n)^2 = 1^3 + 2^3 + \ldots + n^3$ $(*)$.

Induktionsschritt: Wir müssen die Behauptung für $n+1$ beweisen. Zunächst eine Nebenrechnung:

$$\begin{aligned} 2(1+2+\ldots+n) &= (1+2+\ldots+n) + (n+(n-1)+\ldots+1) \\ &= (1+n) + (2+n-1) + \ldots + (n+1) = n(n+1) \end{aligned}$$

(das hat Gauss als 9-Jähriger erkannt[15]). Nach der binomischen Formel gilt nun

$$\begin{aligned} ((1+2+\ldots+n)+(n+1))^2 &= (1+2+\ldots+n)^2 + 2(1+2+\ldots+n)(n+1) + (n+1)^2 \\ &\overset{(*)}{=} 1^3 + 2^3 + \ldots + n^3 + n(n+1)(n+1) + (n+1)^2 \\ &= 1^3 + 2^3 + \ldots + n^3 + (n+1)^3. \end{aligned}$$

$\square$

[15] siehe `https://de.wikipedia.org/wiki/Gau%C3%9Fsche_Summenformel#Herkunft_der_Bezeichnung`

2 Kartesische Produkte und Funktionen

2.1 Paare und Tupel

Bemerkung 2.1. Nach Bemerkung 1.8 sind die Elemente einer Menge ungeordnet. Wir führen eine geordnete Variante ein.

Definition 2.2.

- Seien A und B Mengen. Das *kartesische Produkt* von A und B ist die Menge $A \times B$ bestehend aus allen (*geordneten*) *Paaren*[1] (a, b) mit $a \in A$, $b \in B$, sodass gilt

$$(a, b) = (a', b') \iff (a = a' \wedge b = b').$$

  Es gilt $\boxed{|A \times B| = |A||B|,}$ sofern man die Regeln $\infty \cdot 0 = 0$ und $\infty \cdot n = \infty$ für $n \in \mathbb{N} \cup \{\infty\}$ benutzt.

- Analog definiert man *Tripel* (a, b, c) und *n-Tupel* $(a_1, \ldots, a_n)$ für $n \geq 2$. Für Mengen $A_1, \ldots, A_n$ setzt man

$$A_1 \times \ldots \times A_n := \big\{(a_1, \ldots, a_n) : a_1 \in A_1, \ldots, a_n \in A_n\big\}.$$

  Gilt $A := A_1 = \ldots = A_n$, so benutzt man die Abkürzung $A^n := A_1 \times \ldots \times A_n$.

- Kartesische Produkte lassen sich auch für beliebige Familien von Mengen definieren. Sei I eine Indexmenge und $(A_i : i \in I)$ eine Familie von Mengen. Man definiert $\bigtimes_{i \in I} A_i := \{(a_i)_{i \in I} : \forall i \in I : a_i \in A_i\}$.

Beispiel 2.3.

(a) Das kartesische Produkt $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ besteht aus allen Koordinaten in der 2-dimensionalen Ebene.

(b) Es gilt

$$\{1, 2\} \times \{2, 3, 4\} = \big\{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4)\big\}.$$

(c) Das kartesische Produkt $\bigtimes_{i \in \mathbb{N}} \mathbb{R}$ ist die Menge aller reellen Folgen aus der Analysis.

(d) Sei $(A_i : i \in I)$ eine beliebige Familie von nicht-leeren Mengen. Das bereits erwähnte *Auswahlaxiom* besagt $\bigtimes_{i \in I} A_i \neq \varnothing$, d. h. man kann aus jeder Menge A_i *gleichzeitig* ein Element auswählen.

Definition 2.4. Eine *Relation* auf einer nicht-leeren Menge A ist eine Teilmenge $R \subseteq A \times A$. Üblicherweise wählt man ein Symbol, zum Beispiel $\sim$, und schreibt $a \sim b$, falls $(a, b) \in R$. Man nennt R

- *reflexiv*, falls $\forall a \in A : a \sim a$.
- *symmetrisch*, falls $\forall a, b \in A : (a \sim b \Rightarrow b \sim a)$.

[1] Die formale Definition von Paaren kann man auf Mengen zurückführen: $(a, b) := \{\{a\}, \{a, b\}\}$.

- *antisymmetrisch*, falls $\forall a, b \in A : ((a \sim b \wedge b \sim a) \Rightarrow a = b)$.[2]
- *transitiv*, falls $\forall a, b, c \in A : ((a \sim b \wedge b \sim c) \Rightarrow a \sim c)$.
- *Äquivalenzrelation*, falls R reflexiv, symmetrisch und transitiv ist.
- (*partielle*) *Ordnungsrelation*, falls R reflexiv, antisymmetrisch und transitiv ist.

Ist R eine Äquivalenzrelation auf der Menge A und $a \in A$, so nennt man $[a] := \{b \in A : a \sim b\} \subseteq A$ die *Äquivalenzklasse* von a.

Beispiel 2.5.

(a) Die *triviale* Relation $R = A \times A$ ist eine (uninteressante) Äquivalenzrelation.

(b) Die Gleichheitsrelation $\{(a, a) : a \in A\}$ mit dem Symbol $=$ ist die „kleinste" reflexive Relation auf A. Trivialerweise handelt es sich um eine Äquivalenzrelation.[3] Man kann viele weitere Äquivalenzrelationen auf die Gleichheit zurückführen. Sei beispielsweise A die Menge aller Menschen und $a \sim b$ falls $a, b \in A$ im gleichen Land leben. Die Äquivalenzklassen entsprechen dann den Ländern.

(c) Man kann durch einfache Beispiele zeigen, dass die Eigenschaften reflexiv, symmetrisch und transitiv unabhängig voneinander sind. Zum Beispiel ist die Relation

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

auf $A = \{1, 2, 3\}$ reflexiv und symmetrisch, aber nicht transitiv.

(d) Auf $\mathbb{R}$ ist die Kleinergleichrelation $\leq$ eine Ordnungsrelation. Sie hat zusätzlich die Eigenschaft, dass je zwei Zahlen a und b in Relation stehen, d. h. es gilt $a \leq b$ oder $b \leq a$ (man spricht von einer *totalen* Ordnungsrelation).

(e) Auf der Potenzmenge jeder Menge A ist die Inklusionsrelation $\subseteq$ eine Ordnungsrelation. Im Fall $A = \mathbb{N}$ stehen $\{1\}$ und $\{2\}$ nicht in Relation ($\{1\} \nsubseteq \{2\} \nsubseteq \{1\}$). Im Gegensatz zu $\leq$ ist $\subseteq$ also nicht total.

Lemma 2.6. *Sei R eine Äquivalenzrelation auf einer Menge A. Dann existiert eine Teilmenge $T \subseteq A$, sodass die Äquivalenzklassen $[t]$ mit $t \in T$ eine Partition von A bilden, d. h. $A = \dot{\bigcup}_{t \in T} [t]$.*

Beweis. Sei $\sim$ das Symbol von R. Seien $a, b \in A$ und $c \in [a] \cap [b]$. Dann gilt $a \sim c$ und $b \sim c$. Da $\sim$ symmetrisch ist, gilt $c \sim b$. Da $\sim$ transitiv ist, gilt $a \sim b$. Für jedes $d \in [b]$ gilt also $a \sim b \sim d$ und $a \sim d$. Dies zeigt $[b] \subseteq [a]$ und analog erhält man $[a] \subseteq [b]$. Es folgt $[a] = [b]$. Somit sind je zwei Äquivalenzklassen entweder gleich oder disjunkt. Die Existenz von T folgt nun aus dem Auswahlaxiom. □

Bemerkung 2.7.

(a) In der Situation von Lemma 2.6 nennt man T ein *Repräsentantensystem* für die Äquivalenzklassen.

(b) Ist $A = \dot{\bigcup}_{i \in I} A_i$ eine Partition von A, so definiert

$$a \sim b \;:\Longleftrightarrow\; \exists i \in I : a, b \in A_i$$

eine Äquivalenzrelation auf A (nachrechnen). Daher entsprechen sich Partitionen und Äquivalenzrelationen.

[2] Achtung: Es gibt auch die stärkere Eigenschaft *asymmetrisch*: $\forall a, b \in A : (a \sim b \Rightarrow b \not\sim a)$.

[3] Für symmetrische Relationen sollte man „symmetrische" Symbole wählen. Die Umkehrung ist leider nicht immer gegeben, z. B. das Symbol $\mid$ für die antisymmetrische Teilbarkeitsrelation auf $\mathbb{N}$.

Beispiel 2.8. Für die Äquivalenzklasse auf der Menge aller Menschen aus Beispiel 2.5 bilden die Präsidenten jedes Landes ein Repräsentantensystem.

2.2 Injektive und surjektive Funktionen

Definition 2.9.

- Seien A und B Mengen. Eine *Funktion* oder *Abbildung* f von A nach B ist eine Vorschrift, die jedem $a \in A$ genau ein $f(a) \in B$ zuordnet.[4] Man schreibt dann[5]

$$f\colon A \to B, \qquad a \mapsto f(a).$$

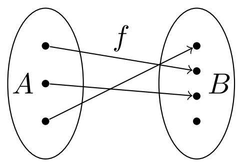
  Die Menge aller Abbildungen $A \to B$ bezeichnen wir mit $\mathrm{Abb}(A, B)$.

- Man nennt A den *Definitionsbereich* und B den *Wertebereich* von f. Außerdem ist $f(a)$ das *Bild* von a unter f und $f(A) := \{f(a) : a \in A\} \subseteq B$ ist das *Bild* von f. Für $B' \subseteq B$ ist

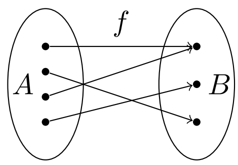
$$f^{-1}(B') := \{a \in A : f(a) \in B'\} \subseteq A$$

  das *Urbild* von B' unter f.

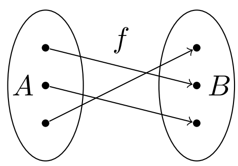
- Man nennt $f\colon A \to B$
  - *injektiv*, falls $\forall a, a' \in A : (f(a) = f(a') \Longrightarrow a = a')$.



  - *surjektiv*, falls $\forall b \in B : \exists a \in A : f(a) = b$, d. h. $f(A) = B$.



  - *bijektiv* (oder *Bijektion*), falls f injektiv und surjektiv ist. Ggf. nennt man A und B *gleichmächtig*.



- Die *Einschränkung* von $f\colon A \to B$ auf eine Teilmenge $A' \subseteq A$ ist die Funktion

$$f_{|A'}\colon A' \to B, \qquad a \mapsto f(a).$$

  Für eine weitere Funktion $g\colon B \to C$ nennt man die Abbildung

$$g \circ f\colon A \to C, \qquad a \mapsto g(f(a))$$

  die *Komposition* (oder *Hintereinanderausführung*, *Verkettung*) von f und g.

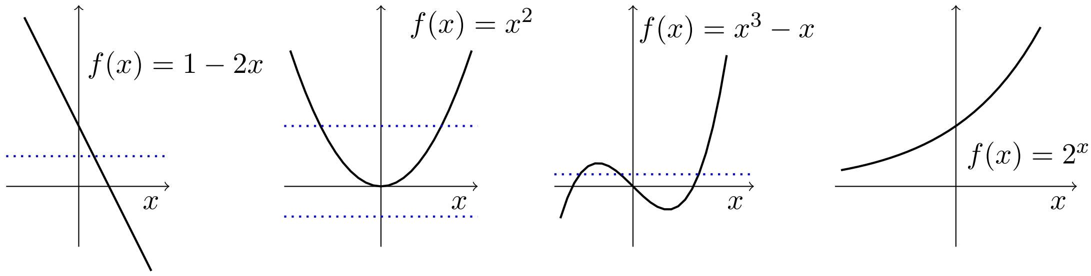
---

[4] Formal: Eine Funktion ist eine Teilmenge $f \subseteq A \times B$, sodass für jedes $a \in A$ genau ein $b \in B$ mit $(a, b) \in f$ existiert.
[5] Man beachte die unterschiedlichen Pfeile $\to$ und $\mapsto$.

Beispiel 2.10.

(a) Für jede Menge A und $B \subseteq A$ ist $f\colon B \to A$, $b \mapsto b$ eine injektive Funktion, die man *Inklusionsabbildung* nennt. Im Fall $B = A$ ist f sogar bijektiv und man nennt $f = \mathrm{id}_A$ die *Identität* auf A.

(b) Abbildungen $f\colon \mathbb{R} \to \mathbb{R}$ lassen sich grafisch darstellen:



Injektiv (bzw. surjektiv) bedeutet, dass der Graph von f jede horizontale Gerade höchstens (bzw. mindestens) einmal schneidet. Wir lesen ab:

Funktion	injektiv	surjektiv	bijektiv
$f(x) = 1 - 2x$	✓	✓	✓
$f(x) = x^2$	✗	✗	✗
$f(x) = x^3 - x$	✗	✓	✗
$f(x) = 2^x$	✓	✗	✗

Bemerkung 2.11.

(a) Sind A und B endliche Mengen, so gilt $|\mathrm{Abb}(A, B)| = |B|^{|A|}$, denn für $f\colon A \to B$ und jedes $a \in A$ hat man $|B|$ Möglichkeiten $f(a) \in B$ zu wählen.

(b) Achtung: Injektiv ist nicht das Gegenteil von surjektiv (ein häufiger Anfängerfehler)!

(c) Man kann jede Funktion $f\colon A \to B$ surjektiv machen, indem man auf den Wertebereich auf das Bild einschränkt: $f\colon A \to f(A)$.

(d) Für $f\colon A \to B$ gilt

$$\begin{aligned} f \text{ injektiv} &\implies |A| = |f(A)| \le |B| \\ f \text{ surjektiv} &\implies |B| = |f(A)| \le |A| \\ f \text{ bijektiv} &\implies |A| = |B| \end{aligned}$$

(wobei $\infty \le \infty$).

(e) Zwei endliche Mengen A und B sind genau dann gleichmächtig, wenn $|A| = |B|$. In diesem Fall sind die Eigenschaften injektiv, surjektiv und bijektiv nach (d) äquivalent. Für unendliche Mengen ist dies im Allgemeinen falsch (Beispiel 2.10).

(f) Obwohl $\mathbb{N}$ nur „halb so viele“ Zahlen wie $\mathbb{Z}$ enthält, sind $\mathbb{N}$ und $\mathbb{Z}$ durch die Bijektion

$$\mathbb{N} \to \mathbb{Z}, \qquad n \mapsto \begin{cases} \frac{n-1}{2} & \text{falls } n \text{ ungerade,} \\ -\frac{n}{2} & \text{falls } n \text{ gerade} \end{cases}$$

gleichmächtig. Eine Menge, die gleichmächtig zu $\mathbb{N}$ ist, nennt man *abzählbar*[6]. Nach Cantors *Diagonalisierungsargumenten* ist $\mathbb{Q}$ abzählbar, aber $\mathbb{R}$ nicht, d. h. $\mathbb{R}$ ist *überabzählbar*. Die (nicht beweisbare) *Kontinuumshypothese* besagt, dass jede unendliche Teilmenge von $\mathbb{R}$ entweder zu $\mathbb{N}$ oder zu $\mathbb{R}$ gleichmächtig ist. Der folgende Satz zeigt, dass es beliebig „große" Mengen gibt (*Kardinalzahlen*), die man in der Praxis aber selten antrifft.

Satz 2.12 (Cantor). *Jede Menge M ist „kleiner" als ihre Potenzmenge, d. h. es existiert eine injektive Abbildung $M \to \mathcal{P}(M)$, aber keine Bijektion. Ist M endlich, so gilt $|\mathcal{P}(M)| = 2^{|M|}$.*

Beweis. Die Abbildung $M \to \mathcal{P}(M)$, $a \mapsto \{a\}$ ist sicher injektiv. Nehmen wir an es existiert eine Bijektion $f\colon M \to \mathcal{P}(M)$. Sei

$$A := \{x \in M : x \notin f(x)\} \in \mathcal{P}(M).$$

Dann existiert ein $a \in M$ mit $f(a) = A$. Es folgt der Widerspruch $a \in A = f(a) \Longleftrightarrow a \notin f(a)$. Für die zweite Behauptung sei $M = \{x_1, \ldots, x_n\}$. Dann ist die Abbildung

$$\mathcal{P}(M) \to \{0,1\}^n, \qquad A \mapsto (a_1, \ldots, a_n)$$

mit $a_i = 1 \Longleftrightarrow x_i \in A$ eine Bijektion. Also ist $|\mathcal{P}(M)| = |\{0,1\}^n| = 2^n = 2^{|M|}$. □

Lemma 2.13. *Seien $f\colon A \to B$, $g\colon B \to C$, $h\colon C \to D$ Funktionen. Dann gilt:*

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

(b) Sind f und g injektiv, so auch $g \circ f$.

(c) Sind f und g surjektiv, so auch $g \circ f$.

(d) Ist $g \circ f$ injektiv, so auch f.

(e) Ist $g \circ f$ surjektiv, so auch g.

(f) Genau dann ist f bijektiv, wenn eine Funktion $g\colon B \to A$ mit $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$ existiert. Ggf. ist g eindeutig bestimmt und man nennt $f^{-1} := g$ die Umkehrfunktion *von f.*

Beweis.

(a) Für $a \in A$ ist $((h \circ g) \circ f)(a) = (h \circ g)(f(a)) = h(g(f(a))) = h((g \circ f)(a)) = (h \circ (g \circ f))(a)$.

(b) Für $a, a' \in A$ mit $(g \circ f)(a) = (g \circ f)(a')$ gilt $g(f(a)) = g(f(a'))$, also $f(a) = f(a')$ und $a = a'$.

(c) Es gilt $(g \circ f)(A) = g(f(A)) = g(B) = C$.

(d) Sei $f(a) = f(a')$ für $a, a' \in A$. Dann ist $(g \circ f)(a) = g(f(a)) = g(f(a')) = (g \circ f)(a')$. Da $g \circ f$ injektiv ist, folgt $a = a'$.

(e) Es gilt $C = (g \circ f)(A) = g(f(A)) \subseteq g(B) \subseteq C$, also $g(B) = C$.

(f) Ist $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$, so ist f injektiv nach (d) und surjektiv nach (e), also auch bijektiv. Sei umgekehrt f bijektiv. Für jedes $b \in B$ existiert dann genau ein $g(b) \in A$ mit $f(g(b)) = b$. Daher ist $g\colon B \to A$ die einzige Abbildung mit $f \circ g = \mathrm{id}_B$. Aus $f(a) = f(g(f(a)))$ folgt $g(f(a)) = a$ für alle $a \in A$, da f injektiv ist. Dies zeigt $g \circ f = \mathrm{id}_A$. □

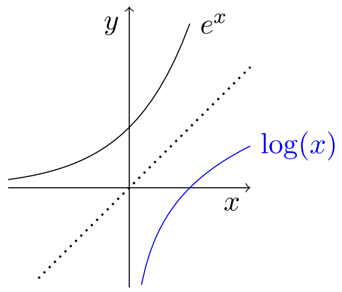
[6] In manchen Büchern zählen die endlichen Mengen auch zu den abzählbaren Mengen.

Bemerkung 2.14. Verwechseln Sie die Umkehrfunktion nicht mit dem Urbild. Der Zusammenhang beider Konzepte ist $f^{-1}(\{b\}) = \{f^{-1}(b)\}$ für jede Bijektion $f\colon A \to B$ und $b \in B$.

Beispiel 2.15.

(a) Die Abbildung $f\colon \mathbb{Q} \to \mathbb{Q}$, $x \mapsto 2x + 1$ ist eine Bijektion mit Umkehrabbildung $f^{-1}\colon \mathbb{Q} \to \mathbb{Q}$, $x \mapsto \frac{x-1}{2}$ (nachrechnen).

(b) Die Umkehrabbildung der *Exponentialfunktion* $\exp\colon \mathbb{R} \to \mathbb{R}_{>0}$, $x \mapsto e^x$ ist der *natürliche Logarithmus* $\log\colon \mathbb{R}_{>0} \to \mathbb{R}$. Man erhält den Graphen von log durch Spiegelung an der Geraden $y = x$:



Man beachte, dass die reine Existenz der Umkehrfunktion noch lange keine konkrete Formel für $f^{-1}(x)$ liefert. Diesen Umstand macht man sich in der Kryptographie zunutze (*Einwegfunktion*).

3 Körper und Vektorräume

3.1 Gruppen und Körper

Bemerkung 3.1. In fast allen Anwendungen der linearen Algebra wird nur von den vier Grundrechenarten (Addition, Subtraktion, Multiplikation und Division) Gebrauch gemacht. Damit man nicht jede Aussage für jeden Zahlbereich ($\mathbb{Q}, \mathbb{R}, \ldots$) neu beweisen muss, ersetzt man Zahlbereiche durch abstrakte *Gruppen* (mit einer Operation) und *Körper* (mit zwei Operationen). Zur Beschreibung von Lösungsmengen von Gleichungssystemen führt man *Vektorräume* ein. Beachten Sie, dass dies lediglich Modelle zur Untersuchung linearer Probleme sind, die sich im Laufe der Zeit bewährt haben (so wie metrische Räume in der Analysis oder das Bohrsche Atommodell in der Chemie).

Definition 3.2. Eine *Verknüpfung* $\cdot$ auf einer Menge G ist eine Abbildung $G \times G \to G$, $(x, y) \mapsto x \cdot y$. Man nennt das Paar $(G, \cdot)$ (oder auch nur G) eine *Gruppe*, falls

- $\forall x, y, z \in G : (x \cdot y) \cdot z = x \cdot (y \cdot z)$ (*Assoziativgesetz*),
- $\exists e \in G : (\forall x \in G : e \cdot x = x = x \cdot e)$ (*neutrales Element*),
- $\forall x \in G : (\exists y \in G : y \cdot x = e = x \cdot y)$ (*inverses Element*).

Gilt zusätzlich

- $\forall x, y \in G : x \cdot y = y \cdot x$ (*Kommutativgesetz*),

so heißt G *abelsch*.

Bemerkung 3.3. Sei G eine Gruppe mit neutralem Element e.

(a) Aus Bequemlichkeit schreiben wir oft xy anstatt $x \cdot y$.

(b) Ist auch $e' \in G$ ein neutrales Element, so gilt $e' = e' \cdot e = e$. Also ist e eindeutig bestimmt und wir schreiben oft $e = 1_G = 1$ oder $e = 0_G = 0$, falls die Verknüpfung $+$ ist.

(c) Seien $y, y' \in G$ invers zu $x \in G$. Dann ist

$$y' = y'e = y'(xy) = (y'x)y = ey = y.$$

Somit hat x genau ein Inverses und wir schreiben $y = x^{-1}$ oder $y = -x$, falls die Verknüpfung $+$ ist. Im letzten Fall schreiben wir $x - y := x + (-y)$ für beliebige $x, y \in G$.[1]

(d) Für $x, y \in G$ ist $\boxed{(x^{-1})^{-1} = x}$ und $\boxed{(xy)^{-1} = y^{-1}x^{-1}}$ (Achtung Reihenfolge!).

Beispiel 3.4.

(a) Wegen $e \in G$ ist eine Gruppe niemals leer. Andererseits gibt es die *triviale* Gruppe $G = \{e\}$.

[1] In nicht-abelschen Gruppen ist die Schreibweise $\frac{x}{y}$ problematisch, denn es könnte sowohl xy^{-1} als auch $y^{-1}x$ gemeint sein.

(b) Nach den üblichen Rechenregeln sind $(\mathbb{Z},+)$, $(\mathbb{Q},+)$ und $(\mathbb{R},+)$ abelsche Gruppen mit neutralem Element 0. Andererseits ist $(\mathbb{Z},-)$ *keine* Gruppe, denn das Assoziativgesetz ist verletzt:

$$(1-2)-3=-4\neq 2=1-(2-3).$$

Ebenso besitzt $(\mathbb{N},+)$ kein neutrales Element und in $(\mathbb{N}_0,+)$ hat nicht jedes Element ein Inverses (z. B. $-1\notin\mathbb{N}_0$).

(c) Offenbar sind $(\mathbb{Q}\setminus\{0\},\cdot)$ und $(\mathbb{R}\setminus\{0\},\cdot)$ abelsche Gruppen mit neutralem Element 1, aber nicht $(\mathbb{Z}\setminus\{0\},\cdot)$, denn $2^{-1}=\frac{1}{2}\notin\mathbb{Z}$.

(d) Für Gruppen $G_1,\ldots,G_n$ ist auch $G_1\times\ldots\times G_n$ eine Gruppe mit

$$(x_1,\ldots,x_n)\cdot(y_1,\ldots,y_n):=(x_1y_1,\ldots,x_ny_n)$$

für $(x_1,\ldots,x_n),(y_1,\ldots,y_n)\in G_1\times\ldots\times G_n$ (Aufgabe 8). Das neutrale Element ist $(1_{G_1},\ldots,1_{G_n})$. Man spricht dann vom *direkten Produkt* von $G_1,\ldots,G_n$ (anstelle vom kartesischen Produkt).

Definition 3.5. Ein *Körper* ist eine Menge K mit Verknüpfungen $+$ und $\cdot$, sodass folgende Eigenschaften gelten:

- $(K,+)$ ist eine abelsche Gruppe mit neutralem Element 0.
- $(K\setminus\{0\},\cdot)$ ist eine abelsche Gruppe mit neutralem Element 1. Man setzt $K^\times:=K\setminus\{0\}$.
- $\forall x,y,z\in K: x\cdot(y+z)=(x\cdot y)+(x\cdot z)$ (*Distributivgesetz*).

Bemerkung 3.6. Im Folgenden sei K stets ein Körper.

(a) Durch die Vereinbarung „Punktrechnung vor Strichrechnung“ sparen wir Klammern ein. Zum Beispiel sei $xy+z:=(x\cdot y)+z$ für $x,y,z\in K$.

(b) Für alle $x\in K$ gilt $\boxed{x\cdot 0=0=0\cdot x,}$ denn $x0=x(0+0)=x0+x0$. Es folgt $(-x)y=-(xy)$ für $x,y\in K$.

(c) Für $x,y,z\in K$ und $z\neq 0$ gilt die *Kürzungsregel* $xz=yz\Longrightarrow x=y$, denn

$$x=x\cdot 1=x(zz^{-1})=(xz)z^{-1}=(yz)z^{-1}=\ldots=y.$$

Beispiel 3.7.

(a) Nach den gewohnten Rechenregeln sind $\mathbb{Q}$ und $\mathbb{R}$ Körper. Es gibt außerdem unendlich viele Körper „zwischen“ $\mathbb{Q}$ und $\mathbb{R}$ (vgl. Aufgabe 14). Andererseits ist $(\mathbb{Z},+,\cdot)$ kein Körper, da $(\mathbb{Z}\setminus\{0\},\cdot)$ keine Gruppe ist.

(b) Jeder Körper besitzt mindestens die beiden Elemente 0 und 1. Tatsächlich ist $\mathbb{F}_2=\{0,1\}$ bereits ein Körper, wenn man $1+1:=0$ definiert. Die Verknüpfungstabellen sind dadurch vollständig bestimmt:

$+$	0	1
0	0	1
1	1	0

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

Auf Computern werden alle Rechnungen in $\mathbb{F}_2$ durchgeführt, indem man 0 und 1 als *Bits* interpretiert. In der Algebra[2] konstruiert man für jede Primzahlpotenz q ein Körper mit genau q Elementen (vgl. Aufgabe 9).

[2] siehe Algebra-Skript

3.2 Vektorräume und Unterräume

Definition 3.8. Ein *Vektorraum* V über einem Körper K (kurz: K-Vektorraum) ist eine abelsche Gruppe bzgl. $+$ zusammen mit einer *Skalarmultiplikation* $K \times V \to V$, $(\lambda, v) \mapsto \lambda \cdot v$ mit folgenden Eigenschaften:

- $\forall v \in V : 1 \cdot v = v$,
- $\forall v, w \in V,\ \lambda \in K : \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$,
- $\forall v \in V,\ \lambda, \mu \in K : (\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$,
- $\forall v \in V,\ \lambda, \mu \in K : (\lambda \cdot \mu) \cdot v = \lambda \cdot (\mu \cdot v)$.

Die Elemente von V heißen *Vektoren* und die Elemente in K *Skalare* (in diesem Kontext). Das neutrale Element 0_V in V nennt man den *Nullvektor*.

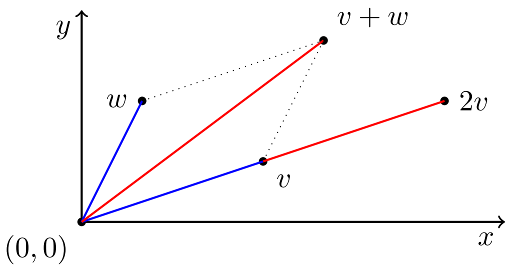
Bemerkung 3.9. Man beachte, dass $+$ sowohl die Addition in K als auch in V bezeichnet. Ebenso steht $\cdot$ für die Multiplikation in K und für die Skalarmultiplikation (das ist ungenau, aber durchaus üblich). Wir werden in beiden Fällen das Symbol $\cdot$ oft einsparen. Falls Missverständnisse ausgeschlossen sind, schreiben wir auch 0 anstatt 0_V. Sie müssen im Zweifel in der Lage sein zu entscheiden, ob das Nullelement in K oder V gemeint ist.

Beispiel 3.10.

(a) Der *Nullraum* $V = \{0_V\}$ mit der Skalarmultiplikation $\lambda \cdot 0_V := 0_V$ für alle $\lambda \in K$.

(b) Für K-Vektorräume $V_1, \ldots, V_n$ ist auch das direkte Produkt (bzgl. $+$) $V_1 \times \ldots \times V_n$ ein Vektorraum mit komponentenweiser Skalarmultiplikation: $\lambda(v_1, \ldots, v_n) := (\lambda v_1, \ldots, \lambda v_n)$ für $v_i \in V_i$ und $\lambda \in K$ (nachrechnen).

(c) Offenbar ist K selbst ein Vektorraum, in dem die Skalarmultiplikation mit der gewöhnlichen Multiplikation übereinstimmt. Nach (b) ist auch K^n für $n \geq 1$ ein Vektorraum. In $\mathbb{R}^2$ lassen sich Vektoraddition und Skalarmultiplikation geometrisch deuten:



(d) Sind $v_1, \ldots, v_n$ Vektoren aus V und $\lambda_1, \ldots, \lambda_n \in K$, so liegt auch die *Linearkombination* $\lambda_1 v_1 + \ldots + \lambda_n v_n$ in V (Nachweis durch Induktion nach n). Man benutzt dafür das Summenzeichen:

$$\boxed{\sum_{i=1}^{n} \lambda_i v_i := \lambda_1 v_1 + \ldots + \lambda_n v_n.}$$

Sind $v_1, \ldots, v_n$ paarweise verschieden (d. h. $v_i \neq v_j$ für $i \neq j$)[3] und mindestens ein $\lambda_i \neq 0$, so nennt man die Linearkombination *nicht-trivial*. Manchmal tritt die *leere Summe* ohne

[3] „paarweise verschieden" ist stärker als die Formulierung „nicht alle sind gleich"

Summanden auf. Diese wird stets als 0_V interpretiert. Zum Beispiel $\sum_{i=1}^{0} v_i = 0$. Sei auch $v_i = \mu_{i1}w_{i1} + \mu_{i2}w_{i2} + \ldots + \mu_{im}w_{im}$ eine Linearkombination für $i = 1, \ldots, n$.[4] Dann erhält man eine *Doppelsumme*:

$$\sum_{i=1}^{n} v_i = \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}w_{ij}.$$

Da $(V, +)$ abelsch ist, darf man die Summanden beliebig umordnen und somit die Summenzeichen vertauschen:

$$\boxed{\sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}w_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} \mu_{ij}w_{ij}.}$$

Ein Vorzug der Algebra gegenüber der Analysis ist, dass alle Summen endlich sind und man keine Konvergenzbetrachtungen anstellen muss.

Satz 3.11. *Sei $A \neq \varnothing$ eine Menge und V ein K-Vektorraum. Dann ist* $\mathrm{Abb}(A, V)$ *mit den folgenden Verknüpfungen ein K-Vektorraum:*

$$\begin{aligned} (f+g)(a) &:= f(a) + g(a) && (f, g \in \mathrm{Abb}(M, V),\ a \in M) \\ (\lambda f)(a) &:= \lambda f(a) && (\lambda \in K) \end{aligned}$$

Beweis. Offenbar liegen $f + g$ und λf in $\mathrm{Abb}(A, V)$. Die triviale Abbildung $f(a) = 0$ für alle $a \in A$ ist das neutrale Element bzgl. $+$. Für $f\colon A \to V$ ist $-f\colon A \to V$, $a \mapsto -f(a)$ invers zu f bzgl. $+$. Für $f, g, h\colon A \to V$ und $a \in A$ ist

$$\begin{aligned} ((f+g)+h)(a) &= (f+g)(a) + h(a) = (f(a) + g(a)) + h(a) = f(a) + (g(a) + h(a)) \\ &= f(a) + (g+h)(a) = (f + (g+h))(a). \end{aligned}$$

Daher ist $+$ assoziativ. Auf die gleiche Weise übertragen sich die verbleibenden Vektorraumaxiome von V nach $\mathrm{Abb}(A, V)$. □

Definition 3.12.

- Eine Teilmenge H einer Gruppe G heißt *Untergruppe*, falls H mit der eingeschränkten Verknüpfung selbst eine Gruppe ist, d. h.
  - $1_G \in H$,
  - $\forall g, h \in H : gh \in H$,
  - $\forall h \in H : h^{-1} \in H$.

  Wir schreiben ggf. $H \leq G$. Im Fall $H \neq G$ nennt man H eine *echte* Untergruppe und schreibt $H < G$.

- Eine Teilmenge U eines Vektorraums V heißt *Unterraum*, falls U mit den eingeschränkten Verknüpfungen selbst einen Vektorraum ist, d. h.
  - $(U, +)$ ist eine Untergruppe von $(V, +)$,
  - $\forall v \in U,\ \lambda \in K : \lambda v \in U$.

[4] Im Zweifel sollte man doppelte Indizes durch Komma trennen $\mu_{i,1}$.

Wir schreiben dann $U \le V$ wie bei Untergruppen. Im Fall $U \neq V$ nennt man U einen *echten* Unterraum und schreibt $U < V$.

Bemerkung 3.13.

(a) Die Bedingungen garantieren, dass H unter Multiplikation bzw. U unter Addition und Skalarmultiplikation *abgeschlossen* ist. Somit sind die Verknüpfungen auf H bzw. U *wohldefiniert*. Die verbleibenden Gruppenaxiome bzw. Vektorraumaxiome muss man nicht prüfen, da sie bereits in der größeren Menge G bzw. V gelten.

(b) Im Folgenden beschränken wir uns auf die Untersuchung von Unterräumen. Die meisten Aussagen gelten sinngemäß auch für (abelsche) Gruppen.

(c) Für Vektorräume kann man die Bedingungen wie folgt zusammenfassen: Eine *nichtleere* Teilmenge $U \subseteq V$ ist genau dann ein Unterraum, wenn für alle $u, v \in U$ und $\lambda \in K$ gilt: $\lambda u + v \in U$ (Aufgabe 10).

Beispiel 3.14.

(a) Jeder Vektorraum V besitzt die Unterräume $\{0_V\}$ und V.

(b) Aus $U \le W \le V$ folgt $U \le V$. Aus $U, W \le V$ und $U \subseteq W$ folgt sicher auch $U \le W$.

(c) Der Durchschnitt von beliebig vielen Unterräumen ist wieder ein Unterraum (nachrechnen).

(d) Wir beweisen $U := \big\{(x, 0) : x \in \mathbb{R}\big\} \le \mathbb{R}^2$ mit Hilfe von Bemerkung 3.13: Wegen $(0, 0) \in U$ ist $U \neq \varnothing$. Für $(x_1, 0), (x_2, 0) \in U$ und $\lambda \in \mathbb{R}$ gilt

$$\lambda(x_1, 0) + (x_2, 0) = (\lambda x_1, 0) + (x_2, 0) = (\lambda x_1 + x_2, 0) \in U.$$

Geometrisch entspricht U der x-Achse in der Ebene. Analog ist die xy-Ebene

$$U := \big\{(x, y, 0) \in \mathbb{R}^3 : x, y \in \mathbb{R}\big\}$$

ein Unterraum von $\mathbb{R}^3$.

(e) Die Teilmenge $U := \big\{(x, x^2) : x \in \mathbb{Q}\big\}$ von $\mathbb{Q}^2$ ist *kein* Unterraum, denn $(1, 1) \in U$, aber $2 \cdot (1, 1) = (2, 2) \notin U$. Wir zeigen in Bemerkung 7.19, dass sich jeder Unterraum durch *lineare* Gleichungen beschreiben lässt.

(f) Offenbar sind $U := \big\{(0, 0), (1, 0)\big\}$ und $W := \big\{(0, 0), (0, 1)\big\}$ Unterräume von $\mathbb{F}_2^2$, aber nicht $U \cup W$ (Warum?)

Lemma 3.15. *Sei V ein K-Vektorraum und $U \le V$. Für $v \in V$ sei $v + U := \{v + u : u \in U\} \subseteq V$. Dann wird $V/U := \{v + U : v \in V\}$ mit*

$$\begin{aligned} (v + U) + (w + U) &:= (v + w) + U \qquad (v, w \in V), \\ \lambda(v + U) &:= \lambda v + U \qquad (\lambda \in K) \end{aligned}$$

zu einem K-Vektorraum.

Beweis. Sei $\overline{v} := v + U$ für $v \in V$. Für $\overline{v} = \overline{v'}$ und $\overline{w} = \overline{w'}$ gilt

$$\overline{v + w} = v + w + U = v + w' + U = w' + v + U = w' + v' + U = \overline{v' + w'}.$$

Dies zeigt, dass die Addition auf V/U wohldefiniert ist. Das neutrale Element ist $0 + U = U$. Für $\lambda \in K$ gilt analog

$$\overline{\lambda v} = \lambda v + U = \lambda v + \lambda U = \lambda(v + U) = \lambda(v' + U) = \lambda v' + U = \overline{\lambda v'}.$$

Also ist auch die Skalarmultiplikation wohldefiniert. Die Vektorraumaxiome für V/U folgen unmittelbar aus den Axiomen für V. $\square$

Bemerkung 3.16. Man nennt V/U den *Faktorraum* von V nach U. Die Mengen $v+U$ werden manchmal als *affine Räume* bezeichnet. Sie sind die Äquivalenzklassen der Relation $v \sim w :\Longleftrightarrow v - w \in U$.

4 Basen und Dimension

4.1 Lineare Unabhängigkeit und Erzeugendensysteme

Bemerkung 4.1. Um unendlich große Vektorräume vergleichen zu können, führen wir die Dimension als feinere Kenngröße ein. Es wird sich zeigen, dass Vektorräume allein durch ihre Dimension weitestgehend bestimmt sind (Satz 7.10).

Definition 4.2. Sei V ein Vektorraum.

(a) Für $S \subseteq V$ sei $\langle S \rangle \subseteq V$ die Menge aller Linearkombinationen von Elementen aus S. Man nennt $\langle S \rangle$ den *Spann* von S.[1] Im Fall $S = \{s_1, \ldots, s_n\}$ schreiben wir auch $\langle s_1, \ldots, s_n \rangle$ anstatt $\langle S \rangle$ (d. h. wir sparen die Mengenklammern ein).

(b) Für Unterräume $U, W \leq V$ sei

$$U + W := \{u + w : u \in U, w \in W\} \subseteq V$$

die (Minkowski-)*Summe* von U und W. Im Fall $U \cap W = \{0\}$ nennt man die Summe *direkt* und schreibt $U \oplus W$ anstatt $U + W$.[2]

Lemma 4.3. *Sei V ein Vektorraum, $S \subseteq V$ und $U, W \leq V$. Dann sind $\langle S \rangle$ und $U + W$ Unterräume von V.*

Beweis. Offenbar ist 0 eine Linearkombination von Elementen aus S, d. h. $0 \in \langle S \rangle$ (im Fall $S = \varnothing$ wähle man die leere Summe). Addition und Skalarmultiplikation von Linearkombinationen sind wieder Linearkombinationen. Dies zeigt $\langle S \rangle \leq V$. Wegen $0 \in U \cap W$ ist $0 = 0 + 0 \in U + W$. Seien $u_1 + w_1, u_2 + w_2 \in U + W$ und $\lambda \in K$. Dann gilt

$$\lambda(u_1 + w_1) + (u_2 + w_2) = \underbrace{(\lambda u_1 + u_2)}_{\in U} + \underbrace{(\lambda w_1 + w_2)}_{\in W} \in U + W.$$

Also ist auch $U + W \leq V$. □

Beispiel 4.4.

(a) Es gilt $\langle \varnothing \rangle = \{0\}$, denn die leere Summe ist die einzige Linearkombination aus $\varnothing$.

(b) Für $U \leq W \leq V$ gilt $U + W = W$ und $\langle U \rangle = U = U \oplus \{0\}$.

(c) Für $s_1, \ldots, s_n \in V$ gilt $\langle s_i \rangle = \{\lambda s_i : \lambda \in K\} =: K s_i$ und $\langle s_1, \ldots, s_n \rangle = K s_1 + \ldots + K s_n$. Insbesondere ist $\mathbb{R}^2 = \mathbb{R}(1, 0) \oplus \mathbb{R}(0, 1)$.

---

[1] In manchen Büchern schreibt man $\mathrm{Span}(S)$ anstatt $\langle S \rangle$.

[2] Dies ersetzt die (disjunkte) Vereinigung, siehe Aufgabe 12.

Definition 4.5. Eine Teilmenge S eines Vektorraums V heißt

- *Erzeugendensystem*, falls $\langle S\rangle = V$. Im Fall $|S| < \infty$ nennt man V *endlich erzeugt*.
- *linear abhängig*, falls 0_V eine nicht-triviale Linearkombination von Elementen aus S ist.
- *linear unabhängig*, falls nicht linear abhängig, d. h. für paarweise verschiedene Elemente $s_1, \ldots, s_n \in S$ und $\lambda_1, \ldots, \lambda_n \in K$ gilt:

$$\sum_{i=1}^{n} \lambda_i s_i = 0 \qquad \Longrightarrow \qquad \lambda_1 = \ldots = \lambda_n = 0.$$

- *Basis*, falls S ein linear unabhängiges Erzeugendensystem ist.

Bemerkung 4.6. Da Basen Mengen sind, besitzen ihre Elemente keine feste Anordnung. Tatsächlich hängen aber viele Sätze von der Reihenfolge der Basiselemente ab. Wir führen daher folgende Sprechweise ein: Vektoren $s_1, \ldots, s_n$ heißen linear unabhängig (bzw. bilden eine Basis), falls sie paarweise verschieden sind und $\{s_1, \ldots, s_n\}$ linear unabhängig (bzw. eine Basis) ist.

Beispiel 4.7.

(a) Die leere Menge ist stets linear unabhängig und bildet eine Basis des Nullraums.

(b) Wegen $1_K \cdot 0_V = 0_V$ ist der Nullvektor niemals Bestandteil einer linear unabhängigen Menge. Ein einzelner Vektor $v \neq 0$ ist hingegen stets linear unabhängig, denn aus $\lambda v = 0$ mit $\lambda \in K^\times$ folgt der Widerspruch
$$v = 1v = (\lambda^{-1}\lambda)v = \lambda^{-1}(\lambda v) = \lambda^{-1}0 = 0.$$

(c) Vektoren $v, w \in V \setminus \{0\}$ sind genau dann linear abhängig, wenn $Kv = Kw$, d. h. v ist ein skalares Vielfache von w und umgekehrt.

(d) Jede Teilmenge einer linear unabhängigen Menge ist linear unabhängig.

(e) Für $n \geq 1$ seien
$$\begin{aligned} e_1 &:= (1, 0, \ldots, 0),\\ e_2 &:= (0, 1, 0, \ldots, 0),\\ &\;\vdots\\ e_n &:= (0, \ldots, 0, 1) \end{aligned}$$
Vektoren aus K^n. Da sich jeder Vektor $v = (v_1, \ldots, v_n) \in K^n$ in der Form $v = \sum_{i=1}^{n} v_i e_i$ schreiben lässt, ist $\{e_1, \ldots, e_n\}$ ein Erzeugendensystem von K^n. Aus $v = 0 \Longleftrightarrow v_1 = \ldots = v_n = 0$ folgt die lineare Unabhängigkeit von $\{e_1, \ldots, e_n\}$. Man nennt $e_1, \ldots, e_n$ die *Standardbasis* von K^n.

4.2 Charakterisierung und Existenz von Basen

Satz 4.8. *Sei $b_1, \ldots, b_n$ eine Basis eines K-Vektorraums V. Dann lässt sich jedes $v \in V$ eindeutig in der Form $v = \sum_{i=1}^n \lambda_i b_i$ mit $\lambda_1, \ldots, \lambda_n \in K$ schreiben. Insbesondere ist die Abbildung*

$$_B[.]\colon V \to K^n, \qquad v \mapsto {}_B[v] :=(\lambda_1, \ldots, \lambda_n)$$

eine Bijektion.

Beweis. Wegen $V = \langle b_1, \ldots, b_n \rangle$ ist jedes $v \in V$ eine Linearkombination der angegebenen Form. Seien $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n \in K$ mit

$$v = \sum_{i=1}^n \lambda_i b_i = \sum_{i=1}^n \mu_i b_i.$$

Dann ist $0 = v - v = \sum_{i=1}^n (\lambda_i - \mu_i) b_i$. Da $\{b_1, \ldots, b_n\}$ linear unabhängig ist, folgt $\lambda_i = \mu_i$ für $i = 1, \ldots, n$. □

Definition 4.9. In der Situation von Satz 4.8 nennt man $_B[v]$ die *Koordinatendarstellung* von v bzgl. B.

Lemma 4.10. *Für einen Vektorraum V und $B \subseteq V$ sind äquivalent:*

(1) B ist eine Basis von V.

(2) B ist ein minimales Erzeugendensystem, d. h. für alle $b \in B$ ist $B \setminus \{b\}$ kein Erzeugendensystem.

(3) B ist maximal linear unabhängig, d. h. für alle $v \in V \setminus B$ ist $B \cup \{v\}$ linear abhängig.

Beweis. Wir führen einen *Ringbeweis.*[3]

(1) ⇒ (2): Sei B eine Basis, also insbesondere ein Erzeugendensystem von V. Nehmen wir an, dass auch $B \setminus \{b\}$ für ein $b \in B$ ein Erzeugendensystem ist. Dann existieren $\lambda_1, \ldots, \lambda_n \in K$ und $b_1, \ldots, b_n \in B \setminus \{b\}$ mit $b = \sum_{i=1}^n \lambda_i b_i$. Wegen $-b + \sum_{i=1}^n \lambda_i b_i = 0$ wäre B dann linear abhängig. Widerspruch.

(2) ⇒ (3): Sei B ein minimales Erzeugendensystem. Sei $\sum_{i=1}^n \lambda_i b_i = 0$ für $\lambda_1, \ldots, \lambda_n \in K$ und paarweise verschiedene $b_1, \ldots, b_n \in B$. Ist $\lambda_i \neq 0$ für ein i, so gilt

$$b_i = -\lambda_i^{-1} \sum_{j \neq i} \lambda_j b_j = \sum_{j \neq i} (-\lambda_i^{-1} \lambda_j) b_j \in \langle B \setminus \{b_i\} \rangle.$$

Dann wäre aber auch $B \setminus \{b_i\}$ ein Erzeugendensystem. Also ist $\lambda_1 = \ldots = \lambda_n = 0$ und B ist linear unabhängig. Sei nun $v \in V \setminus B$. Wegen $\langle B \rangle = V$ existieren $\lambda_1, \ldots, \lambda_n \in K$ und $b_1, \ldots, b_n \in B$ mit $v = \sum_{i=1}^n \lambda_i b_i$ und $-v + \sum_{i=1}^n \lambda_i b_i = 0$. Insbesondere ist $B \cup \{v\}$ linear abhängig.

(3) ⇒ (1): Sei B maximal linear unabhängig. Wir müssen $\langle B \rangle = V$ zeigen. Sei $v \in V$. Im Fall $v \in B$ ist $v \in \langle B \rangle$. Sei also $v \notin B$. Dann ist $B \cup \{v\}$ linear abhängig. Also existieren $\lambda_1, \ldots, \lambda_n \in K^\times$, $\mu \in K$, $b_1, \ldots, b_n \in B$ mit $\mu v + \sum_{i=1}^n \lambda_i b_i = 0$. Da B linear unabhängig ist, muss $\mu \neq 0$ gelten. Dies liefert

$$v = -\mu^{-1} \sum_{i=1}^n \lambda_i b_i = \sum_{i=1}^n (-\mu^{-1} \lambda_i) b_i \in \langle B \rangle.$$

Insgesamt ist $V = \langle B \rangle$. □

[3] Ein *Zirkelschluss* hingehen ist eine fehlerhafte Argumentation, bei der die Behauptung bereits vorausgesetzt wird.

Satz 4.11 (Basisergänzungssatz)**.** *Sei V ein Vektorraum mit einem endlichen Erzeugendensystem $E \subseteq V$. Dann lässt sich jede linear unabhängige Menge $U \subseteq V$ durch Hinzunahme von Elementen aus E zu einer Basis von V ergänzen.*

Beweis. Sei $E = \{s_1, \ldots, s_n\}$. Im Fall $E \subseteq \langle U \rangle$ ist $V = \langle E \rangle \subseteq \langle U \rangle$, d. h. U ist bereits eine Basis. Sei also $E \nsubseteq \langle U \rangle$ und o. B. d. A. $s_1 \notin \langle U \rangle$. Wie üblich ist dann $U_1 := U \cup \{s_1\}$ linear unabhängig. Wir können nun das Argument mit U_1 anstelle von U wiederholen. Im Fall $E \subseteq \langle U_1 \rangle$ ist U_1 eine Basis und anderenfalls können wir $s_2 \notin \langle U_1 \rangle$ annehmen. Dann ist $U_2 := U_1 \cup \{s_2\}$ linear unabhängig usw. Da E endlich ist, erhält man nach endlich vielen Schritten eine Basis von V. □

Beispiel 4.12. Die linear unabhängige Menge $U := \big\{(1, 2, 0), (2, 1, 0)\big\} \subseteq \mathbb{R}^3$ lässt sich mit dem Standardbasisvektor e_3 zu einer Basis ergänzen (aber nicht mit e_1 oder e_2).

Satz 4.13 (Steinitzer Austauschsatz)**.** *Sei V ein Vektorraum mit Erzeugendensystem E. Für jede linear unabhängige Teilmenge $U \subseteq V$ gilt $|U| \le |E|$.*

Beweis. O. B. d. A. sei E endlich, sagen wir $E = \{s_1, \ldots, s_n\}$. Seien $u_1, \ldots, u_m \in U$ paarweise verschieden. Wir müssen $m \le n$ zeigen. Da U linear unabhängig ist, gilt $0 \neq u_1 = \sum_{i=1}^{n} \lambda_i s_i$, wobei nicht alle $\lambda_1, \ldots, \lambda_n \in K$ verschwinden. Sei also o. B. d. A. $\lambda_1 \neq 0$ und daher

$$s_1 = \lambda_1^{-1} u_1 + \sum_{i=2}^{n} (-\lambda_1^{-1} \lambda_i) s_i \in \langle u_1, s_2, \ldots, s_n \rangle.$$

Folglich ist auch $\{u_1, s_2, \ldots, s_n\}$ ein Erzeugendensystem mit n Elementen (wir haben s_1 gegen u_1 *ausgetauscht*). Schreibe nun $u_2 = \mu_1 u_1 + \sum_{i=2}^{n} \mu_i s_i$ mit $\mu_1, \ldots, \mu_n \in K$. Wegen $u_2 \notin \langle u_1 \rangle$ muss mindestens ein μ_i mit $i \ge 2$ ungleich 0 sein. Sagen wir $\mu_2 \neq 0$. Wegen

$$s_2 = -\mu_2^{-1} \mu_1 u_1 + \mu_2^{-1} u_2 - \sum_{i=3}^{n} \mu_2^{-1} \mu_i s_i \in \langle u_1, u_2, s_3, \ldots, s_n \rangle$$

kann man s_2 auf die gleiche Weise gegen u_2 austauschen. Wiederholt man diesen Prozess, so erhält man schließlich das Erzeugendensystem $\{u_1, \ldots, u_m, s_{m+1}, \ldots, s_n\}$ von V. Insbesondere ist $m \le n$. □

Beispiel 4.14. Die Menge $\big\{(1, 2, 3, 4), (-1, 4, 0, 2), (0, 5, 2, 1), (0, 0, -7, 1), (-3, 4, 1, 0)\big\} \subseteq \mathbb{R}^4$ muss linear abhängig sein, da $\{e_1, e_2, e_3, e_4\}$ ein Erzeugendensystem von $\mathbb{R}^4$ ist (beachten Sie, dass man nichts rechnen muss).

Satz 4.15. *Jeder endlich erzeugte Vektorraum besitzt eine endliche Basis und je zwei Basen sind gleichmächtig.*

Beweis. Sei V ein Vektorraum mit endlichem Erzeugendensystem E. Nach dem Basisergänzungssatz kann man die linear unabhängige Menge $\varnothing$ mit Elementen aus E zu einer Basis B von V ergänzen. Insbesondere ist $|B| \le |E| < \infty$. Sei auch C eine Basis von V. Nach dem Austauschsatz gilt $|C| \le |B| \le |C|$, also $|C| = |B|$. Da B und C endlich sind, müssen sie nach Bemerkung 2.11(e) gleichmächtig sein. □

Folgerung 4.16. *Jeder Unterraum U eines endlich erzeugten Vektorraums V ist endlich erzeugt und besitzt ein* Komplement *$W \le V$, d. h. es gilt $V = U \oplus W$.*

Beweis. Sei B eine Basis von V und $S \subseteq U$ linear unabhängig. Nach dem Austauschsatz gilt $|S| \le |B| < \infty$. Insbesondere besitzt U eine maximal linear unabhängige Teilmenge C. Nach Lemma 4.10 ist C eine (endliche) Basis von U. Also ist U endlich erzeugt. Nach dem Basisergänzungssatz lässt sich C zu einer Basis D von V ergänzen. Die zweite Behauptung folgt dann mit $W := \langle D \setminus C \rangle$. □

Bemerkung 4.17.

(a) In der Situation von Folgerung 4.16 ist W im Allgemeinen nicht eindeutig bestimmt. Zum Beispiel gilt
$$\mathbb{R}^2 = \mathbb{R}(1,0) \oplus \mathbb{R}(0,1) = \mathbb{R}(1,0) \oplus \mathbb{R}(1,1).$$

(b) Mit dem Auswahlaxiom (genauer mit ZORNs *Lemma*) kann man zeigen, dass jeder Vektorraum eine (möglicherweise unendliche) Basis besitzt und je zwei Basen gleichmächtig sind. Zum Beispiel hat $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum unendliche Basen, von denen man keine explizit angeben kann. Wir überlegen uns in Satz 6.12 wie man Basen von endlich-erzeugten Vektorräumen effizient berechnet.

4.3 Dimension

Definition 4.18. Sei B eine Basis eines endlich erzeugten K-Vektorraums V. Dann nennt man
$$d = \dim_K V = \dim V := |B| \in \mathbb{N}_0$$
die *Dimension* von V. Nach Satz 4.15 hängt d nicht von der Wahl von B ab. Anstatt „endlich erzeugt“ kann man nun *endlich-dimensional* oder genauer *d-dimensional* sagen.

Beispiel 4.19.

(a) Für jeden Körper K und $n \ge 1$ hat K^n Dimension n (wähle die Standardbasis). Der Unterraum $U := \{(x,x) \in K^2 : x \in K\} \le K^2$ ist 1-dimensional mit Basis $\{(1,1)\}$.

(b) Im $\mathbb{R}^3$ beschreibt $\{(x,y,0) : x,y \in \mathbb{R}\}$ eine 2-dimensionale Ebene. Allgemeiner nennt man einen $(d-1)$-dimensionalen Unterraum eines d-dimensionalen Raums eine *Hyperebene*.

(c) Sei V ein d-dimensionaler $\mathbb{F}_2$-Vektorraum. Die Koordinatendarstellung bzgl. einer Basis zeigt
$$|V| = |\mathbb{F}_2^d| = |\mathbb{F}_2 \times \ldots \times \mathbb{F}_2| = 2^d.$$

Bemerkung 4.20.

(a) Aus den obigen Sätzen folgen einige nützliche Fakten:

- Für $U \le V$ gilt $\dim U \le \dim V$ mit Gleichheit genau dann, wenn $U = V$ (ergänze eine Basis von U zu einer Basis von V).
- Jedes Erzeugendensystem E von V enthält eine Basis von V (reduziere zu einem minimalen Erzeugendensystem). Insbesondere ist $|E| \ge \dim V$.
- $d+1$ Vektoren eines d-dimensionalen Vektorraums sind linear abhängig.

(b) In der linearen Algebra stehen endlich-dimensionale Vektorräume im Vordergrund, während unendlich-dimensionale Vektorräume Gegenstand der *Funktionalanalysis* sind.

(c) Die folgende Formel entspricht der Gleichung $|A \cup B| = |A| + |B| - |A \cap B|$ für endliche Mengen A und B (Lemma 1.12).

Satz 4.21 (Dimensionsformel)**.** *Für Unterräume U und W eines endlich-dimensionalen Vektorraums V gilt*

$$\boxed{\dim(U + W) = \dim U + \dim W - \dim(U \cap W).}$$

Ist die Summe direkt, so gilt $\dim(U \oplus W) = \dim U + \dim W$.

Beweis. Sei $\{b_1, \ldots, b_n\}$ eine Basis von $U \cap W$. Wir ergänzen zu einer Basis $\{b_1, \ldots, b_n, c_1, \ldots, c_s\}$ von U und einer Basis $\{b_1, \ldots, b_n, d_1, \ldots, d_t\}$ von W. Da $U + W$ aus den Elementen der Form $u + w$ mit $u \in U$ und $w \in W$ besteht, wird $U + W$ von $b_1, \ldots, b_n, c_1, \ldots, c_s, d_1, \ldots, d_t$ erzeugt.

Für die lineare Unabhängigkeit seien $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_s, \rho_1, \ldots, \rho_t \in K$ mit

$$\underbrace{\sum_{i=1}^{n} \lambda_i b_i}_{=:v} + \underbrace{\sum_{j=1}^{s} \mu_j c_j}_{=:u} + \underbrace{\sum_{k=1}^{t} \rho_k d_k}_{=:w} = 0.$$

Dann ist $v + u = -w \in U \cap W$. Also lässt sich $v + u$ als Linearkombination von $b_1, \ldots, b_n$ ausdrücken. Andererseits ist die Darstellung von $v + u$ bzgl. der Basis $\{b_1, \ldots, b_n, c_1, \ldots, c_s\}$ eindeutig nach Satz 4.8. Dies zeigt $\mu_1 = \ldots = \mu_s = 0$. Nun ist $v + w = 0$ eine Linearkombination der Basis $\{b_1, \ldots, b_n, d_1, \ldots, d_t\}$. Dies geht nur falls $\lambda_1 = \ldots = \lambda_n = \rho_1 = \ldots = \rho_t = 0$. Daher ist $\{b_1, \ldots, b_n, c_1, \ldots, c_s, d_1, \ldots, d_t\}$ linear unabhängig und folglich eine Basis von $U + W$. Man erhält

$$\dim(U + W) = n + s + t = (n + s) + (n + t) - n = \dim U + \dim W - \dim(U \cap W).$$

Ist die Summe direkt, so gilt $U \cap W = \{0\}$ und die zweite Behauptung folgt. ☐

Beispiel 4.22. Sei

$$U := \langle (1,1,0), (0,2,1) \rangle \le \mathbb{R}^3,$$
$$W := \langle (1,1,1) \rangle \le \mathbb{R}^3.$$

Offenbar gilt $\dim U = 2$ und $\dim W = 1$. Für $v \in U \cap W$ existieren $\lambda, \mu, \rho \in \mathbb{R}$ mit

$$v = \lambda(1,1,0) + \mu(0,2,1) = \rho(1,1,1).$$

Es folgt $(\lambda, \lambda + 2\mu, \mu) = (\rho, \rho, \rho)$. Ein Koeffizientenvergleich liefert $\lambda = \rho = \mu$ und $3\rho = \lambda + 2\mu = \rho$. Dies kann nur für $\rho = 0$ gelten. Also ist $v = 0(1,1,1) = 0$ und $U \cap W = \{0\}$. Man erhält $\dim(U + W) = \dim U + \dim W = 2 + 1 = 3$. Wegen $U + W \le \mathbb{R}^3$ ist daher $\mathbb{R}^3 = U \oplus W$.

Satz 4.23. *Für Vektorräume $U \le V$ gilt* $\boxed{\dim V = \dim U + \dim(V/U).}$

Beweis. Sei W ein Komplement von U in V. Sei B eine Basis von W. Es genügt zu zeigen, dass $\overline{B} := \{b + U : b \in B\}$ eine Basis von U/V ist. Jedes Element in V hat die Form $v = w + u$ mit $u \in U$ und $w \in W$. Wegen $v + U = w + U$ ist $\overline{B}$ ein Erzeugendensystem von V/U. Seien nun $\lambda_b \in K$ mit $\sum_{b \in B} \lambda_b(b + U) = 0_{V/U}$. Dann ist $\sum_{b \in B} \lambda_b b \in U$. Aus $U \cap W = \{0\}$ folgt $\lambda_b = 0$ für alle $b \in B$. Also ist $\overline{B}$ linear unabhängig. ☐

5 Matrizen

5.1 Der Matrizen-Vektorraum

Bemerkung 5.1. Sofern nichts Gegenteiliges gesagt wird, setzen wir von nun an stillschweigend voraus, dass alle Vektorräume endlich-dimensional sind. Eine Matrix ist ein Schema zur expliziten Berechnung von Basen von Vektorräumen und Lösungen von linearen Gleichungssystemen. Matrizen treten auch als eigenständige Objekte in zahlreichen anderen Gebieten auf.

Definition 5.2. Sei K ein Körper und $n, m \in \mathbb{N}$. Eine ($n \times m$-)*Matrix* über K ist ein rechteckig angeordnetes nm-Tupel

$$A = (a_{ij})_{i,j=1}^{n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

mit $a_{ij} \in K$ für $i = 1, \ldots, n$ und $j = 1, \ldots, m$. Die Menge der $n \times m$-Matrizen über K bezeichnet man mit $K^{n \times m}$. Im Fall $n = m$ nennt man A *quadratisch*.

Beispiel 5.3.

(a) Die 1×1-Matrizen entsprechen genau den Elementen aus K. Die Vektoren aus K^n kann man als $1 \times n$-Matrizen auffassen. Man spricht dann von *Zeilenvektoren*. Die $m \times 1$-Matrizen heißen demnach *Spaltenvektoren*. Wenn Missverständnisse ausgeschlossen sind, benutzen wir die Standardbasis $e_1, \ldots, e_n$ sowohl als Zeilen- als auch Spaltenvektoren.

(b) Die $n \times m$-*Nullmatrix* $0_{n \times m} := (0)_{i,j} \in K^{n \times m}$, $0_n := 0_{n \times n}$ (wie üblich lassen wir die Indizes weg, wenn Missverständnisse ausgeschlossen sind).

(c) Die (quadratische) $n \times n$-*Einheitsmatrix*

$$1_n = (\delta_{ij}) := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Das Symbol δ_{ij} nennt man das Kronecker-*Delta*. Es gilt

$$\delta_{ij} := \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{falls } i \neq j. \end{cases}$$

Die Zeilen von 1_n bilden die Standardbasis $\{e_1, \ldots, e_n\}$ von K^n.

(d) Für $\lambda_1, \ldots, \lambda_n \in K$ nennt man

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) := (\delta_{ij}\lambda_i) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

eine *Diagonalmatrix*. Die Einträge $\lambda_1, \ldots, \lambda_n$ bilden die *Hauptdiagonale*. Im Spezialfall $\lambda_1 = \ldots = \lambda_n$ spricht man von *Skalarmatrizen*.

(e) Die $n \times m$-Matrix E_{st} mit einer 1 an Position (s, t) und sonst nur Nullen. Man nennt sie *Standardmatrizen*. Mit dem Kronecker-Delta gilt $E_{st} = (\delta_{is}\delta_{jt})_{i,j}$.

(f) Für $A \in K^{n\times m}$ ist $A^{\mathrm{t}} := (a_{ji})_{i,j} \in K^{m\times n}$ die zu A *transponierte Matrix*. Sie entsteht aus A durch Spiegelung an der Hauptdiagonale:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \longrightarrow \quad A^{\mathrm{t}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Offensichtlich ist $(A^{\mathrm{t}})^{\mathrm{t}} = A$. Durch Transponieren werden Zeilenvektoren zur Spaltenvektoren und umgekehrt.

(g) Quadratische Matrizen A heißen *symmetrisch*, falls $A^{\mathrm{t}} = A$.

Lemma 5.4. *Mit komponentenweisen Verknüpfungen wird $K^{n\times m}$ zu einem nm-dimensionalen K-Vektorraum:*

$$\begin{aligned} A + B &:= (a_{ij} + b_{ij})_{i,j}, \\ \lambda \cdot A &:= (\lambda a_{ij})_{i,j} \end{aligned}$$

für $A = (a_{ij}), B = (b_{ij}) \in K^{n\times m}$ und $\lambda \in K$. Die Standardmatrizen E_{st} bilden eine Basis von $K^{n\times m}$. Insbesondere ist $\dim(K^{n\times m}) = nm$.

Beweis. Die definierten Verknüpfungen auf $K^{n\times m}$ entsprechen genau den Verknüpfungen in K^{nm}, indem man die Vektoren aus K^{nm} als $n \times m$-Matrix anordnet. Da K^{nm} ein Vektorraum ist, muss auch $K^{n\times m}$ ein Vektorraum sein. Die Standardbasisvektoren $e_1, \ldots, e_{nm}$ von K^{nm} entsprechen (bis auf Reihenfolge) genau den Standardmatrizen. □

Beispiel 5.5.

(a) Achtung: Nur Matrizen vom gleichen Format können addiert werden. Zum Beispiel:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 1 \\ 1 & -2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 4 \\ 5 & 3 & 8 \end{pmatrix} \qquad 2 \cdot \begin{pmatrix} 1 & 2 \\ 0 & -3 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & -6 \end{pmatrix}$$

Die Skalarmatrizen sind genau die skalaren Vielfachen der Einheitsmatrix.

(b) Die symmetrischen Matrizen bilden einen Unterraum S von $K^{n\times n}$. Eine Basis erhält man durch die Matrizen $E_{11}, \ldots, E_{nn}$ und $E_{ij} + E_{ji}$ für $i < j$. Insbesondere ist

$$\dim S = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$

5.2 Matrizenmultiplikation

Definition 5.6. Für $A = (a_{ij}) \in K^{n\times m}$ und $B = (b_{ij}) \in K^{m\times k}$ sei $A \cdot B := (c_{ij})_{i,j} \in K^{n\times k}$ mit

$$\boxed{c_{ij} := \sum_{l=1}^{m} a_{il} b_{lj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{im}b_{mj}.}$$

Bemerkung 5.7.

(a) Merkregel: c_{ij} entsteht, indem man die i-Zeile von A mit der j-Spalte von B „verrechnet“:

$$\begin{pmatrix} * & * & * \\ * & * & * \\ a & b & c \\ * & * & * \end{pmatrix} \cdot \begin{pmatrix} * & a' \\ * & b' \\ * & c' \end{pmatrix} = \begin{pmatrix} * & * \\ * & * \\ * & aa' + bb' + cc' \\ * & * \end{pmatrix}$$

(die Sterne bezeichnen beliebige Einträge). Oft ist es hilfreich sich folgendes Schema vorzustellen:

$$\boxed{4 \times 3} \cdot \boxed{3 \times 2} = \boxed{4 \times 2}$$

(b) Die Multiplikation von Diagonalmatrizen ist einfach:

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) \cdot \operatorname{diag}(\mu_1, \ldots, \mu_n) = \operatorname{diag}(\lambda_1\mu_1, \ldots, \lambda_n\mu_n).$$

(c) Als Vektorraum ist $(K^{n\times n}, +)$ eine abelsche Gruppe. Das folgende Lemma zeigt, dass $(K^{n\times n}, +, \cdot)$ einige, aber nicht alle Körperaxiome erfüllt.[1]

Lemma 5.8. *Für alle Matrizen A, B, C mit „passendem“ Format und $\lambda \in K$ gilt*

$$\boxed{\begin{array}{lll} A \cdot 1_m = A = 1_n \cdot A, & (AB)^{\mathrm{t}} = B^{\mathrm{t}} A^{\mathrm{t}}, & \lambda(AB) = (\lambda A)B = A(\lambda B), \\ A(BC) = (AB)C, & A(B+C) = AB + AC, & (A+B)C = AC + BC. \end{array}}$$

Beweis. Wie üblich sei $A = (a_{ij})$, $B = (b_{ij})$ und $C = (c_{ij})$. Für eine beliebige Matrix M sei M_{ij} der Eintrag an Position (i, j). Dann gilt

$$\begin{aligned}
(A1_m)_{ij} &= \sum_{k=1}^{m} a_{ik}\delta_{kj} = a_{ij} = \sum_{k=1}^{n} \delta_{ik} a_{kj} = (1_n A)_{ij}, \\
((AB)^{\mathrm{t}})_{ij} &= \sum_{k=1}^{m} a_{jk} b_{ki} = \sum_{k=1}^{m} (B^{\mathrm{t}})_{ik} (A^{\mathrm{t}})_{kj} = (B^{\mathrm{t}} A^{\mathrm{t}})_{ij}, \\
(\lambda(AB))_{ij} &= \lambda \sum_{k=1}^{m} a_{ik} b_{kj} = \sum_{k=1}^{m} (\lambda a_{ik}) b_{kj} = ((\lambda A)B)_{ij} = \sum_{k=1}^{m} a_{ik} (\lambda b_{kj}) = (A(\lambda B))_{ij}, \\
(A(BC))_{ij} &= \sum_{k=1}^{m} a_{ik} (BC)_{kj} = \sum_{k=1}^{m} a_{ik} \sum_{l=1}^{n} b_{kl} c_{lj} = \sum_{k=1}^{m} \sum_{l=1}^{n} a_{ik} b_{kl} c_{lj}
\end{aligned}$$

[1]Man nennt diese schwächere Struktur einen *Ring*.

$$
\begin{aligned}
&= \sum_{l=1}^{n}\Big(\sum_{k=1}^{m} a_{ik}b_{kl}\Big)c_{lj} = \sum_{l=1}^{n}(AB)_{il}c_{lj} = ((AB)C)_{ij},\\
(A(B+C))_{ij} &= \sum_{k=1}^{m} a_{ik}(B+C)_{kj} = \sum_{k=1}^{m} a_{ik}(b_{kj}+c_{kj}) = \sum_{k=1}^{m}(a_{ik}b_{kj}+a_{ik}c_{kj})\\
&= \sum_{k=1}^{m} a_{ik}b_{kj} + \sum_{k=1}^{m} a_{ik}c_{kj} = (AB)_{ij} + (AC)_{ij},\\
((A+B)C)_{ij} &= \sum_{k=1}^{m}(a_{ik}+b_{ik})c_{kj} = \sum_{k=1}^{m} a_{ik}c_{kj} + \sum_{k=1}^{m} b_{ik}c_{kj} = (AC)_{ij} + (BC)_{ij}.
\end{aligned}
$$
□

Bemerkung 5.9.

(a) Für $A \in K^{n\times m}$ gilt selbstverständlich $0_{k\times n}\cdot A = 0_{k\times m}$ und $A\cdot 0_{m\times k} = 0_{n\times k}$.

(b) Wir nennen Matrizen $A, B \in K^{n\times n}$ *vertauschbar*, falls $AB = BA$. Das ist für $n \geq 2$ in der Regel nicht erfüllt:
$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix} = 0_2 \neq \begin{pmatrix}0&1\\0&0\end{pmatrix} = \begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}0&1\\0&0\end{pmatrix}$$
Daher ist $(K^{n\times n}, +, \cdot)$ *kein* Körper. Wir sehen auch, dass nicht jede (von 0_n verschiedene) Matrix ein Inverses bzgl. $\cdot$ besitzt (selbst die Kürzungsregel gilt nicht für Matrizen).

(c) Eine quadratische Matrix $A \in K^{n\times n}$ heißt *invertierbar*, falls eine Matrix $B \in K^{n\times n}$ mit
$$AB = 1_n = BA$$
existiert.[2] Gilt auch $AC = 1_n = CA$, so ist $C = C1_n = C(AB) = (CA)B = 1_nB = B$. Daher ist B eindeutig bestimmt und man schreibt wie bei Gruppen $A^{-1} := B$. Wir zeigen in Lemma 5.15, dass aus $AB = 1_n$ bereits die Invertierbarkeit folgt, d. h. $BA = 1_n$ muss nicht geprüft werden.

(d) Ist A invertierbar, so auch A^{t}, denn
$$A^{\mathrm{t}}(A^{-1})^{\mathrm{t}} = (A^{-1}A)^{\mathrm{t}} = 1_n^{\mathrm{t}} = 1_n = (AA^{-1})^{\mathrm{t}} = (A^{-1})^{\mathrm{t}}A^{\mathrm{t}}.$$
Man setzt daher $\boxed{A^{-\mathrm{t}} := (A^{-1})^{\mathrm{t}} = (A^{\mathrm{t}})^{-1}.}$

(e) Manchmal ist es nützlich Matrizen in rechteckige *Blöcke* zu unterteilen:
$$A = \begin{pmatrix}A_1 & A_2\\ A_3 & A_4\end{pmatrix} \qquad \text{mit } A_1 \in K^{n_1\times m_1},\ A_2 \in K^{n_1\times m_2},\ A_3 \in K^{n_2\times m_1},\ A_4 \in K^{n_2\times m_2}$$
Für eine weitere Matrix $B = \left(\begin{smallmatrix}B_1 & B_2\\ B_3 & B_4\end{smallmatrix}\right)$ mit Blöcken $B_1 \in K^{m_1\times k_1}$, $B_2 \in K^{m_1\times k_2}$, $B_3 \in K^{m_2\times k_1}$, $B_4 \in K^{m_2\times k_2}$ gilt dann
$$AB = \begin{pmatrix}A_1B_1 + A_2B_3 & A_1B_2 + A_2B_4\\ A_3B_1 + A_4B_3 & A_3B_2 + A_4B_4\end{pmatrix}.$$
Sind A_1 und A_4 quadratisch und $A_2 = A_3 = 0$, so nennt man $A = \left(\begin{smallmatrix}A_1 & 0\\ 0 & A_4\end{smallmatrix}\right) = \operatorname{diag}(A_1, A_4)$ eine *Blockdiagonalmatrix*.

[2] Invertierbare (bzw. nicht invertierbare) Matrizen nennt man auch *regulär* (bzw. *singulär*).

Lemma 5.10. *Die invertierbaren Matrizen in $K^{n\times n}$ bilden eine Gruppe* $\mathrm{GL}(n,K)$ *bzgl. Multiplikation. Man nennt sie* allgemeine lineare Gruppe vom Grad n *über* K*.*

Beweis. Das neutrale Element 1_n ist offenbar invertierbar. Mit A ist auch A^{-1} invertierbar. Für invertierbare Matrizen A und B gilt

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A1_nA^{-1} = AA^{-1} = 1_n$$

und analog $(B^{-1}A^{-1})(AB) = 1_n$. Dies zeigt, dass auch AB invertierbar ist mit $(AB)^{-1} = B^{-1}A^{-1}$. Das Assoziativgesetz der Multiplikation folgt aus Lemma 5.8. $\square$

Beispiel 5.11. In $\mathrm{GL}(2,\mathbb{F}_2)$ gilt

$$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix} = 1_2 = \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}.$$

5.3 Der Rang einer Matrix

Satz 5.12. *Seien $z_1,\ldots,z_n$ die Zeilen und $s_1,\ldots,s_m$ die Spalten einer Matrix $A\in K^{n\times m}$. Dann gilt $\dim\langle z_1,\ldots,z_n\rangle = \dim\langle s_1,\ldots,s_m\rangle$.*

Beweis. Nach Bemerkung 4.20 existiert $I\subseteq\{1,\ldots,n\}$, sodass $\{z_i : i\in I\}$ eine Basis von $\langle z_1,\ldots,z_n\rangle$ ist. Sei analog $\{s_j : j\in J\}$ eine Basis von $\langle s_1,\ldots,s_m\rangle$ ist. Sei $A = (a_{ij})$. Wir zeigen, dass die Zeilen der Matrix

$$B := (a_{ij})_{i\in I, j\in J}\in K^{|I|\times|J|}$$

linear unabhängig sind. Seien dazu $\lambda_i\in K$ mit $\sum_{i\in I}\lambda_i a_{ij} = 0$ für alle $j\in J$. Für $k\in\{1,\ldots,m\}\setminus J$ existieren $\mu_j\in K$ mit $s_k = \sum_{j\in J}\mu_j s_j$, d. h. $a_{ik} = \sum_{j\in J}\mu_j a_{ij}$ für alle $i\in\{1,\ldots,n\}$. Es folgt

$$\sum_{i\in I}\lambda_i a_{ik} = \sum_{i\in I}\lambda_i\sum_{j\in J}\mu_j a_{ij} = \sum_{j\in J}\mu_j\sum_{i\in I}\lambda_i a_{ij} = 0.$$

Also gilt $\sum_{i\in I}\lambda_i a_{ij} = 0$ für alle $j\in\{1,\ldots,m\}$, d. h. $\sum_{i\in I}\lambda_i z_i = 0$. Aus der linearen Unabhängigkeit von $\{z_i : i\in I\}$ folgt $\lambda_i = 0$ für $i\in I$. Daher sind die Zeilen von B linear unabhängig. Da sie im $|J|$-dimensionalen Vektorraum $K^{|J|}$ liegen, gilt $|I|\le|J|$. Die Zeilen (bzw. Spalten) von A sind die Spalten (bzw. Zeilen) von A^{t}. Benutzt man das obige Argument mit A^{t}, so folgt $|J|\le|I|$. Insgesamt ist $\dim\langle z_1,\ldots,z_n\rangle = |I| = |J| = \dim\langle s_1,\ldots,s_m\rangle$. $\square$

Definition 5.13. In der Situation von Satz 5.12 nennt man

$$\mathrm{rk}(A) := \dim\langle z_1,\ldots,z_n\rangle = \dim\langle s_1,\ldots,s_m\rangle$$

den *Rang* von A. Im Fall $\mathrm{rk}(A) = \min\{n,m\}$ sagt man: A hat *vollen Rang.*

Beispiel 5.14.

(a) Die Einheitsmatrix 1_n hat (vollen) Rang n, denn ihre Zeilen bilden die Standardbasis. Es gilt $\mathrm{rk}\left(\begin{smallmatrix}1 & 2\\ 2 & 4\end{smallmatrix}\right) = 1$, denn die zweite Zeile ist das Doppelte der ersten. Wir überlegen uns in Bemerkung 6.13 wie man den Rang einer beliebigen Matrix effizient berechnet.

(b) Für jede Matrix A gilt $\mathrm{rk}(A) = \mathrm{rk}(A^{\mathrm{t}})$ und $\mathrm{rk}(A) = 0 \Longleftrightarrow A = 0$.

Lemma 5.15.

(a) Für Matrizen A und B mit „passendem" Format gilt $\mathrm{rk}(AB) \le \min\{\mathrm{rk}(A), \mathrm{rk}(B)\}$.

(b) Eine quadratische Matrix ist genau dann invertierbar, wenn sie vollen Rang hat.

Beweis.

(a) Seien $s_1, \ldots, s_m$ die Spalten von A und sei $(\lambda_1, \ldots, \lambda_m)^{\mathrm{t}}$ die i-te Spalte von B. Dann ist $\lambda_1 s_1 + \ldots + \lambda_m s_m$ die i-te Spalte von AB. Also sind die Spalten von AB Linearkombinationen der Spalten von A. Dies zeigt $\mathrm{rk}(AB) \le \mathrm{rk}(A)$. Aus Beispiel 5.14 folgt

$$\mathrm{rk}(AB) = \mathrm{rk}((AB)^{\mathrm{t}}) = \mathrm{rk}(B^{\mathrm{t}} A^{\mathrm{t}}) \le \mathrm{rk}(B^{\mathrm{t}}) = \mathrm{rk}(B).$$

(b) Ist $A \in K^{n \times n}$ invertierbar, so gilt

$$n = \mathrm{rk}(1_n) = \mathrm{rk}(AA^{-1}) \overset{\text{(a)}}{\le} \mathrm{rk}(A) \le n,$$

d. h. A hat vollen Rang. Sei umgekehrt $\mathrm{rk}(A) = n$. Dann lassen sich die Standardbasisvektoren $e_1, \ldots, e_n$ als Linearkombinationen der Spalten von A ausdrücken. Also existiert $B \in K^{n \times n}$ mit $AB = 1_n$. Wegen $\mathrm{rk}(A^{\mathrm{t}}) = \mathrm{rk}(A)$ existiert auch ein $C \in K^{n \times n}$ mit $A^{\mathrm{t}} C = 1_n$, d. h. $C^{\mathrm{t}} A = (A^{\mathrm{t}} C)^{\mathrm{t}} = 1_n^{\mathrm{t}} = 1_n$. Wegen $C^{\mathrm{t}} = C^{\mathrm{t}}(AB) = (C^{\mathrm{t}} A)B = B$ ist A invertierbar. □

6 Der Gauß-Algorithmus

6.1 Gleichungssysteme

Definition 6.1. Ein (*lineares*) *Gleichungssystem* ist eine Matrixgleichung der Form $Ax = b$, wobei die *Koeffizientenmatrix* $A \in K^{n\times m}$ und der Vektor $b \in K^{n\times 1}$ gegeben sind. Gesucht ist die *Lösungsmenge*

$$L := \{x \in K^{m\times 1} : Ax = b\} \subseteq K^{m\times 1}.$$

- Im Fall $L \neq \varnothing$ nennt man das Gleichungssystem *lösbar*.
- Im Fall $b = 0$ nennt man das Gleichungssystem *homogen* und anderenfalls *inhomogen*.
- Durch Anfügen der Spalte b zu A erhält man die *erweiterte* Koeffizientenmatrix $(A|b) \in K^{n\times(m+1)}$.

Beispiel 6.2. Das Gleichungssystem

$$\begin{aligned} 2x_1 + 3x_2 &= 5, \\ -x_1 \qquad &= 2 \end{aligned}$$

entspricht der Matrixgleichung

$$\begin{pmatrix} 2 & 3 \\ -1 & 0 \end{pmatrix} x = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

mit genau einer Lösung $x = (-2, 3)^{\mathrm{t}}$.

Bemerkung 6.3. Jedes homogene Gleichungssystem ist lösbar, denn der Nullvektor ist eine Lösung.

Satz 6.4 (Kronecker-Capelli[1])**.** *Genau dann ist das Gleichungssystem $Ax = b$ lösbar, wenn $\operatorname{rk}(A) = \operatorname{rk}(A|b)$.*

Beweis. Seien $s_1, \ldots, s_m$ die Spalten von A. Dann gilt

$$\begin{aligned} \operatorname{rk}(A) = \operatorname{rk}(A|b) &\iff \dim\langle s_1, \ldots, s_m\rangle = \dim\langle s_1, \ldots, s_m, b\rangle \\ &\iff \langle s_1, \ldots, s_m\rangle = \langle s_1, \ldots, s_m, b\rangle \iff b \in \langle s_1, \ldots, s_m\rangle \\ &\iff \exists x_1, \ldots, x_m \in K : b = \sum_{i=1}^{m} x_i s_i \iff \exists x \in K^{m\times 1} : Ax = b. \end{aligned}$$

□

[1] mitunter auch nach Rouché und Frobenius benannt

Bemerkung 6.5. Sei $A \in K^{n\times m}$ mit vollem Rang $n \le m$. Dann ist

$$\mathrm{rk}(A) \le \mathrm{rk}(A|b) \le \min\{n, m+1\} = n = \mathrm{rk}(A)$$

und $Ax = b$ ist stets lösbar. Im Fall $n = m$ ist A invertierbar nach Lemma 5.15 und $x = A^{-1}b$ ist die einzige Lösung. Im Fall $n < m$ nennt man das Gleichungssystem $Ax = b$ *unterbestimmt*, d. h. es gibt weniger Gleichungen als Unbekannte. Wir zeigen, dass es dann mehrere Lösungen gibt.

Satz 6.6. *Sei $A \in K^{n\times m}$ und $b \in K^{n\times 1}$.*

(a) Die Lösungsmenge des homogenen Gleichungssystems $Ax = 0$ ist ein Unterraum L_0 von $K^{m\times 1}$ der Dimension $m - \mathrm{rk}(A)$.

(b) Besitzt das Gleichungssystem $Ax = b$ eine Lösung $\widetilde{x}$, so ist

$$\widetilde{x} + L_0 := \{\widetilde{x} + y : y \in L_0\}$$

die Lösungsmenge.

Beweis.

(a) Wegen $0 \in L_0$ ist $L_0 \neq \varnothing$. Für $x, y \in L_0$ und $\lambda \in K$ gilt

$$A(\lambda x + y) = \lambda Ax + Ay = 0,$$

d. h. $\lambda x + y \in L_0$. Dies zeigt $L_0 \le K^{m\times 1}$ (Bemerkung 3.13). Sei $b_1, \ldots, b_k$ eine Basis von L_0. Wir ergänzen zu einer Basis $b_1, \ldots, b_m$ von $K^{m\times 1}$. Die i-te Spalte von A ist Ae_i mit dem Standardbasisvektor e_i. Da e_i eine Linearkombination von $b_1, \ldots, b_m$ ist, liegt jede Spalte von A in $\langle Ab_1, \ldots, Ab_m\rangle$. Insbesondere ist $\mathrm{rk}(A) = \dim\langle Ab_1, \ldots, Ab_m\rangle$. Aus $b_1, \ldots, b_k \in L_0$ folgt $Ab_1 = \ldots = Ab_k = 0$ und

$$\mathrm{rk}(A) = \dim\langle Ab_{k+1}, \ldots, Ab_m\rangle.$$

Es genügt zu zeigen, dass die $m - k$ Vektoren $Ab_{k+1}, \ldots, Ab_m$ linear unabhängig sind, denn dann ist $\mathrm{rk}(A) = m - k$. Seien $\lambda_i \in K$ mit $\sum_{i=k+1}^{m} \lambda_i Ab_i = 0$. Dann ist auch $A\sum_{i=k+1}^{m} \lambda_i b_i = 0$, d. h. $\sum_{i=k+1}^{m} \lambda_i b_i \in L_0 = \langle b_1, \ldots, b_k\rangle$. Da $b_1, \ldots, b_m$ eine Basis von $K^{m\times 1}$ ist, erhält man $\lambda_{k+1} = \ldots = \lambda_m = 0$ wie gewünscht.

(b) Es gilt

$$Ax = b \iff Ax = A\widetilde{x} \iff A(x - \widetilde{x}) = 0 \iff x - \widetilde{x} \in L_0 \iff x \in \widetilde{x} + L_0. \qquad \square$$

Bemerkung 6.7.

(a) Hat $A \in K^{n\times m}$ vollem Rang $m \le n$, so besitzt das Gleichungssystem $Ax = b$ höchstens eine Lösung. Im Fall $m < n$ spricht man von *überbestimmten* Gleichungssystemen. Im Folgenden beschäftigen wir uns mit der expliziten Konstruktion der Lösungsmenge eines beliebigen Gleichungssystems.

(b) Da die Abbildung $L_0 \to \widetilde{x} + L_0$, $v \mapsto \widetilde{x} + v$ eine Bijektion ist, besitzt ein lösbares Gleichungssystem genauso viele Lösungen wie das entsprechende homogene Gleichungssystem. Ist K endlich (z. B. $K = \mathbb{F}_2$), so ist die Anzahl dieser Lösungen eine Potenz von $|K|$ (Satz 4.8). Für unendliche Körper wie $\mathbb{Q}$ oder $\mathbb{R}$ besitzt jedes Gleichungssystem keine, genau eine oder unendlich viele Lösungen.

6.2 Elementare Zeilenoperationen

Definition 6.8. Die folgenden Transformationen einer Matrix $A \in K^{n\times m}$ werden (*elementare*) *Zeilenoperationen* genannt:

- Multiplikation einer Zeile von A mit einem Skalar $\lambda \in K^\times$. Dies entspricht der Multiplikation mit einer *Elementarmatrix* der Form

$$\begin{pmatrix} 1_{s-1} & & 0 \\ & \lambda & \\ 0 & & 1_{n-s} \end{pmatrix} = 1_n + (\lambda - 1)E_{ss}$$

  von links an A.

- Vertauschen zweier Zeilen von A. Dies entspricht der Multiplikation mit einer Elementarmatrix der Form

$$\begin{pmatrix} 1_{s-1} & & & & \\ & 0 & & 1 & \\ & & 1_{t-s-1} & & \\ & 1 & & 0 & \\ & & & & 1_{n-t} \end{pmatrix} = 1_n - E_{ss} - E_{tt} + E_{st} + E_{ts}$$

  von links an A.

- Addieren eines Vielfaches einer Zeile von A zu einer anderen Zeile. Dies entspricht der Multiplikation mit einer Elementarmatrix der Form

$$\begin{pmatrix} 1_{s-1} & & & & \\ & 1 & & \lambda & \\ & & 1_{t-s-1} & & \\ & & & 1 & \\ & & & & 1_{n-t} \end{pmatrix} = 1_n + \lambda E_{st} \qquad (\lambda \in K,\ s \neq t)$$

  von links an A.

Matrizen A und B heißen *zeilen-äquivalent*, falls sich A durch endlich viele elementare Zeilenoperationen in B überführen lässt. Ggf. schreiben wir $A \sim B$.

Bemerkung 6.9.

(a) Alle elementaren Zeilenoperationen sind umkehrbar. Aus $A \sim B$ folgt somit $B \sim A$. Außerdem sind die Elementarmatrizen invertierbar. Nach Lemma 5.10 ist auch das Produkt von Elementarmatrizen invertierbar. Aus $A \sim B$ folgt daher die Existenz einer Matrix $S \in \mathrm{GL}(n, K)$ mit $SA = B$.

(b) Nach (a) ist die Zeilen-Äquivalenz eine Äquivalenzrelation auf $K^{n\times m}$. Im nächsten Satz bestimmen wir ein besonders einfaches Repräsentantensystem für die Äquivalenzklassen.

(c) Analog definiert man (*elementare*) *Spaltenoperationen*. Diese entsprechen der Multiplikation von Elementarmatrizen von *rechts* an A (probieren Sie es aus). Spaltenoperationen lassen sich auch durch Zeilenoperation mit A^{t} realisieren. Wir benutzen die Schreibweise $A \sim B$ auch, wenn B aus A durch Spaltenoperationen hervorgeht.

Satz 6.10 (Gauss-Algorithmus[2])**.** *Jede Matrix $A \in K^{n\times m}$ ist zeilen-äquivalent zu genau einer Matrix $\widehat{A}$ in* Zeilenstufenform[3], *d. h.*

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

Beweis. Der folgende Algorithmus überführt A in $\widehat{A}$:

(1) Setze $z := 1$ (Zeilenindex).

(2) Für $s = 1, \ldots, m$ (Spaltenindex) tue:

- Falls $\exists i \geq z : a_{is} \neq 0$, dann:
  - Tausche i-te mit z-ter Zeile. Anschließend gilt $a_{zs} \neq 0$.
  - Dividiere z-te Zeile durch a_{zs}. Anschließend gilt $a_{zs} = 1$.
  - Für $j = 1, \ldots, z-1, z+1, \ldots, n$ subtrahiere das a_{js}-Fache der z-ten Zeile von der j-ten Zeile. Anschließend gilt $a_{js} = 0$.
  - Erhöhe z um 1.

Für die Eindeutigkeit von $\widehat{A}$ seien B und C Matrizen in Zeilenstufenform mit $A \sim B$ und $A \sim C$. Dann ist auch $B \sim C$ und es existiert $S \in \mathrm{GL}(n, K)$ mit $SB = C$. Sei b_i (bzw. s_i) die i-te Spalte von B (bzw. S). Dann ist $c_i = Sb_i$ die i-te Spalte von C. Sei $e_1, \ldots, e_n$ die Standardbasis von K^n. Sei $b_i \in \langle e_1, \ldots, e_k\rangle$. Wir zeigen $b_i = c_i$ und $s_k = e_k^{\mathrm{t}}$ durch Induktion nach k. Im Fall $k = 0$ ist $b_i \in \langle\varnothing\rangle = \{0\}$, also $b_i = 0$. Sicher ist dann auch $c_i = Sb_i = 0$. Sei nun die Behauptung bis $k-1$ bereits bewiesen. Die erste Spalte (von links) von B, die nicht in $\langle e_1, \ldots, e_{k-1}\rangle$ liegt, ist $b_i = e_k$ wegen der Zeilenstufenform. Die Spalten von S sind linear unabhängig, da S invertierbar ist. Dies zeigt

$$c_i = Sb_i = Se_k^{\mathrm{t}} = s_k \notin \langle s_1, \ldots, s_{k-1}\rangle = \langle e_1, \ldots, e_{k-1}\rangle.$$

Also ist c_i die erste Spalte von C, die nicht in $\langle e_1, \ldots, e_{k-1}\rangle$ liegt, d. h. $c_i = e_k = s_k = b_i$ (Zeilenstufenform). Für jede weitere Spalte $b_j \in \langle e_1, \ldots, e_k\rangle$ gilt nun $c_j = Sb_j = (e_1, \ldots, e_k, s_{k+1}, \ldots, s_n)b_j = b_j$. Somit ist $b_i = c_i$ für $i = 1, \ldots, m$, d. h. $B = C$. □

Beispiel 6.11.

$$\begin{pmatrix} 0 & 1 & 1 & 3 \\ 2 & 1 & 3 & 0 \\ 3 & -1 & 2 & 1 \end{pmatrix} \sim \begin{pmatrix} 2 & 1 & 3 & 0 \\ 0 & 1 & 1 & 3 \\ 3 & -1 & 2 & 1 \end{pmatrix} \begin{array}{l} \mid :2 \\ \\ \\ \end{array} \sim \begin{pmatrix} 1 & 1/2 & 3/2 & 0 \\ 0 & 1 & 1 & 3 \\ 3 & -1 & 2 & 1 \end{pmatrix} \begin{array}{l} -3 \\ \\ + \end{array}$$

$$\sim \begin{pmatrix} 1 & 1/2 & 3/2 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & -5/2 & -5/2 & 1 \end{pmatrix} \begin{array}{l} + \\ -1/2 \quad 5/2 \\ + \end{array} \sim \begin{pmatrix} 1 & 0 & 1 & -3/2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 17/2 \end{pmatrix} \begin{array}{l} \\ \\ \mid : 17/2 \end{array}$$

[2] auch *Gauß-Elimination* oder *Gauß-Jordan-Algorithmus* genannt

[3] Jede von 0 verschiedene Zeile enthält eine *führende Eins*. Alle Einträge links, ober- und unterhalb einer führenden Eins sind 0. Die führenden Einsen rutschen mit jeder Zeile weiter nach rechts. Nullzeilen (falls vorhanden) stehen unten.

$$\sim \begin{pmatrix} 1 & 0 & 1 & -3/2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

6.3 Anwendungen

Satz 6.12. *Sei $U := \langle u_1, \ldots, u_n \rangle \le K^m$. Sei $A \in K^{n \times m}$ die Matrix mit Zeilen $u_1, \ldots, u_n$. Dann bilden die von 0 verschiedenen Zeilen von $\widehat{A}$ eine Basis von U. Insbesondere ist $\dim U = \operatorname{rk}(\widehat{A}) = \operatorname{rk}(A)$.*

Beweis. Durch elementare Zeilenoperationen werden Zeilen von A durch Linearkombinationen von Zeilen ersetzt. Die Zeilen von $\widehat{A}$ erzeugen daher einen Unterraum $W \le U$. Da alle Zeilenoperationen umkehrbar sind, gilt sogar $W = U$ und $\dim U = \operatorname{rk}(A) = \operatorname{rk}(\widehat{A})$. Sei k die Anzahl der von 0 verschiedenen Zeilen in $\widehat{A}$. Dann gilt $\operatorname{rk}(\widehat{A}) \le k$. Andererseits besitzt $\widehat{A}$ die linear unabhängigen Spalten $e_1, \ldots, e_k$. Dies zeigt $\operatorname{rk}(\widehat{A}) = k$ und die Behauptung folgt. □

Bemerkung 6.13. Zur Ermittelung einer Basis von U muss man beim Gauß-Algorithmus keine Nullen oberhalb der führenden Einsen erzeugen (die von 0 verschiedenen Zeilen sind trotzdem linear unabhängig, denn deren Anzahl bleibt gleich). Außerdem ist es ratsam Divisionen zu vermeiden, indem man mit Zeilen tauscht, die bereits eine führende Eins in der aktuellen Spalte haben. Ist man nur an $\dim U$ (oder allgemeiner an $\operatorname{rk}(A)$) interessiert, so darf man wegen $\operatorname{rk}(A) = \operatorname{rk}(A^{\mathrm{t}})$ auch elementare Spaltenoperationen verwenden. Dies kann nützlich sein, wenn A weniger Spalten als Zeilen besitzt (viele Möglichkeiten führen zum Ziel).

Beispiel 6.14. Eine Art Schach-Rätsel: Rang in zwei Zügen!

$$\begin{pmatrix} 1 & -1 & 0 \\ 2 & 0 & 2 \\ 3 & -1 & 2 \\ -1 & 2 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 2 & 2 & 2 \\ 3 & 2 & 2 \\ -1 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 3 & 2 & 0 \\ -1 & 1 & 0 \end{pmatrix} \Longrightarrow \operatorname{rk}(A) = 2$$

Satz 6.15. *Sei $Ax = b$ ein Gleichungssystem mit $A \in K^{n \times m}$. Sei $(\widehat{A}|c)$ die Zeilenstufenform von $(A|b)$. Dann gilt:*

(a) Genau dann ist $Ax = b$ lösbar, wenn e_{m+1} keine Zeile von $(\widehat{A}|c)$ ist.

Ggf. erhält man die Lösungsmenge wie folgt: Seien $(1, n_1), \ldots, (k, n_k)$ die Positionen der führenden Einsen in $(\widehat{A}|c)$. Die $n-k$ Nullzeilen werden gestrichen. Für alle $i \in \{1, \ldots, m\} \setminus \{n_1, \ldots, n_k\}$ fügen wir die Zeile $-e_i$ an Position i ein, sodass die resultierende Matrix $M \in K^{m \times (m+1)}$ auf der Hauptdiagonale nur Einträge ± 1 besitzt.

(b) Die letzte Spalte $\widetilde{x}$ von M erfüllt $A\widetilde{x} = b$.

(c) Die $m-k$ Spalten von M, die nicht zu den Indizes $n_1, \ldots, n_k$ gehören, bilden eine Basis von $L_0 := \{x \in K^{m \times 1} : Ax = 0\}$.

(d) Die Lösungsmenge von $Ax = b$ ist $\widetilde{x} + L_0$.

Beweis. Sei $S \in \mathrm{GL}(n,K)$ mit $(SA|Sb) = S(A|b) = (\widehat{A}|c)$. Für $x \in K^{m\times 1}$ gilt

$$Ax = b \iff SAx = Sb \iff \widehat{A}x = c.$$

Ist e_{m+1} eine Zeile von $(\widehat{A}|c)$, so erhält man in der Gleichung $\widehat{A}x = c$ den Widerspruch $0 = 1$, d. h. es gibt keine Lösung.

Sei nun e_{m+1} keine Zeile von $(\widehat{A}|c)$. Wir verifizieren die Behauptungen an folgendem Beispiel

$$(\widehat{A}|c) = \left(\begin{array}{ccccc|c} . & 1 & a_1 & . & a_2 & c_1 \\ . & . & . & 1 & a_3 & c_2 \\ . & . & . & . & . & . \end{array}\right) \qquad M = \left(\begin{array}{ccccc|c} -1 & . & . & . & . & . \\ . & 1 & a_1 & . & a_2 & c_1 \\ . & . & -1 & . & . & . \\ . & . & . & 1 & a_3 & c_2 \\ . & . & . & . & -1 & . \end{array}\right)$$

(hier steht . für 0 zur besseren Übersicht). Man sieht leicht, dass $\widehat{A}\widetilde{x} = c$ gilt. Somit gilt auch $A\widetilde{x} = b$. Damit sind (a) und (b) bewiesen. Genauso sieht man, dass die (rot markierten) Spalten s_1, s_3 und s_5 von M in L_0 liegen. Die verschiedenen Positionen der Einträge -1 in s_i implizieren die lineare Unabhängigkeit von $\{s_1, s_3, s_5\}$. Nach Satz 6.6 und Bemerkung 6.13 ist andererseits

$$\dim L_0 = m - \mathrm{rk}(A) = m - \mathrm{rk}(\widehat{A}) = m - k = 3.$$

Dies zeigt (c). Aus Satz 6.6 folgt (d). $\square$

Beispiel 6.16.

$$\begin{pmatrix} -1 & 3 & 1 & 1 & 0 \\ -2 & 1 & -3 & 2 & -4 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix} x = \begin{pmatrix} 11 \\ -6 \\ 4 \end{pmatrix}$$

$$(A|b) = \left(\begin{array}{ccccc|c} -1 & 3 & 1 & 1 & 0 & 11 \\ -2 & 1 & -3 & 2 & -4 & -6 \\ 0 & 1 & 1 & 0 & 0 & 4 \end{array}\right) \sim \cdots \sim \begin{pmatrix} 1 & 0 & 2 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

$$M = \begin{pmatrix} 1 & 0 & 2 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 4 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

$$L = \widetilde{x} + L_0 = \begin{pmatrix} 1 \\ 4 \\ 0 \\ 0 \\ 2 \end{pmatrix} + \left\langle \begin{pmatrix} 2 \\ 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} \right\rangle$$

Kontrolle:

$$A\widetilde{x} = \begin{pmatrix} -1 & 3 & 1 & 1 & 0 \\ -2 & 1 & -3 & 2 & -4 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 11 \\ -6 \\ 4 \end{pmatrix} = b.$$

Satz 6.17 (Matrizeninversion)**.** *Für $A \in K^{n\times n}$ sei $(\widehat{A}|B)$ die Zeilenstufenform von $(A|1_n) \in K^{n\times 2n}$. Genau dann ist A invertierbar, wenn $\widehat{A} = 1_n$. Ggf. ist $A^{-1} = B$.*

Beweis. Es gilt

$$A \text{ invertierbar} \overset{5.15}{\iff} \operatorname{rk}(A) = n \overset{6.12}{\iff} \operatorname{rk}(\widehat{A}) = n \iff \widehat{A} = 1_n.$$

Sei nun $S \in \mathrm{GL}(n, K)$ mit

$$(SA|S) = S(A|1_n) = (\widehat{A}|B) = (1_n|B).$$

Dann ist $B = S = S(AA^{-1}) = (SA)A^{-1} = 1_nA^{-1} = A^{-1}$. □

Folgerung 6.18. *Jede invertierbare Matrix ist ein Produkt von Elementarmatrizen.*

Beweis. Für $A \in \mathrm{GL}(n, K)$ gilt $A \sim 1_n$ nach Satz 6.17. □

Beispiel 6.19.

$$(A|1_3) = \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 & 1 & 0 \\ -1 & 1 & -2 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 1 & -3 & 1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -3 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -3 & 1 & 0 & 1 \\ 0 & 0 & -3 & 2 & 1 & 1 \end{array}\right) \quad | : (-3)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -3 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2/3 & -1/3 & -1/3 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/3 & -1/3 & -1/3 \\ 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -2/3 & -1/3 & -1/3 \end{array}\right)$$

$$A^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 0 \\ -2 & -1 & -1 \end{pmatrix}$$

Bemerkung 6.20.

(a) Ergibt sich während des Gauß-Algorithmus ein „Versatz" der Zeilen

$$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 \;\; 1 & \\ & & & & * \end{pmatrix},$$

so muss die Zeilenstufenform eine Nullzeile aufweisen. Die Matrix kann dann nicht invertierbar sein und man kann den Algorithmus vorzeitig abbrechen.

(b) Matrizen $A, B \in K^{n\times m}$ heißen *äquivalent*, falls man A durch Zeilen- und Spaltenoperationen in B überführen kann (und umgekehrt). Das bedeutet es existieren $S \in \mathrm{GL}(n, K)$ und $T \in \mathrm{GL}(m, K)$ mit $SAT = B$. Offenbar definiert dies eine Äquivalenzrelation. Man sieht leicht, dass jede Matrix A zu einer Matrix der Form

$$\begin{pmatrix} 1_r & 0 \\ 0 & 0 \end{pmatrix}$$

äquivalent ist. Dabei ist $r = \mathrm{rk}(A)$ eindeutig bestimmt. Insbesondere sind A und B genau dann äquivalent, wenn $\mathrm{rk}(A) = \mathrm{rk}(B)$ gilt. Man braucht für die Äquivalenz also kein neues Symbol einführen. Die Anzahl der Äquivalenzklassen von $n \times n$-Matrizen ist $n + 1$.

(c) Der folgende Satz liefert einen effizienten Algorithmus um gleichzeitig Summe und Durchschnitt von Unterräumen zu bestimmen.

Satz 6.21 (Zassenhaus-Algorithmus). *Seien $U := \langle u_1, \ldots, u_s \rangle \le K^n$ und $W := \langle w_1, \ldots, w_t \rangle \le K^n$. Sei*

$$A := \begin{pmatrix} u_1 & u_1 \\ \vdots & \vdots \\ u_s & u_s \\ w_1 & 0 \\ \vdots & \vdots \\ w_t & 0 \end{pmatrix} \in K^{(s+t)\times 2n}, \qquad \widehat{A} = \begin{pmatrix} b_1 & * \\ \vdots & \vdots \\ b_k & * \\ 0 & c_1 \\ \vdots & \vdots \\ \vdots & c_l \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix},$$

wobei $b_1, \ldots, b_k, c_1, \ldots, c_l \in K^n$ mit $b_k \ne 0 \ne c_l$. Dann ist $\{b_1, \ldots, b_k\}$ eine Basis von $U + W$ und $\{c_1, \ldots, c_l\}$ ist eine Basis von $U \cap W$.

Beweis. Wegen $U + W = \langle u_1, \ldots, u_s, w_1, \ldots, w_t \rangle$ ist $\{b_1, \ldots, b_k\}$ eine Basis von $U + W$ nach Satz 6.12. Außerdem ist jede Zeile der Form $(0, c_m)$ von $\widehat{A}$ eine Linearkombination der Zeilen von A, sagen wir

$$(0, c_m) = \sum_{i=1}^{s} \lambda_i (u_i, u_i) + \sum_{j=1}^{t} \mu_j (w_j, 0)$$

mit $\lambda_1, \ldots, \lambda_s, \mu_1, \ldots, \mu_t \in K$. Dies zeigt

$$c_m = \sum_{i=1}^{s} \lambda_i u_i = -\sum_{j=1}^{t} \mu_j w_j \in U \cap W$$

für $m = 1, \ldots, l$. Aufgrund der Zeilenstufenform ist $\{c_1, \ldots, c_l\}$ linear unabhängig. Durch elementare *Spalten*operationen überführt man A zu

$$\begin{pmatrix} 0 & u_1 \\ \vdots & \vdots \\ 0 & u_s \\ w_1 & 0 \\ \vdots & \vdots \\ w_t & 0 \end{pmatrix}.$$

Aus Bemerkung 6.13 folgt nun leicht $\mathrm{rk}(A) = \dim U + \dim W$. Die Dimensionsformel liefert daher

$$\dim(U \cap W) = \dim U + \dim W - \dim(U + W) = \mathrm{rk}(A) - k = \mathrm{rk}(\widehat{A}) - k = l.$$

Also ist $\{c_1, \ldots, c_l\}$ eine Basis von $U \cap W$. $\square$

Beispiel 6.22. Sei $U := \langle (1,1,1,0), (0,-4,1,5) \rangle$ und $W := \langle (0,-2,1,2), (1,-1,1,3) \rangle$. Wie üblich muss man nicht alle Schritte des Gauß-Algorithmus durchführen:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & -4 & 1 & 5 & 0 & -4 & 1 & 5 \\ 0 & -2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 3 & 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & -2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & -4 & 1 & 5 & 0 & -4 & 1 & 5 \\ 0 & -2 & 0 & 3 & -1 & -1 & -1 & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & -2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & -4 & 1 & 5 \\ 0 & 0 & -1 & 1 & -1 & -1 & -1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & -2 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & -4 & 1 & 5 \\ 0 & 0 & 0 & 0 & -1 & 3 & -2 & -5 \end{pmatrix}$$

Es folgt $U + W = \langle (1,1,1,0), (0,-2,1,2), (0,0,-1,1) \rangle$ und $U \cap W = \langle (-1,3,-2,-5) \rangle$.

7 Lineare Abbildungen

7.1 Definitionen und Beispiele

Bemerkung 7.1. Um verschiedene Vektorräume V und W in Beziehung zu setzen, studieren wir Abbildungen $V \to W$, die Addition und Skalarmultiplikation „respektieren". Es wird sich zeigen, dass solche Abbildungen durch Matrizen beschrieben werden können.

Definition 7.2. Eine Abbildung $f\colon V \to W$ zwischen K-Vektorräumen V und W heißt *linear* oder *Homomorphismus*, falls für alle $u, v \in V$ und $\lambda \in K$ gilt:

$$\boxed{f(\lambda u + v) = \lambda f(u) + f(v).}$$

Die Menge der linearen Abbildungen $V \to W$ wird mit $\operatorname{Hom}(V, W)$ bezeichnet. Ist f linear und bijektiv, so nennt man f einen *Isomorphismus*. Ggf. nennt man V und W *isomorph* und schreibt $V \cong W$.

Bemerkung 7.3.

(a) Eine Abbildung $f\colon V \to W$ ist genau dann linear, wenn

$$\begin{aligned} f(u+v) &= f(u) + f(v), \\ f(\lambda u) &= \lambda f(u) \end{aligned}$$

für alle $u, v \in V$ und $\lambda \in K$ gilt (setze $\lambda = 1$ bzw. $v = 0$ in Definition 7.2; vgl. Bemerkung 3.13). Isomorphe Vektorräume unterscheiden sich daher nur durch die Benennung ihrer Elemente.

(b) Für $f \in \operatorname{Hom}(V, W)$ gilt

$$f(0_V) = f(0_K \cdot 0_V) = 0_K f(0_V) = 0_W.$$

(c) Sei $f \in \operatorname{Hom}(U, V)$ und $g \in \operatorname{Hom}(V, W)$. Für $u, u' \in U$ und $\lambda \in K$ gilt

$$(g \circ f)(\lambda u + u') = g(\lambda f(u) + f(u')) = \lambda g(f(u)) + g(f(u')) = \lambda (g \circ f)(u) + (g \circ f)(u').$$

Dies zeigt $g \circ f \in \operatorname{Hom}(U, W)$.

(d) Ist $f\colon V \to W$ ein Isomorphismus, so ist auch $f^{-1}\colon W \to V$ ein Isomorphismus[1], denn für $w, w' \in W$ und $\lambda \in K$ gilt

$$\begin{aligned} f^{-1}(\lambda w + w') &= f^{-1}\bigl(\lambda f(f^{-1}(w)) + f(f^{-1}(w'))\bigr) \\ &= f^{-1}\bigl(f(\lambda f^{-1}(w) + f^{-1}(w'))\bigr) = \lambda f^{-1}(w) + f^{-1}(w'). \end{aligned}$$

[1] In der Analysis ist die Umkehrabbildung einer bijektiven stetigen Funktion im Allgemeinen nicht stetig. Daher gibt es den merkwürdigen Begriff *Homöomorphismus* (kein Tippfehler).

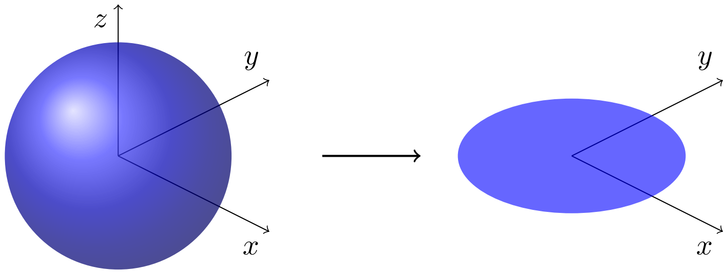
(e) Die Isomorphie von Vektorräumen ist eine Äquivalenzrelation.[2] Die Reflexivität folgt aus dem Isomorphismus id_V (Beispiel 7.4), die Symmetrie folgt aus (c) und die Transitivität folgt aus (d).

Beispiel 7.4.

(a) Die *Nullabbildung* $0\colon V \to W$, $v \mapsto 0_W$ ist stets linear. Die Identität $\mathrm{id}_V\colon V \to V$ ist ein Isomorphismus.

(b) Für $f \in \mathrm{Hom}(V, W)$ und $U \le V$ ist die Einschränkung $f_{|U}$ linear. Insbesondere ist die Inklusionsabbildung $U \to V$ als Einschränkung von id_V linear.

(c) Für $n \le m$ ist die *Projektion* $K^m \to K^n$, $(x_1, \ldots, x_m) \mapsto (x_1, \ldots, x_n)$ ein surjektiver Homomorphismus. Die Projektion $\mathbb{R}^3 \to \mathbb{R}^2$ reduziert ein 3-dimensionales Objekt auf seinen „Schatten“:



(d) Sei $f\colon \mathbb{R} \to \mathbb{R}$ linear und $a := f(1)$. Für $x \in \mathbb{R}$ gilt $f(x) = f(x1) = xf(1) = ax$. Der Graph von f beschreibt daher eine Gerade durch den Koordinatenursprung. Achtung: In der Schulmathematik werden mitunter auch Funktionen der Form $f(x) = ax + b$ als „linear“ bezeichnet (solche Abbildungen heißen *affin*[3], siehe Aufgabe 21).

(e) Für $A \in K^{n\times m}$ ist die Abbildung $K^{m\times 1} \to K^{n\times 1}$, $x \mapsto Ax$ nach Lemma 5.8 linear. Wir zeigen in Satz 7.18, dass jede lineare Abbildung (nach Basiswahl) diese Form besitzt.

(f) Die Transposition $K^{n\times m} \to K^{m\times n}$, $A \mapsto A^{\mathrm{t}}$ ist ein Isomorphismus.

(g) Für jede n-elementige Menge $M = \{x_1, \ldots, x_n\}$ ist die Abbildung $\mathrm{Abb}(M, V) \to V^n$, $f \mapsto (f(x_1), \ldots, f(x_n))$ ein Isomorphismus.

Lemma 7.5. *Für $f \in \mathrm{Hom}(V, W)$, $V_1 \le V$ und $W_1 \le W$ ist $f(V_1) \le W$ und $f^{-1}(W_1) \le V$. Insbesondere ist $f(V) \le W$ und $\mathrm{Ker}(f) := f^{-1}(\{0\}) \le V$.*

Beweis. Wegen $0 = f(0) \in f(V_1)$ ist $f(V_1) \neq \varnothing$. Für $u, v \in V_1$ und $\lambda \in K$ gilt

$$\lambda f(u) + f(v) = f(\lambda u + v) \in f(V_1).$$

Dies zeigt $f(V_1) \le W$ (Bemerkung 3.13). Wegen $0 \in f^{-1}(\{0\}) \subseteq f^{-1}(W_1)$ ist auch $f^{-1}(W_1) \neq \varnothing$. Für $u, w \in f^{-1}(W_1)$ und $\lambda \in K$ gilt $f(\lambda u + w) = \lambda f(u) + f(w) \in W_1$, d. h. $\lambda u + w \in f^{-1}(W_1)$. Dies zeigt $f^{-1}(W_1) \le V$. □

Definition 7.6. In der Situation von Lemma 7.5 nennt man $\mathrm{Ker}(f)$ den *Kern* von f und

$$\mathrm{rk}(f) := \dim f(V)$$

den *Rang* von f. Für $A \in K^{n\times m}$ sei $\mathrm{Ker}(A) := \{x \in K^{m\times 1} : Ax = 0\}$ der *Kern* von A.

[2] Allerdings ist die Gesamtheit aller Vektorräume keine Menge, sondern eine *Klasse*.
[3] Ein Beispiel ist die Umrechnung von Grad Celsius nach Fahrenheit: $f(x) = \frac{9}{5}x + 32$.

Lemma 7.7. *Genau dann ist $f \in \mathrm{Hom}(V,W)$ injektiv, wenn $\mathrm{Ker}(f) = \{0\}$. Ggf. ist $V \to f(V)$, $v \mapsto f(v)$ ein Isomorphismus.*

Beweis. Sei f injektiv und $v \in \mathrm{Ker}(f)$. Aus $f(v) = 0 = f(0)$ folgt $v = 0$, d. h. $\mathrm{Ker}(f) = \{0\}$. Sei umgekehrt $\mathrm{Ker}(f) = \{0\}$ und $u, v \in V$ mit $f(u) = f(v)$. Dann ist $f(u - v) = f(u) - f(v) = 0$, also $u - v \in \mathrm{Ker}(f) = \{0\}$. Daher ist $u = v$ und f ist injektiv. Die zweite Aussage ist trivial. □

Satz 7.8. *Seien V und W Vektorräume. Sei $b_1, \ldots, b_n$ eine Basis von V und seien $c_1, \ldots, c_n \in W$ beliebig. Dann existiert genau eine lineare Abbildung $f\colon V \to W$ mit $f(b_i) = c_i$ für $i = 1, \ldots, n$. Dabei gilt:*

(a) f injektiv $\iff$ $c_1, \ldots, c_n$ linear unabhängig.

(b) f surjektiv $\iff$ $W = \langle c_1, \ldots, c_n \rangle$.

(c) f Isomorphismus $\iff$ $c_1, \ldots, c_n$ Basis von W.

Beweis. Jedes $u \in V$ lässt sich eindeutig in der Form $u = \sum_{i=1}^n \lambda_i b_i$ schreiben. Wir definieren

$$f(u) := \sum_{i=1}^n \lambda_i c_i \in W.$$

Für $v = \sum_{i=1}^n \mu_i b_i$ und $\rho \in K$ gilt

$$f(\rho u + v) = f\Big(\sum_{i=1}^n (\rho\lambda_i + \mu_i) b_i\Big) = \sum_{i=1}^n (\rho\lambda_i + \mu_i) c_i = \rho \sum_{i=1}^n \lambda_i c_i + \sum_{i=1}^n \mu_i c_i = \rho f(u) + f(v).$$

Also ist f linear mit $f(b_i) = c_i$ für $i = 1, \ldots, n$. Ist auch $g \in \mathrm{Hom}(V,W)$ mit $g(b_i) = c_i$ für $i = 1, \ldots, n$, so gilt

$$g(u) = \sum_{i=1}^n \lambda_i g(b_i) = \sum_{i=1}^n \lambda_i c_i = \sum_{i=1}^n \lambda_i f(b_i) = f(u)$$

für alle $u \in V$. Also ist $g = f$ und f ist eindeutig bestimmt.

(a) Sei f injektiv und $\sum_{i=1}^n \lambda_i c_i = 0$ für $\lambda_i \in K$. Dann ist

$$f\Big(\sum_{i=1}^n \lambda_i b_i\Big) = \sum_{i=1}^n \lambda_i c_i = 0.$$

Aus Lemma 7.7 folgt $\sum_{i=1}^n \lambda_i b_i \in \mathrm{Ker}(f) = \{0\}$. Da $b_1, \ldots, b_n$ linear unabhängig sind, erhält man $\lambda_1 = \ldots = \lambda_n = 0$. Also sind $c_1, \ldots, c_n$ linear unabhängig. Seien nun umgekehrt $c_1, \ldots, c_n$ linear unabhängig und $u := \sum_{i=1}^n \lambda_i b_i \in \mathrm{Ker}(f)$. Dann ist $\sum_{i=1}^n \lambda_i c_i = f(u) = 0$ und man erhält $\lambda_1 = \ldots = \lambda_n = 0$. Daher ist $u = 0$ und $\mathrm{Ker}(f) = \{0\}$. Nach Lemma 7.7 ist f injektiv.

(b) Sei f surjektiv und $w \in W$. Dann existiert ein $v = \sum_{i=1}^n \lambda_i b_i \in V$ mit $f(v) = w$. Es folgt

$$w = f(v) = \sum_{i=1}^n \lambda_i c_i \in \langle c_1, \ldots, c_n \rangle.$$

Sei umgekehrt $W = \langle c_1, \ldots, c_n \rangle$ und $w \in W$. Dann existieren $\lambda_i \in K$ mit $w = \sum_{i=1}^n \lambda_i c_i$. Für $v := \sum_{i=1}^n \lambda_i b_i \in V$ gilt dann $f(v) = w$, d. h. f ist surjektiv.

(c) Folgt aus (a) und (b). □

Bemerkung 7.9. Sei $f \in \mathrm{Hom}(V, W)$. Im Fall $\dim V < \dim W$ ist f nicht surjektiv, denn das Bild einer Basis von V kann kein Erzeugendensystem von W sein. Im Fall $\dim V > \dim W$ ist f nicht injektiv, denn das Bild einer Basis kann nicht linear unabhängig sein. Für $\dim V = \dim W$ erhält man

$$f \text{ injektiv} \iff f \text{ surjektiv} \iff f \text{ bijektiv}$$

(vgl. Bemerkung 2.11(e)).

Satz 7.10. *Zwei K-Vektorräume sind genau dann isomorph, wenn sie die gleiche Dimension haben. Insbesondere ist jeder n-dimensionale K-Vektorraum zu K^n isomorph.*

Beweis. Sei $f\colon V \to W$ ein Isomorphismus von Vektorräumen und B eine Basis von V. Nach Satz 7.8 ist $f(B)$ eine Basis von W. Also gilt $\dim V = |B| = |f(B)| = \dim W$. Haben umgekehrt V und W die gleiche Dimension, so gibt es Basen $\{b_1, \ldots, b_n\}$ und $\{c_1, \ldots, c_n\}$ von V bzw. W. Nach Satz 7.8 existiert ein Isomorphismus $f\colon V \to W$ mit $f(b_i) = c_i$ für $i = 1, \ldots, n$. Die zweite Behauptung folgt aus $\dim K^n = n$. Einen expliziten Isomorphismus erhält man durch die Koordinatendarstellung $V \to K^n$, $v \mapsto {}_B[v]$ (sie bildet B auf die Standardbasis von K^n ab). □

Bemerkung 7.11.

(a) Die Vektorräume $\{0\}, K, K^2, \ldots$ bilden ein Repräsentantensystem für die Isomorphieklassen von endlich-dimensionalen K-Vektorräumen.

(b) Für K-Vektorräume $V_1, \ldots, V_n$ mit $d_i := \dim V_i$ gilt $V_1 \times \ldots \times V_n \cong K^{d_1} \times \ldots \times K^{d_n} \cong K^{d_1 + \ldots + d_n}$.

(c) Obwohl $\mathbb{Q}$ und $\mathbb{Q}^2$ gleichmächtig sind, gilt $\mathbb{Q} \ncong \mathbb{Q}^2$ nach Satz 7.10.

(d) Sein $U, V \le W$ Vektorräume. Offenbar gilt $V \le U + V$ und $U \cap V \le U$. Aus Satz 4.23 und der Dimensionsformel folgt

$$\dim((U + V)/V) = \dim(U + V) - \dim(V) = \dim(U) - \dim(U \cap V) = \dim(U/(U \cap V)).$$

Mit Satz 7.10 erhält man den sogenannten *ersten Isomorphiesatz*[4]

$$\boxed{(U + V)/V \cong U/(U \cap V).}$$

(e) Für Vektorräume $U \le V \le W$ gilt offenbar $V/U \le W/U$. Aus

$$\begin{aligned} \dim\bigl((W/U)/(V/U)\bigr) &= \dim(W/U) - \dim(V/U) \\ &= \dim(W) - \dim(U) - \dim(V) + \dim(U) = \dim(W/V) \end{aligned}$$

erhält man den *zweiten Isomorphiesatz*[5]

$$\boxed{(W/U)/(V/U) \cong W/V.}$$

Satz 7.12 (Homomorphiesatz)**.** *Für $f \in \mathrm{Hom}(V, W)$ ist die Abbildung*

$$\begin{aligned} \overline{f}\colon V/\operatorname{Ker}(f) &\to f(V), \\ v + \operatorname{Ker}(f) &\mapsto f(v) \end{aligned}$$

ein Isomorphismus. Also gilt $\boxed{V/\operatorname{Ker}(f) \cong f(V)}$ *und* $\boxed{\dim V = \mathrm{rk}(f) + \dim \operatorname{Ker}(f).}$

[4] Merkregel: Auf einer Seite stehen zwei U, auf der anderen Seite zwei V.

[5] Merkregel: Kürzen eines Doppelbruchs.

Beweis. Für $v, w \in V$ gilt

$$\begin{aligned} v + \operatorname{Ker}(f) = w + \operatorname{Ker}(f) &\iff v - w \in \operatorname{Ker}(f) \iff f(v-w) = 0 \\ &\iff f(v) = f(w) \iff \overline{f}(v + \operatorname{Ker}(f)) = \overline{f}(w + \operatorname{Ker}(f)). \end{aligned}$$

Die Implikation $\Rightarrow$ zeigt, dass $\overline{f}$ wohldefiniert ist, während die Implikation $\Leftarrow$ zeigt, dass $\overline{f}$ injektiv ist. Offenbar ist $\overline{f}$ auch linear und surjektiv. Die Gleichung für $\dim V$ folgt aus Satz 4.23. □

7.2 Darstellungsmatrizen

Satz 7.13. *Für K-Vektorräume V und W ist* $\operatorname{Hom}(V, W)$ *ein Unterraum von* $\operatorname{Abb}(V, W)$.

Beweis. Sicher liegt das neutrale Element $f = 0$ von $\operatorname{Abb}(V, W)$ in $\operatorname{Hom}(V, W)$. Seien $f, g \in \operatorname{Hom}(V, W)$, $u, v \in V$ und $\lambda, \mu \in K$. Wegen

$$\begin{aligned} (f+g)(\mu u + v) &= f(\mu u + v) + g(\mu u + v) = \mu f(u) + f(v) + \mu g(u) + g(v) = \mu(f+g)(u) + (f+g)(v), \\ (\lambda f)(\mu u + v) &= \lambda f(\mu u + v) = \lambda(\mu f(u) + f(v)) = \mu\lambda f(u) + \lambda f(v) = \mu(\lambda f)(u) + (\lambda f)(v) \end{aligned}$$

sind $f + g$ und λf linear. □

Bemerkung 7.14. Ist B eine Basis von V, so ist die Einschränkungsabbildung $\operatorname{Hom}(V, W) \to \operatorname{Abb}(B, W)$, $f \mapsto f_{|B}$ ein Isomorphismus nach Satz 7.8.

Definition 7.15. Seien V und W Vektorräume mit Basen $B = \{b_1, \ldots, b_m\}$ bzw. $C = \{c_1, \ldots, c_n\}$. Sei $f \in \operatorname{Hom}(V, W)$ und $f(b_i) = \sum_{j=1}^n a_{ji} c_j$ mit $a_{ji} \in K$ für $i = 1, \ldots, m$.

(a) Man nennt ${}_C[f]_B := (a_{ij}) \in K^{n \times m}$ die *Darstellungsmatrix* von f bzgl. B und C.

(b) Im Fall $V = W$ und $f = \mathrm{id}_V$ nennt man ${}_C\Delta_B := {}_C[\mathrm{id}_V]_B$ die *Basiswechselmatrix* bzgl. B und C.

(c) Sind B und C die Standardbasen von $V = K^m$ und $W = K^n$, so setzt man $[f] := {}_C[f]_B$. Merkregel: Die Spalten von $[f]$ sind die Bilder der Standardbasis.

Bemerkung 7.16. In der Situation von Definition 7.15 gilt ${}_B\Delta_B = 1_m$.

Beispiel 7.17. Die Abbildung

$$f \colon \mathbb{R}^2 \to \mathbb{R}^3, \qquad (x, y) \mapsto (2x - y,\ y,\ -3x)$$

ist linear mit Matrix

$$[f] = \begin{pmatrix} 2 & -1 \\ 0 & 1 \\ -3 & 0 \end{pmatrix}.$$

Offenbar sind $B := \big\{(1, -1), (0, 2)\big\}$ und $C := \big\{(1, 1, 1), (0, -1, 1), (1, 0, 1)\big\}$ Basen von $\mathbb{R}^2$ bzw. $\mathbb{R}^3$. Wegen

$$\begin{aligned} f(1, -1) &= (3, -1, -3) = -7(1, 1, 1) - 6(0, -1, 1) + 10(1, 0, 1), \\ f(0, 2) &= (-2, 2, 0) = 4(1, 1, 1) + 2(0, -1, 1) - 6(1, 0, 1) \end{aligned}$$

ergibt sich

$$ {}_C[f]_B = \begin{pmatrix} -7 & 4 \\ -6 & 2 \\ 10 & -6 \end{pmatrix} $$

(im Zweifel müssen Sie die Einträge durch ein Gleichungssystem bestimmen).

Satz 7.18. *Sei V ein m-dimensionaler Vektorraum mit Basis B und sei W ein n-dimensionaler Vektorraum mit Basis C. Dann ist die Abbildung*

$$ {}_C[.]_B\colon \operatorname{Hom}(V,W) \to K^{n\times m}, \qquad f \mapsto {}_C[f]_B $$

ein Isomorphismus mit

$$ \boxed{{}_C[f(v)]^{\mathrm{t}} = {}_C[f]_{BB}[v]^{\mathrm{t}}} \qquad \begin{tikzcd} V \arrow[r, "f"] \arrow[d, "{}_B[.]"'] & W \arrow[d, "{}_C[.]"] \\ K^m \arrow[r, "{}_C[f]_B"'] & K^n \end{tikzcd} $$

für alle $v \in V$. Insbesondere ist $\dim \operatorname{Hom}(V,W) = nm$ *und* $\operatorname{rk}(f) = \operatorname{rk}({}_C[f]_B)$.

Beweis. Sei $B = \{b_1, \ldots, b_m\}$ und $C = \{c_1, \ldots, c_n\}$. Nach Satz 7.8 ist ${}_C[.]_B$ eine Bijektion. Seien $f, g \in \operatorname{Hom}(V,W)$ mit ${}_C[f]_B = (a_{ij})$ und ${}_C[g]_B = (b_{ij})$. Für $i = 1, \ldots, m$ und $\lambda \in K$ gilt

$$ (\lambda f + g)(b_i) = \lambda f(b_i) + g(b_i) = \lambda \sum_{j=1}^n a_{ji}c_j + \sum_{j=1}^n b_{ji}c_j = \sum_{j=1}^n (\lambda a_{ji} + b_{ji})c_j. $$

Dies zeigt ${}_C[\lambda f + g]_B = \lambda {}_C[f]_B + {}_C[g]_B$. Also ist ${}_C[.]_B$ ein Isomorphismus. Sei $v = \sum_{i=1}^m v_i b_i$. Dann ist

$$ f(v) = \sum_{i=1}^m v_i f(b_i) = \sum_{i=1}^m v_i \sum_{j=1}^n a_{ji}c_j = \sum_{j=1}^n \Bigl(\sum_{i=1}^m a_{ji}v_i\Bigr) c_j $$

und es folgt

$$ {}_C[f]_{BB}[v]^{\mathrm{t}} = \Bigl(\sum_{i=1}^m a_{ji}v_i\Bigr)_j^{\mathrm{t}} = {}_C[f(v)]^{\mathrm{t}}. $$

Offenbar ist ${}_B[b_i] = e_i$ der i-te Standardbasisvektor. Also ist ${}_C[f(b_i)] = {}_C[f]_{BB}[b_i]^{\mathrm{t}}$ die i-te Spalte von ${}_C[f]_B$. Da ${}_C[.]$ ein Isomorphismus ist, gilt

$$ \operatorname{rk}(f) = \dim\langle f(b_1), \ldots, f(b_m)\rangle = \dim\langle {}_C[f(b_1)], \ldots, {}_C[f(b_m)]\rangle = \operatorname{rk}({}_C[f]_B). $$

□

Bemerkung 7.19.

(a) Für $V = W$ und $f = \mathrm{id}_V$ erhält man ${}_C[v]^{\mathrm{t}} = {}_C\Delta_{BB}[v]^{\mathrm{t}}$. Für $f\colon K^n \to K^m$ gilt $f(v)^{\mathrm{t}} = [f]v^{\mathrm{t}}$ bzgl. der Standardbasen.

(b) Sei $U \le V$. Nach Folgerung 4.16 besitzt U ein Komplement W, d. h. $V = U \oplus W$. Die Projektion $f\colon V \to W$, $u + w \mapsto w$ für $u \in U$ und $w \in W$ ist ein Homomorphismus mit $\operatorname{Ker}(f) = U$. Daher ist jeder Unterraum Kern eines Homomorphismus. Seien B und C Basen von V bzw. W. Für $A := {}_C[f]_B$ gilt $\operatorname{Ker}(f) = \{v \in V : A_B[v]^{\mathrm{t}} = 0\}$ nach Satz 7.18. Also lässt sich U als Lösungsmenge eines linearen Gleichungssystems beschreiben.

Beispiel 7.20. Seien f, B und C wie in Beispiel 7.17. Für $v := (2,4) = 2(1,-1) + 3(0,2) \in \mathbb{R}^2$ gilt

$$f(v)^{\mathrm{t}} = [f]v^{\mathrm{t}} = \begin{pmatrix} 2 & -1 \\ 0 & 1 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ -6 \end{pmatrix},$$

$$_C[f(v)]^{\mathrm{t}} = {}_C[f]_{BB}[v]^{\mathrm{t}} = \begin{pmatrix} -7 & 4 \\ -6 & 2 \\ 10 & -6 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -6 \\ 2 \end{pmatrix}.$$

Kontrolle: $-2(1,1,1) - 6(0,-1,1) + 2(1,0,1) = (0,4,-6) = f(v)$.

Folgerung 7.21. *Sei $f \in \mathrm{Hom}(V,W)$ und $A := {}_C[f]_B \in K^{n\times m}$ bzgl. beliebiger Basen. Dann gilt*

(a) f injektiv $\iff \mathrm{rk}(A) = m$.

(b) f surjektiv $\iff \mathrm{rk}(A) = n$.

Beweis.

(a) $$f \text{ injektiv} \overset{7.7}{\iff} \mathrm{Ker}(A) = \mathrm{Ker}(f) = \{0\} \overset{6.6}{\iff} \mathrm{rk}(A) = m.$$

(b) $$f \text{ surjektiv} \iff f(V) = W \iff \mathrm{rk}(A) = \mathrm{rk}(f) = \dim W = n.$$ □

Satz 7.22. *Seien U, V und W Vektorräume mit Basen B, C bzw. D. Seien $f \in \mathrm{Hom}(U,V)$ und $g \in \mathrm{Hom}(V,W)$. Dann ist*

$$\boxed{{}_D[g \circ f]_B = {}_D[g]_{CC}[f]_B.}$$

Beweis. Nach Bemerkung 7.3 ist $g \circ f \in \mathrm{Hom}(U,W)$. Sei $B = \{b_1,\ldots,b_m\}$, $C = \{c_1,\ldots,c_n\}$ und $D = \{d_1,\ldots,d_k\}$. Sei ${}_C[f]_B = (a_{ij})$ und ${}_D[g]_C = (b_{ij})$. Dann gilt

$$(g \circ f)(b_i) = g\Big(\sum_{j=1}^n a_{ji}c_j\Big) = \sum_{j=1}^n a_{ji}g(c_j) = \sum_{j=1}^n a_{ji}\sum_{l=1}^k b_{lj}d_l = \sum_{l=1}^k \Big(\sum_{j=1}^n b_{lj}a_{ji}\Big)d_l.$$

Darin ist $\sum_{j=1}^n b_{lj}a_{ji}$ der Eintrag von ${}_D[g]_{CC}[f]_B$ an Position (l,i) wie gewünscht. □

Bemerkung 7.23. Merkregel: Die Komposition von linearen Abbildungen entspricht der Multiplikation von Matrizen. Für lineare Abbildungen f,g,h zwischen „passenden“ Räumen übertragen sich die Distributivgesetze von Matrizen:

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

Das kann man natürlich auch direkt nachrechnen.

Beispiel 7.24. Seien f, B und C wie in Beispiel 7.17. Sei $g \in \mathrm{Hom}(\mathbb{R}^3,\mathbb{R}^2)$ mit Matrix ${}_B[g]_C = -\left(\begin{smallmatrix} 3 & 0 & 2 \\ 1 & 1/2 & 1 \end{smallmatrix}\right)$. Dann gilt

$${}_B[g\circ f]_B = {}_B[g]_{CC}[f]_B = -\begin{pmatrix} 3 & 0 & 2 \\ 1 & 1/2 & 1 \end{pmatrix}\begin{pmatrix} -7 & 4 \\ -6 & 2 \\ 10 & -6 \end{pmatrix} = 1_2 = {}_B[\mathrm{id}_{\mathbb{R}^2}]_B,$$

d. h. $g \circ f = \mathrm{id}_{\mathbb{R}^2}$. Umgekehrt ist $f \circ g \neq \mathrm{id}_{\mathbb{R}^3}$, denn f kann nicht surjektiv sein.

Folgerung 7.25. *Sei $f\colon V \to W$ ein Isomorphismus zwischen Vektorräumen V und W mit Basen B bzw. C. Dann ist* $\boxed{{}_B[f^{-1}]_C = {}_C[f]_B^{-1}.}$ *Im Fall $V = W$ ist ${}_C\Delta_B^{-1} = {}_B\Delta_C$.*

Beweis. Nach Bemerkung 7.3 ist $f^{-1} \in \operatorname{Hom}(W, V)$. Aus Satz 7.22 folgt

$$ {}_C[f]_{BB}[f^{-1}]_C = {}_C[\mathrm{id}_W]_C = {}_C\Delta_C = 1_n $$

und ${}_B[f^{-1}]_C = {}_C[f]_B^{-1}$. Die zweite Aussage folgt mit $f = \mathrm{id}_V$. □

Bemerkung 7.26. Die Isomorphismen $f\colon V \to V$ bilden die *allgemeine lineare Gruppe* $\mathrm{GL}(V)$. Sie entsprechen genau den invertierbaren Matrizen, d. h. ${}_B[.]_B\colon \mathrm{GL}(V) \to \mathrm{GL}(n, K)$ ist ein Isomorphismus von Gruppen (anstatt von Vektorräumen).

Folgerung 7.27 (Basiswechsel)**.** *Seien B, B' Basen von V und C, C' Basen von W. Für $f \in \operatorname{Hom}(V, W)$ gilt*

$$ \boxed{{}_{C'}[f]_{B'} = {}_{C'}\Delta_{CC}[f]_{BB}\Delta_{B'}.} $$

Im Fall $V = W$ ist

$$ \boxed{{}_{B'}[f]_{B'} = {}_{B'}\Delta_{BB}[f]_{BB'}\Delta_B^{-1}.} $$

Beweis. Aus Satz 7.22 folgt

$$ {}_{C'}\Delta_{CC}[f]_{BB}\Delta_{B'} = {}_{C'}[\mathrm{id}_W]_{CC}[f]_{BB}[\mathrm{id}_V]_{B'} = {}_{C'}[\mathrm{id}_W]_{CC}[f]_{B'} = {}_{C'}[f]_{B'}. $$

Für $V = W$, $C = B$ und $C' = B'$ erhält man ${}_{B'}[f]_{B'} = {}_{B'}\Delta_{BB}[f]_{BB}\Delta_{B'} = {}_{B'}\Delta_{BB}[f]_{BB'}\Delta_B^{-1}$ mit Folgerung 7.25. □

Beispiel 7.28. Sei wieder $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ wie in Beispiel 7.17. Wir ersetzen die Basis $B = \{(1, -1), (0, 2)\}$ durch $B' := \{(0, 1), (1, 1)\}$. Wegen $(0, 1) = 0(1, -1) + \frac{1}{2}(0, 2)$ und $(1, 1) = (1, -1) + (0, 2)$ gilt

$$ {}_B\Delta_{B'} = \begin{pmatrix} 0 & 1 \\ 1/2 & 1 \end{pmatrix}, \quad {}_C[f]_{B'} = {}_C[f]_{BB}\Delta_{B'} = \begin{pmatrix} -7 & 4 \\ -6 & 2 \\ 10 & -6 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1/2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 1 & -4 \\ -3 & 4 \end{pmatrix}. $$

Definition 7.29. Matrizen $A, B \in K^{n\times n}$ heißen *ähnlich*, falls ein $S \in \mathrm{GL}(n, K)$ mit $B = SAS^{-1}$ existiert. Wir schreiben ggf. $A \approx B$.

Bemerkung 7.30.

(a) Offenbar ist $A \approx A$ (wähle $S = 1_n$). Aus $B = SAS^{-1}$ folgt $A = S^{-1}BS$. Aus $B = SAS^{-1}$ und $C = TBT^{-1}$ mit $T \in \mathrm{GL}(n, K)$ folgt $C = TSAS^{-1}T^{-1} = (TS)A(TS)^{-1}$. Daher ist die Ähnlichkeit von Matrizen eine Äquivalenzrelation.

(b) Nach Satz 7.18 und Folgerung 7.27 bestimmt jeder Endomorphismus von V durch Basiswahl eine Ähnlichkeitsklasse von Darstellungsmatrizen in $K^{n\times n}$. Dies erlaubt konkrete Berechnungen mit abstrakten Abbildungen. In der linearen Algebra II konstruieren wir spezielle Basen, sodass die Darstellungsmatrizen möglichst „einfache“ Gestalt haben (zum Beispiel Diagonalmatrizen). Dies beschleunigt Berechnungen.

Definition 7.31. Für $A = (a_{ij})_{i,j} \in K^{n\times n}$ nennen wir $\operatorname{tr}(A) := \sum_{i=1}^n a_{ii}$ die *Spur* von A. Dies ist die Summe der Hauptdiagonaleinträge.

Lemma 7.32. *Die Abbildung* $\operatorname{tr}\colon K^{n\times n} \to K$ *ist linear mit* $\operatorname{tr}(A^{\mathrm{t}}) = \operatorname{tr}(A)$ *und* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ *für* $A, B \in K^{n\times n}$.

Beweis. Für $A = (a_{ij})$, $B = (b_{ij})$ und $\lambda \in K$ gilt

$$\operatorname{tr}(\lambda A + B) = \operatorname{tr}\big((\lambda a_{ij} + b_{ij})_{i,j}\big) = \sum_{i=1}^n (\lambda a_{ii} + b_{ii}) = \lambda \sum_{i=1}^n a_{ii} + \sum_{i=1}^n b_{ii} = \lambda \operatorname{tr}(A) + \operatorname{tr}(B).$$

Also ist tr linear. Da eine Spiegelung an der Hauptdiagonale diese selbst nicht ändert, gilt $\operatorname{tr}(A^{\mathrm{t}}) = \operatorname{tr}(A)$. Außerdem gilt

$$\operatorname{tr}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ji} = \sum_{j=1}^n \sum_{i=1}^n b_{ji} a_{ij} = \operatorname{tr}(BA). \qquad \square$$

Folgerung 7.33. *Ähnliche Matrizen haben die gleiche Spur.*

Beweis. Für $A \in K^{n\times n}$ und $S \in \mathrm{GL}(n, K)$ gilt $\operatorname{tr}(S(AS^{-1})) = \operatorname{tr}((AS^{-1})S) = \operatorname{tr}(A)$. $\square$

Definition 7.34. Sei V ein Vektorraum mit Basis B und $f \in \operatorname{Hom}(V, V)$. Wir nennen $\operatorname{tr}(f) := \operatorname{tr}({}_B[f]_B)$ die *Spur* von f. Nach Folgerung 7.27 und Folgerung 7.33 hängt $\operatorname{tr}(f)$ nicht von der Wahl der Basis ab.

Satz 7.35 (Fillmore). *Sei* $A \in K^{n\times n} \setminus K1_n$ *und* $d_1, \ldots, d_n \in K$ *mit* $\operatorname{tr}(A) = d_1 + \ldots + d_n$. *Dann ist* A *zu einer Matrix mit Hauptdiagonale* $d_1, \ldots, d_n$ *ähnlich.*

Beweis. Induktion nach n: Wegen $A \notin K1_n$ ist $n \geq 2$. Sei $A = (a_{ij})$. Gilt $a_{ij} \neq 0$ für gewisse $i \neq j$, so ist $Ae_j \notin \langle e_j \rangle$. Ist A eine Diagonalmatrix, so existieren $i \neq j$ mit $a_{ii} \neq a_{jj}$. In diesem Fall ist

$$A(e_i + e_j) = a_{ii} e_i + a_{jj} e_j \notin \langle e_i + e_j \rangle.$$

In jedem Fall existiert ein $b_1 \in K^n$, sodass b_1 und Ab_1 linear unabhängig sind. Dann sind auch b_1 und $b_2 := Ab_1 - d_1 b_1$ linear unabhängig. Wir ergänzen b_1, b_2 zu einer Basis $b_1, \ldots, b_n$ von K^n. Bezüglich dieser Basis hat A die Form $\left(\begin{smallmatrix} d_1 & * \\ * & A_1 \end{smallmatrix}\right)$ mit $A_1 = (a'_{ij}) \in K^{(n-1)\times(n-1)}$. Im Fall $n = 2$ sind wir fertig, denn $\operatorname{tr}(A_1) = \operatorname{tr}(A) - d_1 = d_2$.

Sei nun $n \geq 3$. Ist $A_1 \in K1_{n-1}$, so gilt

$$A(b_1 + b_3) = d_1 b_1 + b_2 + Ab_3 \in b_2 + \langle b_1, b_3 \rangle.$$

Indem wir b_3 durch $b_1 + b_3$ ersetzen, erreichen wir $a'_{23} = 1$ und $A_1 \notin K1_{n-1}$. Nach Induktion existiert $S \in \mathrm{GL}(n-1, K)$, sodass SA_1S^{-1} Hauptdiagonale $d_2, \ldots, d_n$ hat. Nun hat

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} d_1 & * \\ * & A_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix} = \begin{pmatrix} d_1 & * \\ * & SA_1S^{-1} \end{pmatrix}$$

Hauptdiagonale $d_1, \ldots, d_n$. $\square$

Beispiel 7.36. Sei $A := \left(\begin{smallmatrix}1&0\\0&2\end{smallmatrix}\right) \in \mathbb{Q}^{2\times 2}$ und $(d_1, d_2) = (0, 3)$. Wir setzen $b_1 := (1, 1)$ und $b_2 := Ab_1 = (1, 2)$. Sei E die Standardbasis von $\mathbb{Q}^2$ und $B := \{b_1, b_2\}$. Für $S := {}_E\Delta_B = \left(\begin{smallmatrix}1&1\\1&2\end{smallmatrix}\right)$ gilt

$$S^{-1}AS = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$$

nach Folgerung 7.27.

7.3 Dualräume

Definition 7.37. Für einen K-Vektorraum V nennt man $V^* := \operatorname{Hom}(V, K)$ den *Dualraum* von V. Seine Elemente nennt man (*lineare*) *Funktionale*.

Lemma 7.38. *Sei $b_1, \ldots, b_n$ eine Basis von V. Für $i = 1, \ldots, n$ sei $b_i^* \in V^*$ mit $b_i^*(b_j) = \delta_{ij}$ für $j = 1, \ldots, n$. Dann ist $b_1^*, \ldots, b_n^*$ eine Basis von V^*.*

Beweis. Seien $\lambda_1, \ldots, \lambda_n \in K$ mit $f := \lambda_1 b_1^* + \ldots + \lambda_n b_n^* = 0$. Für $i = 1, \ldots, n$ gilt $0 = f(b_i) = \lambda_i$. Daher sind $b_1^*, \ldots, b_n^*$ linear unabhängig. Die Behauptung folgt aus $\dim V^* = \dim V$ (Satz 7.18). □

Beispiel 7.39. Ist $e_1, \ldots, e_n$ die Standardbasis von $V = K^n$, so sind $e_1^*, \ldots, e_n^*$ die Projektionen, d. h. $e_i^*(v_1, \ldots, v_n) = v_i$ für $i = 1, \ldots, n$.

Bemerkung 7.40.

(a) In der Situation von Lemma 7.38 nennt man $b_1^*, \ldots, b_n^*$ die zu $b_1, \ldots, b_n$ *duale Basis*.

(b) Für unendlich-dimensionale Räume V ist $V^* \ncong V$, denn V^* ist „größer" als V (ohne Beweis).

(c) Nach (dem Beweis von) Satz 7.13 existiert ein Isomorphismus $V \to V^*$, der eine Basis auf die entsprechende duale Basis abbildet. Allerdings gibt es keinen *kanonischen* Isomorphismus, der nicht von einer Basiswahl abhängt (vgl. Satz 12.7). Der nächste Satz zeigt, dass die Situation zwischen V und dem *Bidualraum* $V^{**} := (V^*)^*$ besser ist.

Satz 7.41. *Für $v \in V$ sei $F_v \colon V^{**} \to K$, $f \mapsto f(v)$. Dann ist $F \colon V \to V^{**}$, $v \mapsto F_v$ ein kanonischer Isomorphismus.*

Beweis. Für $f_1, f_2 \in V^*$ und $\lambda \in K$ ist

$$F_v(\lambda f_1 + f_2) = (\lambda f_1 + f_2)(v) = \lambda f_1(v) + f_2(v) = \lambda F_v(f_1) + F_v(f_2),$$

d. h. $F_v \in V^{**}$. Für $v, w \in V$ gilt

$$F_{\lambda v+w}(f) = f(\lambda v + w) = \lambda f(v) + f(w) = \lambda F_v(f) + F_w(f) = (\lambda F_v + F_w)(f),$$

d. h. F ist linear. Sei nun $F_v = 0$. Im Fall $v \neq 0$ kann man v zu einer Basis von V fortsetzen. Für die duale Basis wäre dann $0 = F_v(v^*) = v^*(v) = 1$. Dieser Widerspruch zeigt $\operatorname{Ker}(F) = \{0\}$, d. h. F ist injektiv. Wegen $\dim V^{**} = \dim V^* = \dim V < \infty$ muss F auch surjektiv sein. □

Definition 7.42. Für $U \leq V$ und $W \leq V^*$ sei

$$U^0 := \{f \in V^* : f(U) = \{0\}\} \subseteq V^*,$$
$$W_0 := \{v \in V : \forall f \in W : f(v) = 0\} \subseteq V.$$

Wir nennen U^0 (bzw. W_0) das *duale Komplement* von U (bzw. W).

Lemma 7.43.

(a) Für $U \leq V$ ist $U^0 \leq V^*$ mit $\dim V = \dim U + \dim U^0$.

(b) Für $U \leq V^*$ ist $U_0 \leq V$ mit $\dim V = \dim U + \dim U_0$.

(c) Die Abbildungen $U \mapsto U^0$ und $U \mapsto U_0$ sind zueinander invers.

(d) Für $U, W \leq V$ gilt $\boxed{(U + W)^0 = U^0 \cap W^0}$ und $\boxed{(U \cap W)^0 = U^0 + W^0.}$

(e) Für $U, W \leq V^*$ gilt $\boxed{(U + W)_0 = U_0 \cap W_0}$ und $\boxed{(U \cap W)_0 = U_0 + W_0.}$

(f) Es gilt $V = U \oplus W \iff V^* = U^0 \oplus W^0$.

Beweis.

(a) Die Einschränkung $F\colon V^* \to U^*$, $f \mapsto f_{|U}$ ist ein Homomorphismus mit Kern $U^0 \leq V^*$. Da man jedes Funktional in U^* nach V^* fortsetzen kann (Basisergänzung), ist F surjektiv. Aus dem Homomorphiesatz folgt $\dim V = \dim V^* = \dim U^0 + \dim U^* = \dim U^0 + \dim U$.

(b) Die Konstruktion aus (a) liefert zunächst $U^0 = \{f \in V^{**} : f(U) = \{0\}\} \leq V^{**}$ mit $\dim V = \dim U + \dim U^0$. Für den Isomorphismus $F\colon V \to V^{**}$, $v \mapsto F_v$ aus Satz 7.41 gilt

$$v \in U_0 \iff F_v(U) = \{0\} \iff F_v \in U^0.$$

Also ist $U_0 = F^{-1}(U^0) \leq V$ mit $\dim U_0 = \dim U^0$.

(c) Nach Definition ist $U \leq (U^0)_0$ und $U \leq (U_0)^0$. Aus Dimensionsgründen gilt Gleichheit.

(d) Es gilt $f \in (U + W)^0 \iff f(U) = \{0\} = f(W) \iff f \in U^0 \cap W^0$. Dies zeigt die erste Gleichung. Aus $U \cap W \subseteq U, W$ folgt $U^0 + W^0 = \langle U^0 \cup W^0 \rangle \subseteq (U \cap W)^0$. Nach (a) und der Dimensionsformel gilt

$$\begin{aligned}
\dim(U^0 + W^0) &= \dim U^0 + \dim W^0 - \dim(U^0 \cap W^0) \\
&= 2\dim V - \dim U - \dim W - \dim((U + W)^0) \\
&= \dim V - \dim U - \dim W + \dim(U + W) \\
&= \dim V - \dim(U \cap W) = \dim((U \cap W)^0).
\end{aligned}$$

Also ist $(U \cap W)^0 = U^0 + W^0$.

(e) Nach (c) und (d) gilt

$$\begin{aligned}
(U + W)_0 &= \big((U_0)^0 + (W_0)^0\big)_0 = \big((U_0 \cap W_0)^0\big)_0 = U_0 \cap W_0, \\
(U \cap W)_0 &= \big((U_0)^0 \cap (W_0)^0\big)_0 = \big((U_0 + W_0)^0\big)_0 = U_0 + W_0.
\end{aligned}$$

(f) Sei $V = U \oplus W$. Nach (d) ist $U^0 + W^0 = (U \cap W)^0 = \{0\}^0 = V^*$ und $U^0 \cap W^0 = (U+W)^0 = V^0 = \{0\}$. Also ist $V^* = U^* \oplus W^*$. Sei umgekehrt $V^* = U^0 \oplus W^0$. Aus (e) folgt

$$U + W = (U^0)_0 + (W^0)_0 = (U^0 \cap W^0)_0 = \{0\}_0 = V$$

und $U \cap W = (U^0)_0 \cap (W^0)_0 = (U^0 + W^0)_0 = (V^*)_0 = \{0\}$. Dies zeigt $V = U \oplus W$. □

Folgerung 7.44. *Sei V ein n-dimensionaler Vektorraum und $0 \le k \le n$. Dann gibt es eine Bijektion zwischen der Menge der k-dimensionalen Unterräume und der Menge der $(n-k)$-dimensionalen Unterräume von V.*

Beweis. Sei $F\colon V \to V^*$ ein beliebiger Isomorphismus. Sei $U \le V$ mit Dimension k. Nach Lemma 7.43 haben $F(U)_0 \le V$ und $F^{-1}(U^0) \le V$ die Dimension $n-k$. Außerdem gilt

$$\begin{aligned} F^{-1}\big((F(U)_0)^0\big) &= F^{-1}(F(U)) = U, \\ F\big(F^{-1}(U^0)\big)_0 &= (U^0)_0 = U. \end{aligned}$$

Daher sind die Abbildungen $U \mapsto F(U)_0$ und $U \mapsto F^{-1}(U^0)$ zueinander inverse Bijektionen zwischen der Menge der k-dimensionalen Unterräume und der Menge der $(n-k)$-dimensionalen Unterräume von V. □

Satz 7.45. *Sei V ein n-dimensionaler Vektorraum über einem Körper K mit $q < \infty$ Elementen. Für $1 \le k \le n$ besitzt V genau*

$$\frac{(q^n-1)(q^{n-1}-1)\dots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\dots(q-1)}$$

Unterräume der Dimension k.

Beweis. Jeder k-dimensionale Unterraum $U \le V$ wird durch ein k-Tupel linear unabhängiger Vektoren $(v_1, \dots, v_k)$ aufgespannt. Für $v_1 \in V \setminus \{0\}$ hat man $|V| - 1 = q^n - 1$ Möglichkeiten. Wegen der linearen Unabhängigkeit darf v_2 nicht in $\langle v_1 \rangle \cong K$ liegen. Daher gibt es $q^n - q$ Möglichkeiten für $v_2 \in V \setminus \langle v_1 \rangle$. Allgemein hat man $q^n - q^{i-1}$ Möglichkeiten für die Wahl von $v_i \in V \setminus \langle v_1, \dots, v_{i-1} \rangle$. Insgesamt gibt es $D(n,k) := (q^n-1)(q^n-q)\dots(q^n-q^{k-1})$ linear unabhängige k-Tupel $(v_1, \dots, v_k)$ in V. Allerdings spannen viele davon den gleichen Raum auf.

Das gleiche Argument mit U anstelle von V liefert genau $D(k,k) = (q^k-1)(q^k-q)\dots(q^k-q^{k-1})$ linear unabhängige k-Tupel in U. Also spannen genau $D(k,k)$ der k-Tupel in V den gleichen Unterraum auf. Die Anzahl der k-dimensionalen Unterräume ist daher $\frac{D(n,k)}{D(k,k)}$. Die Behauptung folgt, indem man alle Faktoren q kürzt. □

Bemerkung 7.46. Auf den ersten Blick ist nicht klar, warum die in Satz 7.45 angegebene Formel überhaupt eine ganze Zahl ist. Wir haben also eine kombinatorische Aussage mit Hilfe der linearen Algebra bewiesen. Wegen

$$\frac{(q^n-1)(q^{n-1}-1)\dots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\dots(q-1)} = \frac{(q^n-1)(q^{n-1}-1)\dots(q^{k+1}-1)}{(q^{n-k}-1)(q^{n-k-1}-1)\dots(q-1)}$$

erhält man Folgerung 7.44 in diesem Spezialfall.

Beispiel 7.47. Die Anzahl der 2-dimensionalen Unterräume von $\mathbb{F}_2^5$ ist

$$\frac{(2^5-1)(2^4-1)}{(2^2-1)(2-1)} = \frac{31 \cdot 15}{3} = 155.$$

Definition 7.48. Seien V, W Vektorräume und $f \in \mathrm{Hom}(V,W)$. Dann heißt $f^* \colon W^* \to V^*$, $g \mapsto g \circ f$ die zu f *duale Abbildung.*

Satz 7.49. *Für Vektorräume V, W ist $\mathrm{Hom}(V,W) \to \mathrm{Hom}(W^*,V^*)$, $f \mapsto f^*$ ein Isomorphismus. Seien B und C Basen von V bzw. W. Seien B^* und C^* die entsprechenden dualen Basen. Dann gilt*

$$\boxed{{}_{B^*}[f^*]_{C*} = {}_C[f]_B^{\mathrm{t}}.}$$

Insbesondere ist

(a) f injektiv $\Longleftrightarrow$ f^* surjektiv.

(b) f surjektiv $\Longleftrightarrow$ f^* injektiv.

Beweis. Für $g_1, g_2 \in W^*$ und $\lambda \in K$ gilt

$$f^*(\lambda g_1 + g_2) = (\lambda g_1 + g_2) \circ f = \lambda(g_1 \circ f) + (g_2 \circ f) = \lambda f^*(g_1) + f^*(g_2)$$

nach Bemerkung 7.23. Dies zeigt $f^* \in \mathrm{Hom}(W^*, V^*)$. Für $f_1, f_2 \in \mathrm{Hom}(V,W)$ gilt analog

$$(\lambda f_1 + f_2)^*(g) = g \circ (\lambda f_1 + f_2) = \lambda(g \circ f_1) + (g \circ f_2) = (\lambda f_1^* + f_2^*)(g).$$

Daher ist $f \mapsto f^*$ linear. Sei $B = \{b_1, \ldots, b_m\}$ und $C = \{c_1, \ldots, c_n\}$. Sei $f(b_i) = \sum_{j=1}^n a_{ji} c_j$ mit $a_{ji} \in K$. Dann gilt

$$f^*(c_i^*)(b_j) = (c_i^* \circ f)(b_j) = c_i^*\Big(\sum_{k=1}^n a_{kj} c_k\Big) = a_{ij} = \Big(\sum_{k=1}^n a_{ik} b_k^*\Big)(b_j)$$

für $j = 1, \ldots, m$. Es folgt $f^*(c_i^*) = \sum_{k=1}^n a_{ik} b_k^*$ für $i = 1, \ldots, n$. Dies zeigt ${}_C[f]_B^{\mathrm{t}} = (a_{ji})^{\mathrm{t}} = (a_{ij}) = {}_{B^*}[f^*]_{C^*}$. Nach Satz 7.18 ist $f \mapsto f^*$ ein Isomorphismus.

Mit Folgerung 7.21 ergibt sich

$$\begin{aligned} f \text{ injektiv} &\iff \mathrm{rk}({}_C[f]_B) = m \iff \mathrm{rk}({}_{B^*}[f^*]_{C^*}) = m \iff f^* \text{ surjektiv}, \\ f \text{ surjektiv} &\iff \mathrm{rk}({}_C[f]_B) = n \iff \mathrm{rk}({}_{B^*}[f^*]_{C^*}) = n \iff f^* \text{ injektiv}. \end{aligned}$$ □

Bemerkung 7.50. Für $f \in \mathrm{Hom}(V,W)$ und $g \in \mathrm{Hom}(W,U)$ gilt $(g \circ f)^* = f^* \circ g^*$, denn

$$(g \circ f)^*(\varphi) = \varphi \circ (g \circ f) = (\varphi \circ g) \circ f = g^*(\varphi) \circ f = f^*(g^*(\varphi)) = (f^* \circ g^*)(\varphi)$$

für $\varphi \in U^*$. Alternativ kann man die Matrixidentität $(AB)^{\mathrm{t}} = B^{\mathrm{t}} A^{\mathrm{t}}$ aus Lemma 5.8 benutzen.

8 Eigenwerte und Eigenvektoren

8.1 Definitionen und Beispiele

Bemerkung 8.1. In diesem Kapitel untersuchen wir Homomorphismen $f\colon V \to V$ zwischen den gleichen Räumen. Man spricht dann von *Endomorphismen* und schreibt $\mathrm{End}(V) := \mathrm{Hom}(V, V)$. Durch Wahl einer geeigneten Basis B von V werden wir erreichen, dass ${}_B[f]_B$ möglichst einfache Gestalt hat.[1]

Definition 8.2. Sei V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Man nennt $\lambda \in K$ einen *Eigenwert* von f, falls der *Eigenraum*

$$E_\lambda(f) := \{v \in V : f(v) = \lambda v\}$$

nicht der Nullraum ist. Ggf. nennt man $\dim E_\lambda(f)$ die *geometrische Vielfachheit* von λ. Die Vektoren $v \in E_\lambda(f) \setminus \{0\}$ heißen *Eigenvektoren* zum Eigenwert λ.

Bemerkung 8.3.

(a) Für $v \in V$ und $\lambda \in K$ gilt

$$f(v) = \lambda v \iff f(v) - \lambda v = 0 \iff (f - \lambda\,\mathrm{id})(v) = 0 \iff v \in \mathrm{Ker}(f - \lambda\,\mathrm{id}).$$

Daher ist der Eigenraum $E_\lambda(f) = \mathrm{Ker}(f - \lambda\,\mathrm{id})$ tatsächlich ein Unterraum von V. Nach dem Homomorphiesatz ist $\dim(V) - \mathrm{rk}(f - \lambda\,\mathrm{id})$ die geometrische Vielfachheit von λ.

(b) Sei B eine Basis von V, $x := {}_B[v]^{\mathrm{t}} \in K^{n\times 1}$ und $A := {}_B[f]_B$. Dann gilt

$$f(v) = \lambda v \overset{7.18}{\iff} Ax = \lambda x \iff (A - \lambda 1_n)x = 0.$$

Daher lässt sich $E_\lambda(f)$ durch Lösen des homogenen Gleichungssystems $(A - \lambda 1_n)x = 0$ berechnen. Wir sprechen dann auch von Eigenwerten, Eigenräumen und Eigenvektoren von A. Da ähnliche Matrizen den selben Endomorphismus (bzgl. verschiedener Basen) beschreiben, haben sie die gleichen Eigenwerte (aber nicht die gleichen Eigenvektoren).

(c) Aus Lemma 7.7 und Bemerkung 7.9 folgt: $f \in \mathrm{End}(V)$ ist genau dann ein Isomorphismus, wenn 0 *kein* Eigenwert von f ist.

Beispiel 8.4. Sei $f \in \mathrm{End}(\mathbb{R}^3)$ mit

$$A := [f] = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Zieht man $\lambda = 1$ auf der Hauptdiagonale ab, so erhält man eine Matrix vom Rang 1 mit drei identischen Zeilen. Offenbar bilden $b_1 := (1, 0, -1)$ und $b_2 := (0, 1, -1)$ eine Basis von $E_1(f)$. Insbesondere hat

[1] Nach Bemerkung 6.20 findet man stets Basen B, C von V, sodass ${}_C[f]_B = \mathrm{diag}(1_r, 0_{n-r})$ gilt. Das Produkt solcher Matrizen lässt sich allerdings nicht sinnvoll interpretieren.

$\lambda = 1$ geometrische Vielfachheit 2. Da die Zeilensummen von A konstant sind, ist $b_3 := (1, 1, 1)$ ein Eigenvektor zum Eigenwert $\lambda = 4$. Nun ist $B := \{b_1, b_2, b_3\}$ eine Basis von $\mathbb{R}^3$ mit ${}_B[f]_B = \mathrm{diag}(1, 1, 4)$. Damit berechnet man leicht ${}_B[f \circ f \circ f]_B = {}_B[f]_B^3 = \mathrm{diag}(1, 1, 4)^3 = \mathrm{diag}(1, 1, 64)$.

8.2 Diagonalisierbarkeit

Definition 8.5. Man nennt $f \in \mathrm{End}(V)$ *diagonalisierbar*, falls eine Basis B von V existiert, sodass ${}_B[f]_B$ eine Diagonalmatrix ist. Eine Matrix $A \in K^{n\times n}$ heißt *diagonalisierbar*, falls A zu einer Diagonalmatrix ähnlich ist.

Bemerkung 8.6. Offenbar ist $f \in \mathrm{End}(V)$ genau dann diagonalisierbar, wenn V eine Basis aus Eigenvektoren von f besitzt. Eine Matrix A ist genau dann diagonalisierbar, wenn die entsprechende lineare Abbildung $f\colon K^{n\times 1} \to K^{n\times 1}$, $x \mapsto Ax$ diagonalisierbar ist (Folgerung 7.27).

Beispiel 8.7.

(a) Die Abbildung aus Beispiel 8.4 ist diagonalisierbar.

(b) Diagonalmatrizen sind offensichtlich diagonalisierbar.

(c) Sei $A := \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in K^{2\times 2}$. Da $A - \lambda 1_2$ für $\lambda \neq 0$ vollen Rang hat, ist $\lambda = 0$ der einzige Eigenwert von A. Wegen $E_0(A) = \langle (1, 0)\rangle$ existiert keine Basis aus Eigenvektoren und A ist *nicht* diagonalisierbar.

(d) Sei $A := \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right) \in \mathbb{Q}^{2\times 2}$ und $\lambda \in \mathbb{Q}$. Dann ist

$$A - \lambda 1_2 = \begin{pmatrix} -\lambda & 2 \\ 1 & -\lambda \end{pmatrix} \begin{matrix} \leftarrow\!\urcorner^{+} \\ \lrcorner_{\lambda} \end{matrix} \sim \begin{pmatrix} 0 & 2 - \lambda^2 \\ 1 & -\lambda \end{pmatrix}.$$

Wegen $\sqrt{2} \notin \mathbb{Q}$ (Beispiel 1.11) ist $2 - \lambda^2 \neq 0$. Also besitzt A keinen Eigenwert über $\mathbb{Q}$ und kann nicht diagonalisierbar sein. Andererseits ist A als $\mathbb{R}^{2\times 2}$-Matrix diagonalisierbar (Beispiel 8.13).

Definition 8.8. Wir haben in Definition 4.2 die (direkte) Summe $U + W$ (bzw. $U \oplus W$) von zwei Unterräumen $U, W \leq V$ eingeführt. Für $U_1, \ldots, U_n \leq V$ definiert man induktiv

$$U_1 + \ldots + U_n := (U_1 + \ldots + U_{n-1}) + U_n \leq V.$$

Offenbar besteht $U_1 + \ldots + U_n$ aus den Elementen der Form $u_1 + \ldots + u_n$ mit $u_i \in U_i$ für $i = 1, \ldots, n$. Wir nennen die Summe *direkt* und schreiben $U_1 \oplus \ldots \oplus U_n$, falls $U_1 + \ldots + U_{n-1} = U_1 \oplus \ldots \oplus U_{n-1}$ und $(U_1 + \ldots + U_{n-1}) \cap U_n = \{0\}$. Handlicher ist die folgende Charakterisierung.

Lemma 8.9. *Für Unterräume $U_1, \ldots, U_n$ eines Vektorraums V sind die folgenden Aussagen äquivalent:*

(1) $U_1 + \ldots + U_n = U_1 \oplus \ldots \oplus U_n$.

(2) $\dim(U_1 + \ldots + U_n) = \dim(U_1) + \ldots + \dim(U_n)$.

(3) Die Abbildung $U_1 \times \ldots \times U_n \to U_1 + \ldots + U_n$, $(u_1, \ldots, u_n) \mapsto u_1 + \ldots + u_n$ ist ein Isomorphismus.

(4) Jedes Element $u \in U_1 + \ldots + U_n$ lässt sich eindeutig in der Form $u = u_1 + \ldots + u_n$ mit $u_i \in U_i$ für $i = 1, \ldots, n$ schreiben.

(5) Ist $u_1 + \ldots + u_n = 0$ mit $u_i \in U_i$ für $i = 1, \ldots, n$, so folgt $u_1 = \ldots = u_n = 0$.

Beweis.

(1) ⇒ (2): Für $n = 1$ ist (2) trivial. Induktiv dürfen wir annehmen, dass (2) bereits für $n - 1$ gilt. Aus dem Dimensionssatz folgt dann

$$\dim(U_1 + \ldots + U_n) = \dim(U_1 + \ldots + U_{n-1}) + \dim(U_n) = \dim(U_1) + \ldots + \dim(U_n).$$

(2) ⇒ (3): Die gegebene Abbildung ist stets linear und surjektiv. Wegen

$$\dim(U_1 \times \ldots \times U_n) \overset{7.11}{=} \dim(U_1) + \ldots + \dim(U_n) = \dim(U_1 + \times + U_n)$$

muss sie auch injektiv sein.

(3) ⇒ (4): Ergibt sich aus der Injektivität von $U_1 \times \ldots \times U_n \to U_1 + \ldots + U_n$.

(4) ⇒ (5): Die beiden Zerlegungen des Nullvektors $u_1 + \ldots + u_n = 0 + \ldots + 0$ müssen nach (4) identisch sein, d. h. $u_1 = \ldots = u_n = 0$.

(5) ⇒ (1): Für $n = 1$ ist nichts zu zeigen. Induktiv können wir $U_1 + \ldots + U_{n-1} = U_1 \oplus \ldots \oplus U_{n-1}$ annehmen, denn die Voraussetzung (5) überträgt sich auf $U_1, \ldots, U_{n-1}$. Sei nun $u = u_1 + \ldots + u_{n-1} \in (U_1 + \ldots + U_{n-1}) \cap U_n$. Dann ist

$$0 = u_1 + \ldots + u_{n-1} - u \in U_1 + \ldots + U_n$$

und (5) zeigt $u = 0$. Also gilt (1). □

Satz 8.10. *Seien $\lambda_1, \ldots, \lambda_k$ paarweise verschiedene Eigenwerte von $f \in \mathrm{End}(V)$. Dann gilt*

$$E_{\lambda_1}(f) + \ldots + E_{\lambda_k}(f) = E_{\lambda_1}(f) \oplus \ldots \oplus E_{\lambda_k}(f) \leq V. \tag{8.1}$$

Insbesondere ist $k \leq \dim V$.

Beweis. Induktion nach k: Für $k = 1$ ist nichts zu zeigen. Sei also $k \geq 2$ und (8.1) für $k - 1$ bereits bewiesen. Seien $v_i \in E_{\lambda_i}(f)$ mit $v_1 + \ldots + v_k = 0$. Dann gilt

$$\begin{aligned} 0 &= f(v_1 + \ldots + v_k) - \lambda_k(v_1 + \ldots + v_k) = \lambda_1 v_1 + \ldots + \lambda_k v_k - \lambda_k v_1 - \ldots - \lambda_k v_k \\ &= (\lambda_1 - \lambda_k)v_1 + \ldots + (\lambda_{k-1} - \lambda_k)v_{k-1} \in E_{\lambda_1}(f) \oplus \ldots \oplus E_{\lambda_{k-1}}(f). \end{aligned}$$

Aus Lemma 8.9 folgt $(\lambda_i - \lambda_k)v_i = 0$ für $i = 1, \ldots, k - 1$. Wegen $\lambda_i \neq \lambda_k$ gilt $v_1 = \ldots = v_{k-1} = 0$. Schließlich ist auch $v_k = v_1 + \ldots + v_k = 0$. Nun ergibt sich (8.1) aus Lemma 8.9. Die letzte Behauptung folgt aus $\dim E_{\lambda_i}(f) \geq 1$ für $i = 1, \ldots, k$. □

Bemerkung 8.11. Merkregel: Eigenvektoren zu verschiedenen Eigenwerten sind linear unabhängig.

Folgerung 8.12. *Besitzt $A \in K^{n \times n}$ genau n verschiedene Eigenwerte, so ist A diagonalisierbar.*

Beweis. Für die verschiedenen Eigenwerte $\lambda_1, \ldots, \lambda_n$ von A gilt

$$\dim\bigl(E_{\lambda_1}(A) \oplus \ldots \oplus E_{\lambda_n}(A)\bigr) = \dim(E_{\lambda_1}(A)) + \ldots + \dim(E_{\lambda_n}(A)) \geq n = \dim K^{n \times 1}$$

nach Satz 8.10. Dies zeigt $E_{\lambda_1}(A) \oplus \ldots \oplus E_{\lambda_n}(A) = K^{n \times 1}$. Insbesondere besitzt $K^{n \times 1}$ eine Basis aus Eigenvektoren von A. □

Beispiel 8.13.

(a) Die Einheitsmatrix zeigt, dass die Umkehrung von Folgerung 8.12 falsch ist. Wir leiten später eine genaue Charakterisierung der Diagonalisierbarkeit her (Satz 10.34).

(b) Die Matrix $A := \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$ aus Beispiel 8.7 besitzt die Eigenwerte $\pm\sqrt{2}$ und ist daher diagonalisierbar. Wir zeigen über den Umweg der Determinante, dass die Eigenwerte jeder Matrix Nullstellen von Polynomen mit Koeffizienten in K sind (Satz 10.32).

Definition 8.14. Man nennt $A = (a_{ij}) \in K^{n\times n}$ eine (*obere*[2]) *Dreiecksmatrix*, falls alle Einträge unterhalb der Hauptdiagonalen verschwinden, d. h. $a_{ij} = 0$ für alle $i > j$:

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

Gilt zusätzlich $a_{ii} = 0$ für $i = 1, \ldots, n$, so spricht man von einer *strikten* (oberen) Dreiecksmatrix.

Beispiel 8.15. Sei $A = (a_{ij}) \in K^{n\times n}$ eine obere Dreiecksmatrix. Für $\lambda \in \{a_{11}, \ldots, a_{nn}\}$ hat $A - \lambda 1_n$ nicht vollen Rang, denn beim Gauß-Algorithmus tritt ein Versatz der Zeilen auf (Bemerkung 6.20). Ist andererseits $\lambda \notin \{a_{11}, \ldots, a_{nn}\}$, so sind die Hauptdiagonaleinträge von $A - \lambda 1_n$ alle ungleich 0. Daher hat $A - \lambda 1_n$ vollen Rang. Dies zeigt, dass die Eigenwerte von A genau die Einträge auf der Hauptdiagonale sind. Insbesondere ist A diagonalisierbar, wenn $a_{11}, \ldots, a_{nn}$ paarweise verschieden sind.

[2]Analog definiert man *untere* Dreiecksmatrizen.

9 Determinanten

9.1 Rekursive Definition

Bemerkung 9.1.

(a) Mathematiker versuchen oft komplizierte Objekte (wie $f \in \mathrm{End}(V)$) durch einfachere (wie $\mathrm{rk}(f)$ oder $\mathrm{tr}(f)$) zu ersetzen, um wesentliche Informationen sichtbar zu machen. So haben wir in Lemma 5.15 gesehen, dass $\mathrm{rk}(f)$ Rückschluss über die Bijektivität von f liefert, sofern man $\dim V$ kennt. Man nennt solche Größen *Invarianten*, wenn sie unter „natürlichen“ Umformungen (wie Basiswechsel) unverändert bleiben. Wir definieren in diesem Abschnitt als weitere Invariante die *Determinante* $\det(f)$. Wir zeigen, dass f genau dann bijektiv ist, wenn $\det(f) \neq 0$ gilt (im Gegensatz zu $\mathrm{rk}(f)$ hängt diese Kriterium nicht mehr von $\dim(V)$ ab.[1])

(b) In der Maßtheorie versucht man möglichst vielen Mengen $S \subseteq \mathbb{R}^n$ ein „Volumen“ $\mathrm{vol}(S) \in \mathbb{R}_{\geq 0}$ zuzuordnen. Sei $f\colon \mathbb{R}^n \to \mathbb{R}^n$ linear. Die in (a) beschriebene Zahl $\det(f)$ misst wie sehr sich das Volumen durch Anwenden von f verändert, d. h. es gilt $\mathrm{vol}(f(S)) = |\det(f)|\,\mathrm{vol}(S)$ sofern $\mathrm{vol}(S)$ definiert ist. Dem n-dimensionalen *Hyperwürfel*

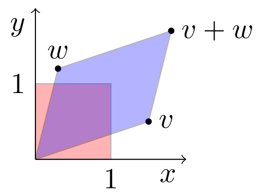
$$H := \big\{(x_1, \dots, x_n) \in \mathbb{R}^n : \forall i : 0 \leq x_i \leq 1\big\}$$

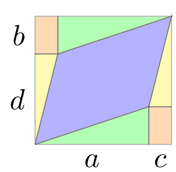
wird das Volumen $\mathrm{vol}(H) = 1$ zugewiesen. Daraus folgt

$$|\det(f)| = \mathrm{vol}(f(H)).$$

Das Vorzeichen von $\det(f)$ beschreibt, ob f orientierungserhaltend (Beispiel: Drehung) oder orientierungsumkehrend (Beispiel: Spiegelung) ist. Mehr dazu in Beispiel 11.22.

Beispiel 9.2. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ mit $v := f(e_1) = (a, b)$ und $w := f(e_2) = (c, d)$, d. h. $[f] = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$. Das Bild des (roten) Quadrats $H = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 1\}$ ist das von den Vektoren v und w aufgespannte (blaue) Parallelogramm $f(H)$:





Die Fläche von $f(H)$ ist

$$\mathrm{vol}(f(H)) = (a + c)(b + d) - 2bc - ab - cd = ad - bc.$$

Die Fläche ist genau dann 0, wenn v und w auf einer Geraden liegen, also linear abhängig sind. Dies ist äquivalent zu $\mathrm{rk}(f) \leq 1$.

[1] Aus dem realen Leben: Wenn Sie Probleme haben sich das Alter einer Person zu merken, merken Sie sich stattdessen das Geburtsjahr, denn diese Invariante ändert sich nicht jedes Jahr.

Definition 9.3. Sei $A = (a_{ij}) \in K^{n\times n}$ und $1 \le s,t \le n$. Durch Streichen der s-ten Zeile und t-ten Spalte von A entsteht die Matrix $A_{st} \in K^{(n-1)\times(n-1)}$. Die *Determinante*[2] von A ist rekursiv definiert:

$$\boxed{\det(A) := \begin{cases} a_{11} & \text{falls } n = 1, \\ \displaystyle\sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(A_{i1}) & \text{falls } n \ge 2. \end{cases}}$$

Beispiel 9.4.

(a) Für $n = 2$ erhält man

$$\det\begin{pmatrix} a & c \\ b & d \end{pmatrix} = a\det(A_{11}) - b\det(A_{21}) = ad - bc$$

(vgl. Beispiel 9.2).

(b) Für jede obere Dreiecksmatrix $A = (a_{ij})$ gilt $\det(A) = a_{11} \dots a_{nn}$. Dies ist klar für $n = 1$. Sei induktiv die Behauptung für $n-1$ bereits bewiesen. Da A_{11} auch eine obere Dreiecksmatrix ist, folgt

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(A_{i1}) = a_{11}\det(A_{11}) = a_{11}a_{22}\dots a_{nn}.$$

Da man mit dem Gauß-Algorithmus jede Matrix in eine obere Dreiecksmatrix überführen kann, untersuchen wir wie sich die Determinante bei elementaren Zeilenoperationen verändert.

Lemma 9.5. *Die Abbildung* $\det\colon K^{n\times n} \to K$ *ist linear in jeder Zeile, d. h. für* $a_1, \dots, a_n, b \in K^n$, $\lambda \in K$ *und* $1 \le k \le n$ *gilt*

$$\det\begin{pmatrix} a_1 \\ \vdots \\ \lambda a_k + b \\ \vdots \\ a_n \end{pmatrix} = \lambda\det\begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix} + \det\begin{pmatrix} a_1 \\ \vdots \\ b \\ \vdots \\ a_n \end{pmatrix}.$$

Beweis. Sei $a_i = (a_{i1}, \dots, a_{in})$ und $b = (b_1, \dots, b_n)$. Für $c \in K^n$ sei $M(c)$ die Matrix mit Zeilen $a_1, \dots, a_{k-1}, c, a_{k+1}, \dots, a_n$. Für $n = 1$ ist $k = 1$ und

$$\det\bigl(M(\lambda a_1 + b)\bigr) = \lambda a_{11} + b_1 = \lambda\det(M(a_1)) + \det(M(b))$$

wie behauptet. Sei nun $n \ge 2$ und die Behauptung für $n-1$ bereits bewiesen. Streichen der k-ten Zeile ergibt

$$M(\lambda a_k + b)_{k1} = M(a_k)_{k1} = M(b)_{k1}.$$

Es folgt

$$\begin{aligned}\det\bigl(M(\lambda a_k + b)\bigr) &= (-1)^{k+1}(\lambda a_{k1} + b_1)\det\bigl(M(\lambda a_k + b)_{k1}\bigr) + \sum_{i\ne k}(-1)^{i+1} a_{i1}\det\bigl(M(\lambda a_k + b)_{i1}\bigr) \\ &= \lambda(-1)^{k+1} a_{k1}\det(M(a_k)_{k1}) + (-1)^{k+1} b_1 \det(M(b)_{k1})\end{aligned}$$

[2] In manchen Büchern schreibt man $|A|$ anstelle von $\det(A)$. Das kann aber mit einer Matrixnorm verwechselt werden.

$$+\sum_{i\neq k}(-1)^{i+1}a_{i1}\big(\lambda\det(M(a_k)_{i1})+\det(M(b)_{i1})\big)$$
$$=\lambda\sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(M(a_k)_{i1})+(-1)^{k+1}b_1\det(M(b)_{k1})+\sum_{i\neq k}(-1)^{i+1}a_{i1}\det(M(b)_{i1})$$
$$=\lambda\det(M(a_k))+\det(M(b)).$$
□

Satz 9.6. *Für $A\in K^{n\times n}$ gilt:*

(a) Durch Multiplikation einer Zeile von A mit $\lambda\in K$ wird auch $\det(A)$ *mit λ multipliziert.*

(b) Vertauschen von zwei Zeilen von A ändert das Vorzeichen von $\det(A)$.

(c) Addieren eines Vielfachen einer Zeile zu einer anderen Zeile von A ändert $\det(A)$ *nicht.*

Beweis.

(a) Setzt man zunächst $\lambda=1$ und $b=0$ in Lemma 9.5, so sieht man, dass die Determinante verschwindet, wenn A eine Nullzeile besitzt. Die Behauptung folgt nun, indem man λ beliebig und $b=0$ in Lemma 9.5 wählt.

(b) Hier ist $n\geq 2$. Seien $a_1,\ldots,a_n$ die Zeilen von A und $s<t$. Vertauschen von a_s und a_t liefert die Matrix A'. Für $n=2$ ist $(s,t)=(1,2)$ und

$$\det(A')=\det\begin{pmatrix}a_{21}&a_{22}\\a_{11}&a_{12}\end{pmatrix}=a_{21}a_{12}-a_{22}a_{11}=-(a_{11}a_{22}-a_{12}a_{21})=-\det(A)$$

nach Beispiel 9.4. Sei nun die Behauptung für $n-1$ bereits bewiesen. Für $s\neq i\neq t$ entsteht A'_{i1} durch Zeilentausch aus A_{i1}. Also ist $\det(A'_{i1})=-\det(A_{i1})$. Andererseits entsteht A'_{s1} aus A_{t1} durch die Vertauschungen $a_s\leftrightarrow a_{s+1}\leftrightarrow a_{s+2}\leftrightarrow\ldots\leftrightarrow a_{t-1}$:

$$A_{t1}=\begin{pmatrix}\vdots\\a_s\\a_{s+1}\\\vdots\\a_{t-1}\\a_{t+1}\\\vdots\end{pmatrix}\sim\begin{pmatrix}\vdots\\a_{s+1}\\a_s\\a_{s+2}\\\vdots\\a_{t-1}\\a_{t+1}\\\vdots\end{pmatrix}\sim\ldots\sim\begin{pmatrix}\vdots\\a_{s+1}\\\vdots\\a_{t-1}\\a_s\\a_{t+1}\\\vdots\end{pmatrix}=A'_{s1}.$$

Dies zeigt $\det(A'_{s1})=(-1)^{t-s-1}\det(A_{t1})$. Analog ist $\det(A'_{t1})=(-1)^{t-s-1}\det(A_{s1})$. Wegen $(-1)^{t-s-1}=(-1)^{s-t-1}$ ergibt sich

$$\begin{aligned}\det(A')&=(-1)^{s+1}a_{t1}\det(A'_{s1})+(-1)^{t+1}a_{s1}\det(A'_{t1})+\sum_{i\notin\{s,t\}}(-1)^{i+1}a_{i1}\det(A'_{i1})\\&=(-1)^{t}a_{t1}\det(A_{t1})+(-1)^{s}a_{s1}\det(A_{s1})+\sum_{i\notin\{s,t\}}(-1)^{i}a_{i1}\det(A_{i1})\\&=-\sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(A_{i1})=-\det(A).\end{aligned}$$

(c) Wir addieren λa_k zu Zeile a_l mit $k \neq l$ und erhalten

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_l + \lambda a_k \\ \vdots \\ a_n \end{pmatrix} \overset{9.5}{=} \det(A) + \lambda \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_k \\ \vdots \end{pmatrix}.$$

Es genügt also $\det(A) = 0$ zu zeigen, falls A zwei identische Zeilen besitzt.[3] Für $n = 2$ gilt

$$\det(A) = \det\begin{pmatrix} a & b \\ a & b \end{pmatrix} = ab - ba = 0.$$

Sei nun $n \geq 3$. Nach (b) können wir annehmen, dass die ersten beiden Zeilen von A identisch sind. Dann hat A_{i1} für $i \geq 3$ ebenfalls zwei identische Zeilen. Mit Induktion nach n folgt

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(A_{i1}) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) = 0. \qquad \square$$

Beispiel 9.7.

$$\det\begin{pmatrix} -4 & -2 & -2 \\ 6 & 3 & 2 \\ 8 & 7 & 6 \end{pmatrix} \;\; | : (-2) \qquad = -2\det\begin{pmatrix} 2 & 1 & 1 \\ 6 & 3 & 2 \\ 8 & 7 & 6 \end{pmatrix} \begin{matrix} \cdot(-3)\;\cdot(-4) \\ + \\ + \end{matrix} \qquad = -2\det\begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 3 & 2 \end{pmatrix}$$

$$= 2\det\begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & -1 \end{pmatrix} \overset{9.4}{=} -12$$

Bemerkung 9.8. Für $A \in K^{n\times n}$ und $\lambda \in K$ gilt $\boxed{\det(\lambda A) = \lambda^n \det(A),}$ denn jede der n Zeilen wird mit λ multipliziert.

Lemma 9.9. *Sei* $A = \left(\begin{smallmatrix} A_1 & A_2 \\ 0 & A_3 \end{smallmatrix}\right) \in K^{(n+m)\times(n+m)}$ *mit* $A_1 \in K^{n\times n}$, $A_2 \in K^{n\times m}$ *und* $A_3 \in K^{m\times m}$. *Dann gilt* $\det(A) = \det(A_1)\det(A_3)$.

Beweis. Führt man den Gauß-Algorithmus an A durch, so wird zuerst A_1 und dann A_2 in eine obere Dreiecksmatrix umgeformt. Am Ende ist auch A eine obere Dreiecksmatrix und die Behauptung folgt. $\square$

9.2 Eigenschaften

Satz 9.10. *Für* $A \in K^{n\times n}$ *gilt*

$$A \textit{ invertierbar} \iff \operatorname{rk}(A) = n \iff \det(A) \neq 0.$$

[3] Für $K = \mathbb{Q}$ folgt dies sofort aus (b), aber nicht für $K = \mathbb{F}_2$.

Beweis. Die erste Äquivalenz stammt aus Lemma 5.15. Da die Zeilenstufenform $\widehat{A}$ eine obere Dreiecksmatrix ist, gilt

$$\det(A) \neq 0 \overset{9.6}{\iff} \det(\widehat{A}) \neq 0 \overset{9.4}{\iff} \widehat{A} = 1_n \iff \operatorname{rk}(A) = \operatorname{rk}(\widehat{A}) = n. \qquad \square$$

Satz 9.11 (Determinantensatz). *Für $A, B \in K^{n\times n}$ gilt* $\boxed{\det(AB) = \det(A)\det(B).}$

Beweis. Ist $\det(A) = 0$, so folgt $\operatorname{rk}(AB) \leq \operatorname{rk}(A) < n$ aus Lemma 5.15. Dann ist auch $\det(AB) = 0$ nach Satz 9.10. Wir können also $A \in \mathrm{GL}(n, K)$ annehmen. Nach Folgerung 6.18 ist A ein Produkt von Elementarmatrizen, sagen wir $A = A_1 \dots A_k$. Sei $M \in K^{n\times n}$ beliebig. Für die drei Arten von elementaren Zeilenoperationen gilt jeweils

$$\det(A_i M) = \begin{Bmatrix} \lambda \det(M) \\ -\det(M) \\ \det(M) \end{Bmatrix} = \det(A_i 1_n)\det(M) = \det(A_i)\det(M)$$

nach Satz 9.6. Insgesamt folgt

$$\begin{aligned} \det(AB) &= \det(A_1 \dots A_k B) = \det(A_1)\det(A_2 \dots A_k B) = \dots = \det(A_1)\dots\det(A_k)\det(B) \\ &= \dots = \det(A_1)\det(A_2 \dots A_k)\det(B) = \det(A)\det(B). \end{aligned} \qquad \square$$

Folgerung 9.12.

(a) Für $A \in K^{n\times n}$ gilt $\boxed{\det(A^{\mathrm{t}}) = \det(A).}$

(b) Für $A \in \mathrm{GL}(n, K)$ gilt $\boxed{\det(A^{-1}) = \det(A)^{-1}.}$

(c) Ähnliche Matrizen haben die gleiche Determinante.

Beweis.

(a) Wegen $\operatorname{rk}(A) = \operatorname{rk}(A^{\mathrm{t}})$ können wir annehmen, dass A invertierbar ist (anderenfalls ist $\det(A) = 0 = \det(A^{\mathrm{t}})$). Wieder ist A ein Produkt von Elementarmatrizen $A = A_1 \dots A_k$. Für die ersten beiden Zeilenoperationen gilt $A_i^{\mathrm{t}} = A_i$. Für die dritte Zeilenoperation ist $\det(A_i) = 1 = \det(A_i^{\mathrm{t}})$. Dies zeigt

$$\begin{aligned} \det(A^{\mathrm{t}}) &\overset{5.8}{=} \det(A_k^{\mathrm{t}} \dots A_1^{\mathrm{t}}) = \det(A_k^{\mathrm{t}})\dots\det(A_1^{\mathrm{t}}) = \det(A_k)\dots\det(A_1) \\ &= \det(A_1)\dots\det(A_k) = \det(A_1 \dots A_k) = \det(A). \end{aligned}$$

(b) Die Behauptung folgt aus $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(1_n) = 1$.

(c) Für $A \in K^{n\times n}$ und $S \in \mathrm{GL}(n, K)$ gilt

$$\det(SAS^{-1}) = \det(S)\det(A)\det(S^{-1}) = \det(S)\det(S)^{-1}\det(A) = \det(A). \qquad \square$$

Definition 9.13. Nach dem Determinantensatz und Folgerung 9.12 bilden die Matrizen mit Determinante 1 eine Untergruppe $\mathrm{SL}(n, K) \leq \mathrm{GL}(n, K)$. Man nennt $\mathrm{SL}(n, K)$ die *spezielle lineare Gruppe* vom Grad n über K. Es gilt $\mathrm{SL}(n, \mathbb{F}_2) = \mathrm{GL}(n, \mathbb{F}_2)$.

Bemerkung 9.14. Wegen $\det(A^{\mathrm{t}}) = \det(A)$ darf man bei der Berechnung von $\det(A)$ auch elementare Spaltenoperationen benutzen.

Definition 9.15. Sei B eine Basis des K-Vektorraums V. Für $f \in \operatorname{End}(V)$ definieren wir die *Determinante* von f durch $\det(f) := \det({}_B[f]_B)$. Nach Folgerung 7.27 und Folgerung 9.12 hängt $\det(f)$ nicht von der Wahl von B ab.

9.3 Laplace-Entwicklung

Satz 9.16 (Laplace-Entwicklung). *Sei $n \geq 2$ und $A \in K^{n \times n}$. Für $1 \leq k \leq n$ gilt*

$$\boxed{\det(A) = \sum_{i=1}^{n} (-1)^{i+k} a_{ik} \det(A_{ik}) = \sum_{i=1}^{n} (-1)^{i+k} a_{ki} \det(A_{ki}).}$$

Beweis. Seien $a_1, \ldots, a_n$ die Spalten von A. Nach Bemerkung 9.14 gilt

$$\begin{aligned}
&\det\begin{pmatrix} \cdots & a_{k-1} & a_k & \cdots \end{pmatrix} = -\det\begin{pmatrix} \cdots & a_{k-2} & a_k & a_{k-1} & \cdots \end{pmatrix} = \ldots \\
&= (-1)^{k-1} \det\begin{pmatrix} a_k & a_1 & \cdots & a_{k-1} & a_{k+1} & \cdots \end{pmatrix} = (-1)^{k-1} \sum_{i=1}^{n} (-1)^{i+1} a_{ik} \det(A_{ik}).
\end{aligned}$$

Die zweite Gleichung folgt aus der ersten, indem man $\det(A^{\mathrm{t}}) = \det(A)$ benutzt. $\square$

Bemerkung 9.17. Die Gleichungen in Satz 9.16 nennt man *Entwicklung nach der k-ten Spalte/Zeile*. Die Vorzeichen $(-1)^{i+k}$ verteilen sich schachbrettartig:

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Beispiel 9.18. Wie die meisten rekursiven Verfahren ist auch die Laplace-Entwicklung in der Regel ineffizient. Sie eignet sich jedoch für sogenannte *dünnbesetzte* Matrizen, d. h. wenn viele Einträge 0 sind. Wir entwickeln zuerst nach der dritten Zeile und anschließend nach der zweiten Spalte:

$$\det\begin{pmatrix} 1 & 2 & -3 & 0 \\ -2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 1 & 0 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 & 0 \\ -2 & 0 & 2 \\ -2 & 0 & 0 \end{pmatrix} = -2\det\begin{pmatrix} -2 & 2 \\ -2 & 0 \end{pmatrix} = -2 \cdot 4 = -8$$

Bemerkung 9.19. Für eine Folge von Zahlen $a_1, \ldots, a_n \in K$ definiert das Produkt $\prod_{i=1}^{n} a_i = a_1 \cdot \ldots \cdot a_n$ analog zum Summenzeichen $\sum$.

Satz 9.20 (Vandermonde). *Für $x_1, \ldots, x_n \in K$ nennt man*

$$A := (x_i^{j-1}) = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} \in K^{n \times n}$$

VANDERMONDE-Matrix.[4] *Es gilt*

$$\det(A) = \prod_{1\le i<j\le n} (x_j - x_i).$$

Insbesondere ist A genau dann invertierbar, wenn $x_1, \ldots, x_n$ paarweise verschieden sind.

Beweis. Induktion nach n: Im Fall $n = 1$ ist $A = 1_1$ und $\prod_{i<j}(x_j - x_i)$ ist das leere Produkt, welches man als 1 interpretiert (so wie die leere Summe als 0 interpretiert wird). Sei also $n \ge 2$. Wir subtrahieren das x_1-Fache der vorletzten Spalte von der letzten Spalte. Anschließend subtrahieren wir das x_1-Fache der $(n-2)$-ten Spalte von der vorletzten Spalte usw. Dadurch erhält man die Matrix

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & x_2 - x_1 & (x_2 - x_1)x_2 & \cdots & (x_2 - x_1)x_2^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n - x_1 & (x_n - x_1)x_n & \cdots & (x_n - x_1)x_n^{n-2} \end{pmatrix}$$

mit der selben Determinante (Satz 9.6). Durch Entwicklung nach der ersten Zeile kann man zur kleineren Matrix

$$\begin{pmatrix} x_2 - x_1 & (x_2 - x_1)x_2 & \cdots & (x_2 - x_1)x_2^{n-2} \\ \vdots & \vdots & & \vdots \\ x_n - x_1 & (x_n - x_1)x_n & \cdots & (x_n - x_1)x_n^{n-2} \end{pmatrix}$$

übergehen. Die Faktoren $(x_k - x_1)$ können für $k = 2, \ldots, n$ aus der Determinante herausgezogen werden. Dies ergibt

$$\det(A) = (x_2 - x_1)(x_3 - x_1)\ldots(x_n - x_1)\det\big((x_{i+1}^{j-1})_{i,j=1}^{n-1}\big).$$

Nun folgt die Behauptung mit Induktion. $\square$

Definition 9.21. Für $A \in K^{n\times n}$ nennt man

$$\widetilde{A} := \begin{cases} 1_1 & \text{falls } n = 1 \\ \big((-1)^{i+j}\det(A_{ji})\big)_{i,j} & \text{falls } n > 1 \end{cases} \qquad \in K^{n\times n}$$

die zu A *komplementäre* Matrix.[5]

Satz 9.22. *Für alle $A \in K^{n\times n}$ gilt* $\boxed{A\widetilde{A} = \det(A)1_n = \widetilde{A}A.}$ *Insbesondere ist* $\boxed{A^{-1} = \frac{1}{\det(A)}\widetilde{A},}$ *falls $A \in \mathrm{GL}(n,K)$.*

Beweis. Für $n = 1$ ist die Behauptung klar. Sei also $n \ge 2$. Sei B_{kl} die Matrix, die aus A entsteht, indem man die l-te Zeile durch die k-te Zeile ersetzt. Für $k \ne l$ hat B_{kl} zwei identische Zeilen und es folgt $\det(B_{kl}) = 0$. Andererseits ist $B_{kk} = A$. Sei $A\widetilde{A} = (c_{ij})$. Entwicklung nach der l-ten Zeile von B_{kl} ergibt

$$\delta_{kl}\det(A) = \det(B_{kl}) = \sum_{i=1}^{n} a_{ki}(-1)^{i+l}\det(A_{li}) = c_{kl}.$$

Dies zeigt $A\widetilde{A} = \det(A)1_n$. Die Gleichung $\widetilde{A}A = \det(A)1_n$ zeigt man analog durch Entwicklung nach einer Spalte. $\square$

[4]In manchen Büchern betrachtet man die transponierte Matrix.
[5]auch *Adjunkte* genannt; Verwechslungsgefahr mit der *adjungierten* Matrix A^* aus Satz 13.7

Beispiel 9.23. Für jede invertierbare 2×2-Matrix $A := \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ erhält man

$$A^{-1} = \frac{1}{\det(A)} \widetilde{A} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Bemerkung 9.24. Die Formel $A^{-1} = \frac{1}{\det(A)} \widetilde{A}$ hat für „große“ n mehr theoretische als praktische Bedeutung (nutzen Sie Satz 6.17 zur Berechnung von A^{-1}). Ist beispielsweise $A \in \mathbb{Z}^{n \times n}$, so ist auch $\det(A)A^{-1} = \widetilde{A} \in \mathbb{Z}^{n \times n}$. Insbesondere ist $A^{-1} \in \mathbb{Z}^{n \times n}$, falls $\det(A) = \pm 1$. Aus dem Gauß-Algorithmus ist diese Beobachtung nicht ersichtlich. Der nächste Satz liefert eine ähnliche Aussage für Gleichungssysteme.

Satz 9.25 (CRAMERsche Regel)**.** *Sei $A \in \mathrm{GL}(n, K)$ und $b \in K^{n \times 1}$. Für $k = 1, \ldots, n$ sei A_k die Matrix, die aus A entsteht, indem man die k-te Spalte durch b ersetzt. Für die eindeutige Lösung $x = (x_1, \ldots, x_n)^{\mathrm{t}}$ des Gleichungssystems $Ax = b$ gilt dann $x_k = \frac{\det(A_k)}{\det(A)}$ für $k = 1, \ldots, n$.*

Beweis. Wir benutzen Satz 9.22 und entwickeln A_k nach der k-ten Spalte:

$$(\det(A_k))_k = \Big(\sum_{i=1}^{n} (-1)^{i+k} \det(A_{ik}) b_i\Big)_k = \widetilde{A} b = \widetilde{A} A x = \det(A) x = (\det(A) x_k)_k. \qquad \square$$

9.4 Die Leibniz-Formel

Bemerkung 9.26. Führt man die Laplace-Entwicklung für $n \times n$-Matrizen bis auf 1×1-Matrizen zurück, so erhält man die Determinante als Summe von $n(n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$ Termen. Wir bestimmen diese Terme explizit.

Definition 9.27. Sei $n \in \mathbb{N}$ und $N := \{1, \ldots, n\}$. Eine Bijektion der Form $N \to N$ heißt *Permutation* von N. Die Menge der Permutationen von N wird mit S_n bezeichnet.

Bemerkung 9.28.

(a) Analog zu $\mathrm{GL}(V)$ ist auch S_n eine Gruppe bzgl. Komposition von Abbildungen. Wir werden daher das Kompositionszeichen $\circ$ oft einsparen. Man nennt S_n die *symmetrische Gruppe von Grad n.*

(b) Sei $\sigma \in S_n$. Für die Wahl von $\sigma(1) \in \{1, \ldots, n\}$ gibt es n Möglichkeiten. Da σ injektiv ist, gilt $\sigma(2) \neq \sigma(1)$. Für die Wahl von $\sigma(2)$ verbleiben also noch $n - 1$ Möglichkeiten usw. Insgesamt hat man $n!$ Möglichkeiten eine Permutation zu definieren, d. h. $|S_n| = n!$.

Beispiel 9.29.

(a) Für $k \geq 2$ nennt man $\sigma \in S_n$ einen (k-)*Zyklus* (oder Zyklus der *Länge* k), falls paarweise verschiedene $1 \leq a_1, \ldots, a_k \leq n$ existieren, sodass

$$\sigma(x) = \begin{cases} a_{i+1} & \text{falls } x = a_i \text{ mit } i < k, \\ a_1 & \text{falls } x = a_k, \\ x & \text{sonst.} \end{cases}$$

$$\begin{tikzcd} & a_1 \arrow[dl, "\sigma"'] & \\ a_2 \arrow[dr, "\sigma"'] & & a_4 \arrow[ul, "\sigma"'] \\ & a_3 \arrow[ur, "\sigma"'] & \end{tikzcd}$$

Man schreibt dann $\sigma = (a_1, \dots, a_k)$. Diese Schreibweise ist eindeutig bis auf „Rotation“, d. h.

$$\sigma = (a_2, \dots, a_k, a_1) = \ldots = (a_k, a_1, \dots, a_{k-1}).$$

Die Komposition von Zyklen geschieht wie bei Abbildungen üblich von rechts nach links:

$$(1,3,4,5) \circ (3,4,5) \circ (1,3,2) = (1,5,4)(2,3).$$

Außerdem ist $(a_1, \dots, a_k)^{-1} = (a_k, a_{k-1}, \dots, a_1)$.

(b) Zyklen der Länge 2 heißen *Transpositionen*. Eine Transposition vertauscht also zwei Elemente und lässt alle anderen Elemente fest. Jeder k-Zyklus ist eine Komposition von $k-1$ Transpositionen:

$$(a_1, \dots, a_k) = (a_1, a_2)(a_2, a_3) \dots (a_{k-1}, a_k).$$

Allerdings gibt es viele Möglichkeiten für eine solche Komposition.

(c) In S_3 ist jedes Element ein Zyklus: $S_3 = \{1, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}$.

(d) Zyklen $\sigma = (a_1, \dots, a_k)$ und $\tau = (b_1, \dots, b_l)$ heißen *disjunkt*, falls

$$\{a_1, \dots, a_k\} \cap \{b_1, \dots, b_l\} = \varnothing.$$

Gegebenenfalls gilt $\sigma\tau = \tau\sigma$.

Satz 9.30. *Jede Permutation $\sigma \in S_n$ ist eine Komposition von paarweise disjunkten Zyklen $\sigma = \sigma_1 \dots \sigma_k$. Dabei sind $\sigma_1, \dots, \sigma_k$ bis auf die Reihenfolge eindeutig bestimmt.*

Beweis. Sei $\sigma \in S_n$. Im Fall $\sigma = \mathrm{id}$ ist σ das leere Produkt von Zyklen. Sei also $\sigma \neq \mathrm{id}$ und

$$a_1 := \min\{1 \le i \le n : \sigma(i) \neq i\}.$$

Sei $a_i := \sigma^{i-1}(a_1)$ für $i \ge 2$. Wegen $n < \infty$ existieren $i < j$ mit $a_i = a_j$, d. h. $\sigma^{j-i}(a_1) = a_1$. Daher existiert

$$k := \min\{1 \le i \le n : \sigma^i(a_1) = a_1\}.$$

Wegen $k \le j - i$ sind die Elemente $a_1, \dots, a_k$ paarweise verschieden. Also ist $\sigma_1 := (a_1, \dots, a_k)$ ein k-Zyklus. Für $\rho := \sigma\sigma_1^{-1}$ gilt

$$\rho(x) = \begin{cases} x & \text{falls } x \in \{a_1, \dots, a_k\}, \\ \sigma(x) & \text{sonst.} \end{cases}$$

Im Fall $\rho = \mathrm{id}$ ist $\sigma = \sigma_1$ und wir sind fertig. Anderenfalls können wir das Verfahren mit ρ anstatt σ wiederholen. Da mit jeder Wiederholung die Anzahl der Fixpunkte von σ wächst, erreicht man nach endlich vielen Schritten $\sigma = \sigma_1 \dots \sigma_k$ mit paarweise disjunkten Zyklen $\sigma_1, \dots, \sigma_k$.

Sei auch $\sigma = \tau_1 \dots \tau_l$ eine Komposition von paarweise disjunkten Zyklen $\tau_1, \dots, \tau_l$. Dann existiert ein i mit $\tau_i(a_1) \neq a_1$. Da die Zyklen disjunkt sind, gilt $\tau_i(a_1) = \sigma(a_1) = \sigma_1(a_1) = a_2$, $\tau_i(a_2) = \sigma_1(a_2) = a_3$ usw. Also ist $\tau_i = \sigma_1$ und $\sigma_2 \dots \sigma_k = \tau_1 \dots \tau_{i-1}\tau_{i+1} \dots \tau_l$. Die Eindeutigkeit der σ_i folgt nun mit Induktion nach k. □

Bemerkung 9.31.

(a) Da nach Beispiel 9.29 jeder Zyklus eine Komposition von Transpositionen ist, ist sogar jede Permutation eine Komposition von (in Regel nicht disjunkten) Transpositionen.

(b) Man kann die Schreibweise in disjunkte Zyklen

$$\sigma = (a_1, \dots, a_s)(b_1, \dots, b_t) \dots$$

vollständig eindeutig machen, indem man $a_1 = \min\{a_1, \dots, a_s\} < b_1 = \min\{b_1, \dots, b_t\} < \dots$ fordert.

Definition 9.32. Für $\sigma \in S_n$ nennt man

$$P_\sigma := (\delta_{i\sigma(j)})_{i,j} \in \mathbb{Q}^{n\times n}$$

die *Permutationsmatrix* von σ. Außerdem heißt $\mathrm{sgn}(\sigma) := \det(P_\sigma)$ das *Signum* oder *Vorzeichen* von σ.

Bemerkung 9.33. Die Permutationsmatrix von σ entsteht, indem man die Zeilen der Einheitsmatrix (also die Standardbasis $e_1, \dots, e_n$) gemäß σ permutiert. Mit dem Gauß-Algorithmus lässt sich diese Permutation durch endlich viele Zeilenvertauschungen realisieren (das entspricht dem Sortieralgorithmus *Selectionsort*). Wegen $\det(1_n) = 1$ ist $\mathrm{sgn}(\sigma) \in \{\pm 1\}$ also tatsächlich ein „Vorzeichen".

Beispiel 9.34. Wir bestimmen die Permutationsmatrix und das Signum für die Permutationen in S_3:

σ	id	$(1,2)$	$(1,3)$	$(2,3)$	$(1,2,3)$	$(1,3,2)$
P_σ	1_3	$\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}$	$\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}$	$\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}$	$\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$	$\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$
$\mathrm{sgn}(\sigma)$	1	-1	-1	-1	1	1

Satz 9.35. *Für $\sigma, \tau \in S_n$ gilt $P_{\sigma\circ\tau} = P_\sigma P_\tau$ und* $\boxed{\mathrm{sgn}(\sigma \circ \tau) = \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\tau).}$

Beweis. Der Eintrag von $P_\sigma P_\tau$ an Position (i, j) ist

$$\sum_{k=1}^{n} \delta_{i\sigma(k)}\delta_{k\tau(j)} = \delta_{i,\sigma(\tau(j))} = \delta_{i,(\sigma\circ\tau)(j)}.$$

Dies zeigt die erste Gleichung. Die zweite folgt aus dem Determinantensatz. □

Bemerkung 9.36.

(a) Die Permutationsmatrix einer Transposition ist genau die Elementarmatrix zur Vertauschung von Zeilen. Insbesondere hat jede Transposition Signum -1. Nach Satz 9.35 ist das Produkt einer geraden Anzahl an Transpositionen niemals ein Produkt einer ungeraden Anzahl an Transpositionen.

(b) Nach Beispiel 9.29 hat jeder k-Zyklus Signum $(-1)^{k-1}$. Ist σ ein Produkt von paarweise disjunkten Zyklen mit Längen $l_1, \dots, l_k$, so gilt

$$\boxed{\mathrm{sgn}(\sigma) = (-1)^{l_1+\dots+l_k-k}.}$$

Zum Beispiel ist

$$\mathrm{sgn}\big((1, 2, 5, 6)(3, 7)(4, 9, 8)\big) = (-1)^{4+2+3-3} = 1.$$

(c) Aus Satz 9.35 folgt, dass die Permutationen mit Signum 1 eine Untergruppe $A_n \leq S_n$ bilden. Man nennt A_n die *alternierende* Gruppe vom Grad n. In gewisser Weise verhält sich A_n zu S_n so wie $\mathrm{SL}(n, K)$ zu $\mathrm{GL}(n, K)$.

Satz 9.37 (LEIBNIZ-Formel)**.** *Für $A = (a_{ij}) \in K^{n \times n}$ gilt*

$$\boxed{\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}.}$$

Beweis. Die Zeilen $a_1, \ldots, a_n$ von A lassen sich als Linearkombination der Standardbasis ausdrücken: $a_i = \sum_{j=1}^n a_{ij} e_j$. Da det in jeder Zeile linear ist (Lemma 9.5), gilt

$$\det(A) = \sum_{i_1=1}^n a_{1i_1} \det \begin{pmatrix} e_{i_1} \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \sum_{i_1=1}^n a_{1i_1} \sum_{i_2=1}^n a_{2i_2} \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ a_3 \\ \vdots \end{pmatrix} = \ldots$$
$$= \sum_{1 \leq i_1, \ldots, i_n \leq n} a_{1i_1} \ldots a_{ni_n} \det \begin{pmatrix} e_{i_1} \\ \vdots \\ e_{i_n} \end{pmatrix}.$$

Existieren $s \neq t$ mit $i_s = i_t$, so verschwindet die entsprechende Determinante. Man muss also nur über die Tupel $(i_1, \ldots, i_n)$ mit paarweise verschiedenen Einträgen summieren. Jedes solche Tupel beschreibt eine Permutation $\sigma \in S_n$ mit $\sigma(j) = i_j$ mit $j = 1, \ldots, n$. Es folgt

$$\det(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \ldots a_{n\sigma(n)} \det(P_\sigma) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \ldots a_{n\sigma(n)}. \qquad \square$$

Folgerung 9.38 (SARRUS-Regel)**.** *Für 3×3-Matrizen gilt*

$$\boxed{\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - gec - hfa - idb.}$$

Beweis. Man benutze die Leibniz-Formel mit Beispiel 9.34. $\square$

Bemerkung 9.39.

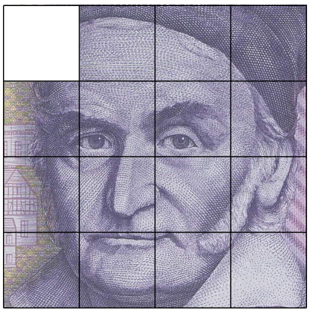
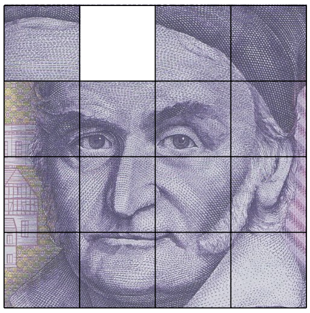
(a) Man kann sich die Sarrus-Regel mit folgendem Schema merken:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$

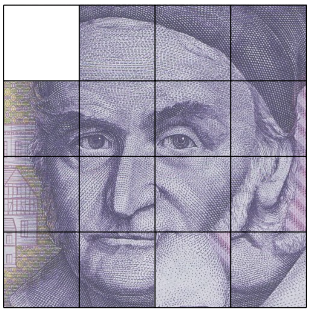
(b) Achtung: Die Sarrus-Regel gilt *nur* für 3×3-Matrizen (für 4×4-Matrizen braucht man $4! = 24$ Summanden).

(c) Sei $A \in K^{n\times n}$ mit Eigenwert λ. Dann ist $E_\lambda(A) = \mathrm{Ker}(A - \lambda 1_n) \neq \{0\}$ und $\det(A - \lambda 1_n) = 0$. Nach der Leibniz-Formel ist $\det(A - \lambda 1_n)$ ein Polynom in λ. Auf diese Weise werden wir alle Eigenwerte von A berechnen.

Beispiel 9.40. Das folgende *Schiebepuzzle* besteht aus 15 beweglichen Quadraten und einem leeren Feld. Ein Quadrat, welches horizontal oder vertikal an das leere Feld grenzt, darf auf dieses geschoben werden (vgl. Cover):

Sam Loyd bot ein Preisgeld von 1000 \$, wem es gelingt die folgende Konfiguration in den Ausgangszustand zu überführen:[6]



Jeder Zug entspricht einer Transposition in S_{16}. Legt man ein Schachbrettmuster zugrunde, so wandert das leere Feld bei jedem Zug von schwarz nach weiß oder umgekehrt. Da das leere Feld in Loyds Konfiguration in der Ausgangsstellung liegt, benötigt man zur Lösung eine gerade Anzahl an Zügen. Andererseits unterscheidet sich Loyds Konfiguration nur um eine Transposition vom Ausgangszustand. Nach Bemerkung 9.36 ist diese Konfiguration also unlösbar und Loyd musste das Preisgeld nie auszahlen.

[6] siehe [D. Slocum und J. Sonneveld, *The 15 puzzle*, The Slocum Puzzle Foundation, Beverly Hills, 2006]

Aufgaben

Aufgabe 1. Welche der folgenden Aussagen sind im Jahr 2025 wahr?

(a) Es gibt einen Monat mit 28 Tagen.

(b) Es gibt einen Monat mit genau 28 Tagen.

(c) Es gibt genau einen Monat mit 28 Tagen.

(d) Es gibt genau einen Monat mit genau 28 Tagen.

Aufgabe 2. Seien $1 \leq a \leq b \leq 9$ natürliche Zahlen. Der Logiker (S)iegfried kennt nur die Summe $a + b$, während sein Kollege (P)etrus nur das Produkt ab kennt. Die beiden führen folgenden Dialog:

S: „Ich kenne a und b nicht.“	P: „Ich kenne a und b nicht.“
S: „Ich kenne a und b nicht.“	P: „Ich kenne a und b nicht.“
S: „Ich kenne a und b nicht.“	P: „Ich kenne a und b nicht.“
S: „Ich kenne a und b nicht.“	P: „Ich kenne a und b nicht.“
S: „Ich kenne a und b nicht.“	P: „Jetzt kenne ich a und b!“

Bestimmen Sie daraus a und b.

Aufgabe 3. Für endliche Mengen A und B gilt $|A \cup B| = |A| + |B| - |A \cap B|$ nach Lemma 1.12. Finden und beweisen Sie eine analoge Gleichung für drei endliche Mengen.

Aufgabe 4. Beweisen Sie mit vollständiger Induktion: Die Summe der ersten n ungeraden Zahlen ist n^2.

Aufgabe 5. Beweisen Sie, dass $\mathbb{N} \times \mathbb{N}$ abzählbar ist.

Aufgabe 6. Konstruieren Sie Relationen mit folgenden Eigenschaften:

(a) reflexiv, aber weder symmetrisch noch transitiv.

(b) symmetrisch, aber weder reflexiv noch transitiv.

(c) transitiv, aber weder reflexiv noch symmetrisch.

Aufgabe 7. Sei $U := \{2z + 1 : z \in \mathbb{Z}\} \cup \{0\}$. Untersuchen Sie, ob $(U, +)$ eine Gruppe ist.

Aufgabe 8. Seien $(G, *)$ und $(H, \circ)$ Gruppen. Zeigen Sie, dass $G \times H$ mit der Verknüpfung

$$(g_1, h_1) \cdot (g_2, h_2) := (g_1 * g_2, h_1 \circ h_2)$$

zu einer Gruppe wird.

Aufgabe 9. Konstruieren Sie einen Körper mit drei Elementen.
Hinweis: Was ist $1 + 1$?

Aufgabe 10. Zeigen Sie:

(a) Eine nichtleere Teilmenge H einer Gruppe G ist genau dann eine Untergruppe, wenn $x, y \in H \Rightarrow xy^{-1} \in H$ gilt.

(b) Eine nichtleere Teilmenge U eines K-Vektorraums V genau dann ein Unterraum ist, wenn für alle $u, v \in U$ und $\lambda \in K$ gilt: $\lambda u + v \in U$.

Aufgabe 11. (Dedekind-Identität) Seien X, Y, Z Unterräume eines Vektorraums V mit $X \subseteq Z$. Zeigen Sie: $(X + Y) \cap Z = X + (Y \cap Z)$.

Aufgabe 12. Sei V ein Vektorraum, $S \subseteq V$ und $U, W \leq V$. Zeigen Sie:

(a) $\langle S \rangle$ ist der Durchschnitt aller Unterräume von V, die S enthalten.

(b) $U + W = \langle U \cup W \rangle$.

(c) $U \cup W \leq V \iff U \cup W \in \{U, W\}$.

Aufgabe 13. Seien u, v, w Vektoren eines Vektorraums V. Beweisen oder widerlegen Sie: Genau dann ist $\{u, v, w\}$ linear unabhängig, wenn $\{u, v\}$, $\{u, w\}$ und $\{v, w\}$ linear unabhängig sind.

Aufgabe 14. Offenbar ist $\mathbb{R}$ ein $\mathbb{Q}$-Vektorraum, in dem die Skalarmultiplikation mit der üblichen Multiplikation in $\mathbb{R}$ übereinstimmt (dies müssen Sie nicht prüfen). Zeigen Sie:

(a) 1 und $\sqrt{2}$ sind linear unabhängig über $\mathbb{Q}$.

(b) $\mathbb{Q}(\sqrt{2}) := \langle 1, \sqrt{2} \rangle = \mathbb{Q} + \mathbb{Q}\sqrt{2}$ ist ein Körper mit den gleichen Verknüpfungen wie in $\mathbb{R}$.

Aufgabe 15. Seien $A \in K^{n \times m}$, $B \in K^{m \times k}$ und $C \in K^{k \times l}$. Zur Berechnung von ABC kann man entweder $(AB)C$ oder $A(BC)$ wählen. Bestimmen Sie für beide Klammerungen die Anzahl der benötigten Multiplikationen von Elementen in K (in Abhängigkeit von n, m, k, l). Welche Variante ist zu bevorzugen, wenn $\frac{1}{n} + \frac{1}{k} < \frac{1}{m} + \frac{1}{l}$?

Aufgabe 16. Zeigen Sie:

(a) Das Vertauschen von zwei Zeilen einer Matrix lässt sich durch die beiden anderen elementaren Zeilenoperationen realisieren.

(b) Jede $n \times n$-Matrix ist ein Produkt von Matrizen der Form $1_n + \lambda E_{ij}$ mit $\lambda \in K$ und $1 \leq i, j \leq n$ (der Fall $i = j$ ist zugelassen).

Aufgabe 17. Sei $n = n_1 + \ldots + n_k$ und $\lambda_1, \ldots, \lambda_k \in K$ paarweise verschieden. Sei

$$A := \mathrm{diag}(\lambda_1 1_{n_1}, \ldots, \lambda_k 1_{n_k}) \in K^{n \times n}$$

und $B \in K^{n \times n}$. Zeigen Sie:

(a) Genau dann ist $AB = BA$, wenn $B_i \in K^{n_i \times n_i}$ für $i = 1, \ldots, k$ mit $B = \mathrm{diag}(B_1, \ldots, B_k)$ existieren.

(b) Genau dann ist $AB = BA$ für alle $B \in K^{n\times n}$, wenn A eine Skalarmatrix ist (d. h. $k = 1$).

Aufgabe 18. Seien $A \in \mathbb{Q}^{n\times m} \subseteq \mathbb{R}^{n\times m}$ und $b \in \mathbb{Q}^{n\times 1} \subseteq \mathbb{R}^{n\times 1}$. Begründen Sie:

(a) Der Rang von A über $\mathbb{Q}$ ist der Rang von A über $\mathbb{R}$.

(b) Ist A über $\mathbb{R}$ invertierbar, so auch über $\mathbb{Q}$.

(c) Besitzt das Gleichungssystem $Ax = b$ eine Lösung in $\mathbb{R}^{m\times 1}$, so existiert auch eine Lösung in $\mathbb{Q}^{m\times 1}$.

(d) Geben Sie ein Beispiel, in dem die Lösungsmengen von $Ax = b$ über $\mathbb{Q}$ und $\mathbb{R}$ unterschiedlich sind.

Aufgabe 19. Seien U, V, W Vektorräume und $f: U \to V$, $g: V \to W$ und $h: W \to X$ lineare Abbildungen.

(a) Zeigen Sie $\mathrm{rk}(g \circ f) + \mathrm{rk}(h \circ g) \leq \mathrm{rk}(g) + \mathrm{rk}(h \circ g \circ f)$ (Frobenius-Ungleichung).
Hinweis: Für $g(V) = g(f(U)) \oplus Y$ gilt $h(g(V)) = h\big(g(f(U))\big) + h(Y)$.

(b) Folgern Sie Lemma 5.15(a) aus Teil (a).

(c) Zeigen Sie $\mathrm{rk}(A) + \mathrm{rk}(B) \leq \mathrm{rk}(AB) + n$ für $A \in K^{m\times n}$ und $B \in K^{n\times k}$ (Sylvester-Ungleichung).

Aufgabe 20. Sei K ein Körper und $n \in \mathbb{N}$. Zeigen Sie:

(a) Die Summe und das Produkt von (strikten) oberen (bzw. unteren) Dreiecksmatrizen in $K^{n\times n}$ sind wieder (strikte) obere (bzw. untere) Dreiecksmatrizen (Definition 8.14).

(b) Die oberen (bzw. unteren) Dreiecksmatrizen bilden einen Unterraum U von $K^{n\times n}$. Berechnen Sie $\dim U$.

(c) Die Menge der invertierbaren oberen (bzw. unteren) Dreiecksmatrizen in $K^{n\times n}$ ist eine Untergruppe von $\mathrm{GL}(n, K)$.

Aufgabe 21. Sei V ein K-Vektorraum. Für $\varphi \in \mathrm{GL}(V)$ und $v \in V$ sei $f_{\varphi,v}: V \to V$, $w \mapsto \varphi(w) + v$. Die Abbildungen der Form $f_{\mathrm{id}_V,v}$ nennt man *Translationen.* Zeigen Sie:

(a)
$$\mathrm{Aff}(V) := \{f_{\varphi,v} : \varphi \in \mathrm{GL}(V), v \in V\} \subseteq \mathrm{Abb}(V, V)$$
ist eine Gruppe bzgl. Komposition von Abbildungen. Handelt es sich um eine Untergruppe von $\mathrm{GL}(V)$?

(b) Die Translationen bilden eine Untergruppe von $\mathrm{Aff}(V)$.

Bemerkung: Man nennt $\mathrm{Aff}(V)$ die *affine* Gruppe von V.

Aufgabe 22. Sei $A \in \mathrm{GL}(n, K)$. Zeigen Sie, dass man die Matrix $\binom{A}{1_n} \in K^{2n\times n}$ durch elementare Spaltenoperationen in die Form $\binom{1_n}{B}$ überführen kann. Dabei ist $B = A^{-1}$.

Aufgabe 23. Seien V, W Vektorräume und $f \in \mathrm{Hom}(V, W)$. Zeigen Sie:

(a) Genau dann ist f injektiv, wenn ein $g \in \mathrm{Hom}(W, V)$ mit $g \circ f = \mathrm{id}_V$ existiert.

(b) Genau dann ist f surjektiv, wenn ein $g \in \mathrm{Hom}(W, V)$ mit $f \circ g = \mathrm{id}_W$ existiert.

(c) Sind die Abbildungen g jeweils eindeutig bestimmt?

Aufgabe 24. Sei $\lambda \in K$ ein Eigenwert von $A \in K^{n\times n}$ und $k \in \mathbb{N}$. Zeigen Sie, dass λ^k ein Eigenwert von A^k ist. Gilt auch die Umkehrung? Zeigen Sie, dass λ^{-1} ein Eigenwert von A^{-1} ist, falls A invertierbar ist.

Aufgabe 25. Sei $A \in K^{n\times n}$ diagonalisierbar. Zeigen Sie, dass auch A^{t} diagonalisierbar ist mit den gleichen Eigenwerten. Stimmen auch die Eigenräume überein?

Aufgabe 26. Seien $a, b \in K$ und

$$A = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix} \in K^{n\times n}.$$

Zeigen Sie:

$$\det(A) = (a-b)^{n-1}(a+(n-1)b).$$

Hinweis: Eigenwerte.

Aufgabe 27. Ist das Schiebepuzzle auf dem Cover lösbar?

Aufgabe 28. Transpositionen der Form $(a, a+1) \in S_n$ nennt man *Basistransposition*. Zeigen Sie, dass jede Permutation ein Produkt von Basistranspositionen ist.
Bemerkung: Dies ist die Grundlage von *Bubblesort*.

Aufgabe 29. Zeigen Sie $|A_n| = \frac{n!}{2}$ für $n \geq 2$.
Hinweis: Wenden Sie die Leibniz-Formel auf die Matrix $(1)_{i,j=1}^n \in \mathbb{Q}^{n\times n}$ an.

Aufgabe 30. Seien $\sigma, \tau \in S_n$ mit $\operatorname{sgn}(\sigma) \neq \operatorname{sgn}(\tau)$. Zeigen Sie $\det(P_\sigma + P_\tau) = 0$.
Hinweis: $P_\sigma + P_\tau = P_\sigma(P_\sigma^{-1} + P_\tau^{-1})P_\tau$.

Aufgabe 31. Sei $\sigma \in S_n$. Zeigen Sie

$$\operatorname{sgn}(\sigma) = \prod_{1\leq i<j\leq n} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Hinweis: Betrachten Sie zunächst Transpositionen σ.

Lineare Algebra II

10 Polynome

10.1 Der Vektorraum der Polynome

Bemerkung 10.1. In Bemerkung 9.39 haben wir angedeutet, dass die Eigenwerte einer Matrix A Lösungen gewisser (nicht-linearer) Polynomgleichungen sind. Wir definieren in dem Kapitel Polynome mit Koeffizienten in einem beliebigen Körper und untersuchen deren Nullstellen. Daraus leiten wir ein notwendiges und hinreichendes Kriterium für die Diagonalisierbarkeit eines Endomorphismus ab.

Definition 10.2. Ein (formales) *Polynom* über einem Körper K in der *Variablen* X ist eine Summe der Form

$$\alpha = \sum_{k=0}^{d} a_k X^k = a_0 + a_1 X + \ldots + a_d X^d$$

mit *Koeffizienten* $a_0, \ldots, a_d \in K$.[1]

- Man nennt a_0 das *Absolutglied* von α.
- Sofern nicht alle Koeffizienten 0 sind, nennt man

$$\deg(\alpha) := \max\{d \in \mathbb{N}_0 : a_d \neq 0\}$$

  den *Grad* von α und a_d den *führenden* Koeffizienten.[2] Im Fall $a_d = 1$ heißt α *normiert*.
- Für das *Nullpolynom* (alle Koeffizienten sind 0) setzt man $\deg(0) := -\infty$.
- Die Menge aller Polynome über K wird mit $K[X]$ bezeichnet.

Bemerkung 10.3.

(a) Kennt man den Grad von $\alpha \in K[X]$ nicht, so schreibt man $\alpha = \sum_{k=0}^{\infty} a_k X^k = \sum a_k X^k$ unter der Annahme, dass nur endlich viele Koeffizienten ungleich 0 sind.

(b) Polynome werden als gleich angesehen, wenn sie die gleichen Koeffizienten haben, d. h.

$$\sum_{k=0}^{\infty} a_k X^k = \sum_{k=0}^{\infty} b_k X^k \iff \forall k \in \mathbb{N}_0 : a_k = b_k.$$

(c) Die Körperelemente $\lambda \in K$ werden mit den *konstanten* Polynomen $\lambda X^0 \in K[X]$ identifiziert. Dies sind genau die Polynome vom Grad ≤ 0. Insbesondere gilt $0, 1 \in K \subseteq K[X]$.

Beispiel 10.4. Das Polynom $\alpha = X^2 - 3X + 1 \in \mathbb{Q}[X]$ ist normiert vom Grad 2 mit Absolutglied 1.

[1] Formal: Ein Polynom ist eine Abbildung $\mathbb{N}_0 \to K$, $k \mapsto a_k$ mit $|\{k \in \mathbb{N}_0 : a_k \neq 0\}| < \infty$.
[2] auch *Leitkoeffizient* genannt

Satz 10.5. *Mit den Verknüpfungen*

$$\sum a_k X^k + \sum b_k X^k := \sum (a_k + b_k) X^k,$$
$$\lambda \sum a_k X^k := \sum (\lambda a_k) X^k$$

wird $K[X]$ ein unendlich-dimensionaler K-Vektorraum mit Basis $1, X, X^2, \ldots$.

Beweis. Seien $\alpha = \sum a_k X^k$ und $\beta = \sum b_k X^k$ mit $d := \deg(\alpha) \geq \deg(\beta)$. Man kann $(a_0, \ldots, a_d)$ und $(b_0, \ldots, b_d)$ als Vektoren in K^{d+1} ansehen. Die Verknüpfungen in $K[X]$ entsprechen genau denen in K^{d+1}. Daher erfüllt $K[X]$ die Vektorraumaxiome. Nach Definition ist jedes Polynom eine endliche Linearkombination von $1, X, X^2, \ldots$, d. h. $K[X] = \langle 1, X, X^2, \ldots \rangle$. Aus der Eindeutigkeit der Koeffizienten (Bemerkung 10.3) folgt die lineare Unabhängigkeit von $\{1, X, X^2, \ldots\}$. □

Bemerkung 10.6.

(a) Für $\alpha, \beta \in K[X]$ und $\lambda \in K$ gilt offenbar $\deg(\alpha + \beta) \leq \max\{\deg(\alpha), \deg(\beta)\}$ und $\deg(\lambda\alpha) \leq \deg(\alpha)$. Daher bilden die Polynome vom Grad kleiner d einen d-dimensionalen Unterraum mit Basis $1, X, \ldots, X^{d-1}$.

(b) Sie wissen vermutlich, dass man Polynome auch multiplizieren kann, z. B.

$$\begin{aligned}(2X^3 - X^2 + 5X - 1)(4X^2 + 3) &= 8X^5 - 4X^4 + (6 + 20)X^3 + (-3 - 4)X^2 + 15X - 3 \\ &= 8X^5 - 4X^4 + 26X^3 - 7X^2 + 15X - 3\end{aligned}$$

Dies lässt sich wie folgt formalisieren.

Satz 10.7. *Für Polynome $\alpha = \sum a_k X^k$, $\beta = \sum b_k X^k$ ist*

$$\alpha \cdot \beta := \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} a_l b_{k-l} \Big) X^k$$

ein Polynom vom Grad $\deg(\alpha) + \deg(\beta)$. Es gelten folgende Rechenregeln:

$$\alpha\beta = \beta\alpha, \qquad \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma, \qquad \alpha(\beta\gamma) = (\alpha\beta)\gamma.$$

Beweis. Im Fall $\alpha = 0$ oder $\beta = 0$ ist $\alpha\beta = 0$ und $\deg(\alpha\beta) = -\infty = \deg(\alpha) + \deg(\beta)$. Sei also $d := \deg(\alpha) \geq 0$ und $e := \deg(\beta) \geq 0$. Für $k > d + e$ ist $\sum_{l=0}^{k} a_l b_{k-l} = 0$ und $\deg(\alpha\beta) \leq d + e$. Für $k = d + e$ ist $\sum_{l=0}^{k} a_l b_{k-l} = a_d b_e \neq 0$. Dies zeigt $\deg(\alpha\beta) = d + e$. Insbesondere ist $\alpha\beta \in K[X]$. Für $\gamma = \sum c_k X^k$ gilt

$$\alpha\beta = \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} a_l b_{k-l} \Big) X^k = \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} b_l a_{k-l} \Big) X^k = \beta\alpha$$

$$\alpha(\beta + \gamma) = \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} a_l (b_{k-l} + c_{k-l}) \Big) X^k = \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} a_l b_{k-l} \Big) X^k + \sum_{k=0}^{\infty} \Big(\sum_{l=0}^{k} a_l c_{k-l} \Big) X^k = \alpha\beta + \alpha\gamma.$$

Der Koeffizient von X^k in $\alpha(\beta\gamma)$ ist

$$\sum_{l=0}^{k} a_l \sum_{m=0}^{k-l} b_m c_{k-l-m} = \sum_{\substack{r,s,t \in \mathbb{N}_0 \\ r+s+t=k}} a_r b_s c_t = \sum_{l=0}^{k} \Big(\sum_{m=0}^{l} a_m b_{l-m} \Big) c_{k-l}.$$

Dies ist auch der Koeffizient von X^k in $(\alpha\beta)\gamma$. Also ist $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. □

Bemerkung 10.8. Im Gegensatz zur Matrizenmultiplikation ist die Multiplikation von Polynomen kommutativ. Das einzige Körperaxiom, welches $K[X]$ nicht erfüllt, ist die Existenz von Inversen. Zum Beispiel existiert kein $\alpha \in K[X]$ mit $X \cdot \alpha = 1$. Dennoch gilt die Kürzungsregel: $\alpha\beta = \alpha\gamma \Rightarrow \beta = \gamma$, falls $\alpha \neq 0$. Dies folgt aus

$$\deg(\beta - \gamma) \leq \deg(\alpha) + \deg(\beta - \gamma) = \deg(\alpha(\beta - \gamma)) = \deg(0) = -\infty.$$

Man kann in $K[X]$ also wie in $\mathbb{Z}$ rechnen.

Satz 10.9 (Division mit Rest)**.** *Für $\alpha, \beta \in K[X]$ mit $\beta \neq 0$ existieren eindeutig bestimmte Polynome $\gamma, \delta \in K[X]$ mit $\alpha = \beta\gamma + \delta$ und $\deg\delta < \deg\beta$.*

Beweis. Existenz: Wähle $\gamma \in K[X]$, sodass

$$\delta := \alpha - \beta\gamma = a_d X^d + a_{d-1}X^{d-1} + \ldots + a_0$$

möglichst kleinen Grad $d \in \mathbb{N}_0 \cup \{-\infty\}$ hat. Sei $\beta = b_e X^e + \ldots + b_0$ und $e := \deg\beta$. Gilt $d \geq e$, so ist

$$a_d b_e^{-1} X^{d-e}\beta = a_d b_e^{-1}(b_e X^d + b_{e-1}X^{d-1} + \ldots + b_0 X^{d-e}) = a_d X^d + \ldots$$

und es folgt

$$\deg\bigl(\alpha - \beta(\gamma + a_d b_e^{-1} X^{d-e})\bigr) = \deg\bigl(\delta - a_d b_e^{-1} X^{d-e}\beta\bigr) < d.$$

Dies ist ein Widerspruch zur Wahl von γ. Also ist $d < e$ und $\alpha = \beta\gamma + \delta$.

Eindeutigkeit: Sei nun $\alpha = \beta\tilde{\gamma} + \tilde{\delta}$ mit $\tilde{\gamma}, \tilde{\delta} \in K[X]$ und $\deg\tilde{\delta} < e$. Nach Satz 10.7 ist

$$e + \deg(\tilde{\gamma} - \gamma) = \deg(\beta) + \deg(\tilde{\gamma} - \gamma) = \deg(\beta(\tilde{\gamma} - \gamma)) = \deg(\delta - \tilde{\delta}) \leq \max\{\deg(\delta), \deg(\tilde{\delta})\} < e.$$

Es folgt $\deg(\tilde{\gamma} - \gamma) = -\infty = \deg(\delta - \tilde{\delta})$. Dies zeigt $\tilde{\gamma} = \gamma$ und $\tilde{\delta} = \delta$. $\square$

Definition 10.10. In der Situation von Satz 10.9 nennt man δ den *Rest* bei der Division von α durch β. Im Fall $\delta = 0$ nennt man β einen *Teiler* von α und schreibt $\beta \mid \alpha$. Ggf. sagt man auch „β *teilt* α“ oder „α ist durch β *teilbar*“.

Beispiel 10.11.

$$\begin{array}{rrrrl}
(2X^3 & -X^2 & +5X & +1) & : (X^2+3) = 2X - 1 =: \gamma \\
-(2X^3 & & +6X) & & \\
 & -X^2 & -X & +1 & \\
 & -(-X^2 & & -3) & \\
 & & -X & +4 =: \delta &
\end{array}$$

Also $\alpha = 2X^3 - X^2 + 5X + 1 = (X^2+3)(2X-1) - X + 4 = \beta\gamma + \delta$ mit $\deg\delta = 1 < 2 = \deg\beta$.

Bemerkung 10.12. Die Division durch normierte Polynome vom Grad 1 lässt sich mit dem Horner-*Schema*[3] effizient gestalten. Sei dazu $\alpha = a_n X^n + a_{n-1}X^{n-1} + \ldots + a_0$ und $\beta = X - b$. Wir berechnen $c_n := 0$, $c_k := a_{k+1} + bc_{k+1}$ für $k = n-1, \ldots, 0$ und $d := a_0 + bc_0$:

$$\begin{array}{cccccc}
 & a_n & a_{n-1} & a_{n-2} & \cdots & a_0 \\
+ & 0 & bc_{n-1} & bc_{n-2} & \cdots & bc_0 \\
\hline
 & c_{n-1} & c_{n-2} & \cdots & c_0 & d
\end{array}$$

[3]auch *Ruffinis Regel* genannt

Für $\gamma := c_{n-1}X^{n-1} + \ldots + c_0$ gilt nun

$$\begin{aligned}\beta\gamma + d &= c_{n-1}X^n + (c_{n-2} - bc_{n-1})X^{n-1} + \ldots + (c_0 - bc_1)X - bc_0 + d \\ &= a_nX^n + a_{n-1}X^{n-1} + \ldots + a_0 = \alpha.\end{aligned}$$

Beispiel 10.13. Für $\alpha = 2X^3 - X^2 + 3X + 1$ und $\beta = X - 2$ erhält man:

$$\begin{array}{r|rrrr} & 2 & -1 & 3 & 1 \\ + & 0 & 4 & 6 & 18 \\ \hline & 2 & 3 & 9 & 19 \end{array}$$

Dies zeigt $\alpha = \beta(2X^2 + 3X + 9) + 19$.

10.2 Nullstellen

Definition 10.14. Sei $\alpha = \sum_{k=0}^{d} a_kX^k \in K[X]$. Man kann ein Element $x \in K$ für X in α *einsetzen*:

$$\alpha(x) := \sum_{k=0}^{d} a_kx^k \in K.$$

Man nennt x eine *Nullstelle* von α, falls $\alpha(x) = 0$.

Lemma 10.15. *Für $\alpha, \beta \in K[X]$ und $x \in K$ gilt*

$$\begin{aligned}(\alpha + \beta)(x) &= \alpha(x) + \beta(x), \\ (\alpha\beta)(x) &= \alpha(x)\beta(x).\end{aligned}$$

Beweis. Seien $\alpha = \sum a_kX^k$ und $\beta = \sum b_kX^k$. Dann gilt

$$\begin{aligned}(\alpha + \beta)(x) &= \sum (a_k + b_k)x^k = \sum a_kx^k + \sum b_kx^k = \alpha(x) + \beta(x), \\ (\alpha\beta)(x) &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_kb_{n-k}x^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n}(a_kx^k)(b_{n-k}x^{n-k}) = \sum a_kx^k \sum b_kx^k = \alpha(x)\beta(x).\end{aligned}$$ □

Bemerkung 10.16. Merkregel: Es ist egal, ob Sie erst addieren/multiplizieren und danach einsetzen oder erst einsetzen und danach addieren/multiplizieren. Achtung: Im Allgemeinen ist $\alpha(x+y) \neq \alpha(x) + \alpha(y)$ und $\alpha(xy) \neq \alpha(x)\alpha(y)$ für $\alpha \in K[X]$ und $x, y \in K$.

Satz 10.17 (Interpolation)**.** *Seien $x_1, \ldots, x_n \in K$ paarweise verschieden und $y_1, \ldots, y_n \in K$ beliebig. Dann existiert genau ein Polynom α vom Grad $< n$ mit $\alpha(x_i) = y_i$ für $i = 1, \ldots, n$.*

Beweis. Sei $\alpha = a_0 + a_1X + \ldots + a_{n-1}X^{n-1}$. Die Bedienung $\alpha(x_i) = y_i$ für $i = 1, \ldots, n$ bedeutet:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Die Koeffizientenmatrix dieses Gleichungssystems ist eine Vandermonde-Matrix. Da die x_i paarweise verschieden sind, ist die Matrix nach Satz 9.20 invertierbar. Also existiert genau eine Lösung $(a_0, \ldots, a_{n-1})$. □

Bemerkung 10.18. Eine explizite Lösung der Interpolationsaufgabe ist durch das LAGRANGE-*Polynom*

$$\alpha := \sum_{i=1}^{n} y_i \prod_{j \neq i} \frac{X - x_j}{x_i - x_j} \in K[X]$$

gegeben (nachrechnen).

Beispiel 10.19.

(a) Für $n = 2$ und $K = \mathbb{R}$ ist Satz 10.17 die geometrische Aussage, dass zwei verschiedene Punkte im $\mathbb{R}^2$ durch genau eine Gerade verbunden sind.

(b) Wir suchen ein Polynom $\alpha \in \mathbb{R}[X]$ durch die Punkte $(-1, 2)$, $(0, 1)$ und $(1, 3)$. Der Beweis von Satz 10.17 führt auf das Gleichungssystem

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

mit der eindeutigen Lösung $\alpha = 1 + \frac{1}{2}X + \frac{3}{2}X^2$.

Folgerung 10.20.

(a) Jedes Polynom $\alpha \in K[X]$ vom Grad $d \geq 0$ besitzt höchstens d Nullstellen.

(b) Sei $|K| = \infty$ und $\alpha, \beta \in K[X]$ mit $\alpha(x) = \beta(x)$ für alle $x \in K$. Dann gilt $\alpha = \beta$.

Beweis.

(a) Angenommen α besitzt paarweise verschiedene Nullstellen $x_1, \ldots, x_{d+1} \in K$. Nach Satz 10.17 mit $y_1 = \ldots = y_{d+1} = 0$ ist α das einzige Polynom vom Grad $\leq d$ mit diesen Nullstellen. Andererseits hat das Nullpolynom auch diese Nullstellen. Also gilt $\alpha = 0$ und $d = -\infty$ im Widerspruch zur Annahme.

(b) Wegen $|K| = \infty$ besitzt $\alpha - \beta$ unendlich viele Nullstellen. Aus (a) folgt $\alpha = \beta$. □

Bemerkung 10.21. Für $K = \mathbb{R}$ ist jedes Polynom $\alpha \in \mathbb{R}[X]$ eindeutig durch die (stetige) Funktion $\mathbb{R} \to \mathbb{R}$, $x \mapsto \alpha(x)$ bestimmt, denn $|K| = \infty$. In der Analysis unterscheidet man daher nicht zwischen Polynom und Funktion. Über endlichen Körpern K würde man dabei Information verlieren, denn es gibt nur endlich viele Abbildungen $K \to K$, aber unendlich viele Polynome. Zum Beispiel entsprechen die Polynome $X, X^2, \ldots \in \mathbb{F}_2[X]$ alle der Identität $\mathrm{id}_{\mathbb{F}_2}$.

Lemma 10.22. *Genau dann ist $x \in K$ eine Nullstelle von α, wenn $(X - x) \mid \alpha$.*

Beweis. Division mit Rest liefert $\gamma, \delta \in K[X]$ mit $\alpha = (X - x)\gamma + \delta$ und $\deg \delta < \deg(X - x) = 1$, d. h. $\delta \in K$. Nun ist

$$\delta = \delta(x) = \big(\alpha - (X - x)\gamma\big)(x) \overset{10.15}{=} \alpha(x) - (x - x)\gamma(x) = \alpha(x).$$ □

Definition 10.23. Sei $x \in K$ eine Nullstelle von α. Man nennt $X - x$ einen *Linearfaktor* von α. Die größte Zahl $e \in \mathbb{N}$ mit $(X - x)^e \mid \alpha$ nennt man die *algebraische Vielfachheit* der Nullstelle x. Im Fall $e = 1$ spricht man von einer *einfachen* Nullstelle und anderenfalls von einer *mehrfachen* Nullstelle.

Lemma 10.24. *Seien $x_1, \ldots, x_n \in K$. Dann hat jeder normierte Teiler von $(X-x_1)\ldots(X-x_n) \in K[X]$ die Form $(X - x_{i_1})\ldots(X - x_{i_k})$ mit $1 \le i_1 < \ldots < i_k \le n$.*

Beweis. Im Fall $n = 1$ sind 1 (das leere Produkt mit $k = 0$) und $X - x_1$ die einzigen normierten Teiler. Sei also $n \ge 2$ und die Behauptung für $n - 1$ bereits bewiesen. Seien $\alpha, \beta \in K[X]$ mit $(X - x_1)\ldots(X - x_n) = \alpha\beta$. Dann ist $\alpha(x_n)\beta(x_n) = (\alpha\beta)(x_n) = 0$, o. B. d. A. sei $\alpha(x_n) = 0$. Nach Lemma 10.22 gilt $\alpha = (X - x_n)\gamma$ für ein $\gamma \in K[X]$. Nach Bemerkung 10.8 darf man $X - x_n$ kürzen und erhält $(X - x_1)\ldots(X - x_{n-1}) = \gamma\beta$. Die Behauptung folgt nun durch Induktion. □

Beispiel 10.25.

(a) Sei $\alpha = X^3 + X^2 - 5X + 3 \in \mathbb{R}[X]$. Eine Nullstelle $x \in \mathbb{R}$ ist eine Lösung der Gleichung

$$x^3 + x^2 - 5x + 3 = 0.$$

Auch wenn es Lösungsformeln für solche Gleichungen (dritten und vierten Grades[4]) gibt, sind diese in der Praxis aufwendig. Wir werden unsere Beispiele (und Übungsaufgaben) daher so wählen, dass man „kleine" ganzzahlige Nullstellen erraten kann. Angenommen es gibt eine Nullstelle $x \in \mathbb{Z}$. Wegen $x(x^2 + x - 5) = -3$ ist x ein Teiler von 3, d. h. $x \in \{\pm 1, \pm 3\}$. Man prüft leicht, dass $x_1 = 1$ tatsächlich eine Nullstelle ist ($1^3 + 1^2 - 5 \cdot 1 + 3 = 0$). Polynomdivision (zum Beispiel mit dem Horner-Schema) ergibt

$$(X^3 + X^2 - 5X + 3) : (X - 1) = X^2 + 2X - 3 =: \gamma.$$

Für jede Nullstelle $y \in \mathbb{R}$ von γ gilt nun $\alpha(y) = (y-1)\gamma(y) = 0$, d. h. y ist auch eine Nullstelle von α. Mit der p-q-Formel $\frac{1}{2}(-p \pm \sqrt{p^2 - 4q})$ für quadratische Gleichungen erhält man die Nullstellen von γ:

$$x_2 = \frac{1}{2}\big(-2 + \sqrt{4 + 12}\big) = 1, \qquad x_3 = \frac{1}{2}\big(-2 - \sqrt{4 + 12}\big) = -3.$$

Daher ist $x_1 = x_2 = 1$ eine Nullstelle von α mit algebraischer Vielfachheit 2 (eine *doppelte* Nullstelle). Außerdem *zerfällt* α in Linearfaktoren $\alpha = (X - 1)^2(X + 3)$.

(b) Offensichtlich ist $\alpha(0)$ das Absolutglied von $\alpha \in K[X]$. Also ist $x = 0$ genau dann eine Nullstelle von α, wenn das Absolutglied von α verschwindet.

(c) Bekanntlich besitzt $X^2 + 1 \in \mathbb{R}[X]$ keine Nullstelle. Wir konstruieren später einen „größeren" Körper, über dem auch dieses Polynom in Linearfaktoren zerfällt (Lemma 11.27).

(d) Das Polynom $X^2 + X + 1 \in \mathbb{F}_2[X]$ hat keine Nullstelle in $\mathbb{F}_2$, denn es kommen nur 0 und 1 in Frage.

10.3 Charakteristische Polynome

Bemerkung 10.26. Im Folgenden betrachten wir Matrizen mit Einträgen in $K[X]$. Aufgrund der Rechenregeln für Polynome (Satz 10.7) überlegt man sich leicht, dass die gewohnten Rechenregeln (Lemma 5.8) für Matrizen auch in $K[X]^{n\times n}$ gelten. Schließlich kann man sogar die Definition der Determinante auf Matrizen in $K[X]^{n\times n}$ anwenden (dabei werden Matrixeinträge nur addiert und multipliziert, aber niemals dividiert). Ebenso bleiben der Determinantensatz, die Laplace-Entwicklung, die Leibniz-Formel und der Satz 9.22 über die komplementäre Matrix in dieser größeren Allgemeinheit richtig. Andererseits funktioniert der Gauß-Algorithmus in $K[X]^{n\times n}$ nicht, denn hier muss dividiert werden.

[4] siehe Algebra-Skript

Definition 10.27. Für $A = (a_{ij}) \in K^{n\times n}$ betrachten wir die Matrix

$$X1_n - A = \begin{pmatrix} X - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & X - a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & -a_{n-1,n} \\ -a_{n1} & \cdots & -a_{n,n-1} & X - a_{nn} \end{pmatrix} \in K[X]^{n\times n}.$$

Man nennt $\chi_A := \det(X1_n - A) \in K[X]$ das *charakteristische Polynom* von A.[5]

Lemma 10.28. *Ähnliche Matrizen haben das gleiche charakteristische Polynom.*

Beweis. Für $A \in K^{n\times n}$ und $S \in \mathrm{GL}(n, K)$ gilt

$$\chi_{SAS^{-1}} = \det(X1_n - SAS^{-1}) = \det(S(X1_n - A)S^{-1}) \overset{10.28}{=} \det(X1_n - A) = \chi_A. \qquad \square$$

Definition 10.29. Sei V ein n-dimensionaler K-Vektorraum mit Basis B. Für $f \in \mathrm{End}(V)$ definiert man $\chi_f := \det(X1_n - {}_B[f]_B)$. Nach Folgerung 7.27 und Lemma 10.28 hängt χ_f nicht von der Wahl von B ab. In den nachfolgenden Sätzen kann man Matrizen also durch Endomorphismen ersetzen (und umgekehrt).

Beispiel 10.30. Das charakteristische Polynom von $A := \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right) \in \mathbb{Q}[X]$ ist

$$\chi_A = \det \begin{pmatrix} X - 1 & -2 \\ -3 & X - 4 \end{pmatrix} = (X-1)(X-4) - (-2)(-3) = X^2 - 5X - 2 = X^2 - \mathrm{tr}(A)X + \det(A).$$

Lemma 10.31. *Für $A \in K^{n\times n}$ gilt $\chi_A = X^n - \mathrm{tr}(A)X^{n-1} + \ldots + (-1)^n \det(A)$. Insbesondere ist χ_A normiert vom Grad n.*

Beweis. Sei $A = (a_{ij})$. Nach der Leibniz-Formel gilt

$$\begin{aligned} \chi_A &= \det(X1_n - A) = (X - a_{11})(X - a_{22}) \ldots (X - a_{nn}) \\ &\quad + \sum_{\sigma \in S_n \setminus \{\mathrm{id}\}} \mathrm{sgn}(\sigma)(\delta_{1\sigma(1)}X - a_{1\sigma(1)}) \ldots (\delta_{n\sigma(n)}X - a_{n\sigma(n)}). \end{aligned}$$

Für $\sigma \in S_n \setminus \{\mathrm{id}\}$ existiert ein $k \in \{1, \ldots, n\}$ mit $l := \sigma(k) \neq k$. Da σ injektiv ist, gilt $\sigma(l) \neq \sigma(k) = l$. Daher ist $\delta_{k\sigma(k)} = 0 = \delta_{l\sigma(l)}$ und

$$(\delta_{1\sigma(1)}X - a_{1\sigma(1)}) \ldots (\delta_{n\sigma(n)}X - a_{n\sigma(n)})$$

ist ein Polynom vom Grad $\le n - 2$. Insgesamt ist

$$\chi_A = (X - a_{11})(X - a_{22}) \ldots (X - a_{nn}) + \alpha$$

mit $\deg(\alpha) \le n-2$. Ausmultiplizieren zeigt $\chi_A = X^n - (a_{11} + \ldots + a_{nn})X^{n-1} + \ldots = X^n - \mathrm{tr}(A)X^{n-1} + \ldots$. Zur Berechnung des Absolutglieds setzt man $X = 0$ und erhält $\chi_A(0) \overset{10.15}{=} \det(-A) = (-1)^n \det(A)$ aus Bemerkung 9.8. $\square$

[5] In manchen Büchern definiert man χ_A durch $\det(A - X1_n) = (-1)^n \det(X1_n - A)$. Das macht keinen großen Unterschied, aber bringt den Nachteil, dass χ_A nicht normiert ist, wenn n ungerade ist.

Satz 10.32. *Die Eigenwerte von $A \in K^{n\times n}$ sind die Nullstellen von χ_A.*

Beweis. Es gilt

$$\operatorname{Ker}(A-\lambda 1_n) \neq \{0\} \iff \det(A-\lambda 1_n)=0 \overset{9.8}{\iff} \det(\lambda 1_n - A)=0 \overset{10.15}{\iff} \chi_A(\lambda)=0. \qquad \square$$

Lemma 10.33. *Sei $\lambda \in K$ ein Eigenwert von $f \in \operatorname{End}(V)$. Dann ist die geometrische Vielfachheit von λ höchstens so groß wie die algebraische Vielfachheit von λ als Nullstelle von χ_f.*

Beweis. Man ergänze eine Basis $b_1,\ldots,b_e$ von $E_\lambda(f)$ zu einer Basis $B := \{b_1,\ldots,b_n\}$ von V. Dann gilt

$$\chi_f = \det(X1_n - {}_B[f]_B) = \det\begin{pmatrix}(X-\lambda)1_e & * \\ 0 & *\end{pmatrix}.$$

Lemma 9.9 zeigt $\chi_f = (X-\lambda)^e\beta$ für ein $\beta \in K[X]$. Also ist die algebraische Vielfachheit von λ als Nullstelle von χ_f mindestens e. $\square$

Satz 10.34. *Genau dann ist $f \in \operatorname{End}(V)$ diagonalisierbar, wenn χ_f in Linearfaktoren zerfällt und für jede Nullstelle von χ_f die algebraische Vielfachheit mit der geometrischen Vielfachheit übereinstimmt.*

Beweis. Seien $\lambda_1,\ldots,\lambda_k \in K$ die verschiedenen Nullstellen von χ_f mit algebraischen Vielfachheiten $m_1,\ldots,m_k$. Dann existiert ein $\alpha \in K[X]$ mit $\chi_f = (X-\lambda_1)^{m_1}\ldots(X-\lambda_k)^{m_k}\alpha$. Sei m_i' die geometrische Vielfachheit von λ_i als Eigenwert von f. Nach Lemma 10.33 gilt

$$m_1'+\ldots+m_k' \le m_1+\ldots+m_k+\deg(\alpha) \overset{10.7}{=} \deg\big((X-\lambda_1)^{m_1}\ldots(X-\lambda_k)^{m_k}\alpha\big) = \deg(\chi_f) = \dim V.$$

Nach Satz 8.10 besitzt V genau dann eine Basis aus Eigenvektoren, wenn $m_1'+\ldots+m_k' = \dim V$. Dies gilt genau dann, wenn χ_f in Linearfaktoren zerfällt (d. h. $\alpha = 1$) und die algebraischen Vielfachheiten mit den geometrischen Vielfachheiten übereinstimmen (d. h. $m_i = m_i'$ für $i = 1,\ldots,k$). $\square$

Bemerkung 10.35. Zerfällt χ_A in Linearfaktoren, so gilt

$$\chi_A = (X-\lambda_1)\ldots(X-\lambda_n) = X^n - (\lambda_1+\ldots+\lambda_n)X^{n-1} + \ldots + (-1)^n\lambda_1\ldots\lambda_n.$$

Ein Vergleich mit Lemma 10.31 zeigt:

$$\boxed{\begin{aligned}\operatorname{tr}(A) &= \lambda_1+\ldots+\lambda_n,\\ \det(A) &= \lambda_1\ldots\lambda_n,\end{aligned}}$$

d. h. die Spur ist die Summe der Eigenwerte und die Determinante ist das Produkt der Eigenwerte (sofern diese existieren). Hat man bereits $\lambda_1,\ldots,\lambda_{n-1}$ bestimmt, so erhält man $\lambda_n = \operatorname{tr}(A) - \lambda_1 - \ldots - \lambda_{n-1}$.

Satz 10.36 (Mirsky)**.** *Sei $d_1,\ldots,d_n,\lambda_1,\ldots,\lambda_n \in K$ mit $d_1+\ldots+d_n = \lambda_1+\ldots+\lambda_n$. Dann existiert eine Matrix $A \in K^{n\times n}$ mit Hauptdiagonale $d_1,\ldots,d_n$ und Eigenwerten $\lambda_1,\ldots,\lambda_n$.*

Beweis. Im Fall $n=1$ erfüllt $A = (d_1) = (\lambda_1)$ die Behauptung. Sei $n \ge 2$ und $A = (a_{ij}) \in K^{n\times n} \setminus K1_n$ eine Dreiecksmatrix mit Hauptdiagonale $\lambda_1,\ldots,\lambda_n$. Dann hat A Eigenwerte $\lambda_1,\ldots,\lambda_n$. Nach dem Satz von Fillmore ist A zu einer Matrix mit Hauptdiagonale $d_1,\ldots,d_n$ ähnlich. Diese hat die gleichen Eigenwerte wie A. $\square$

Beispiel 10.37.

(a) Wir suchen eine Matrix mit Eigenwerten $1, 1, 1$ und Hauptdiagonale $0, 0, 3$. Dafür wenden wir den Satz von Fillmore auf

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

an. Der Übergang zur Basis $\{(1, 0, 0), (1, 1, 0), (1, 2, 1)\}$ liefert

$$A \approx \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{pmatrix}.$$

(b) Die FIBONACCI-*Zahlen* F_k sind rekursiv definiert:

$$F_k := k \quad (k = 0, 1) \qquad F_{k+1} := F_k + F_{k-1} \quad (k \geq 1).$$

k	0	1	2	3	4	5	6	7	8	9	10
F_k	0	1	1	2	3	5	8	13	21	34	55

Wir suchen eine explizite Formel für F_k. Für $A := \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$ gilt

$$\begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix} = A \begin{pmatrix} F_k \\ F_{k-1} \end{pmatrix} = A^2 \begin{pmatrix} F_{k-1} \\ F_{k-2} \end{pmatrix} = \ldots = A^k \begin{pmatrix} F_1 \\ F_0 \end{pmatrix} = A^k \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Um A^k zu berechnen, diagonalisieren wir A. Wegen $\chi_A = (X - 1)X - 1 = X^2 - X - 1$ hat A die Eigenwerte $\varphi := \frac{1+\sqrt{5}}{2}$ und $\psi := \frac{1-\sqrt{5}}{2}$ (man nennt $\varphi \approx 1.618$ den *goldenen Schnitt*). Man berechnet

$$E_\varphi(A) = \operatorname{Ker}(A - \varphi 1_2) = \langle \begin{pmatrix} \varphi \\ 1 \end{pmatrix} \rangle,$$
$$E_\psi(A) = \langle \begin{pmatrix} \psi \\ 1 \end{pmatrix} \rangle.$$

Für $S := \begin{pmatrix} \varphi & \psi \\ 1 & 1 \end{pmatrix}$ gilt also $S^{-1}AS = \operatorname{diag}(\varphi, \psi)$ und

$$A^k = (S \operatorname{diag}(\varphi, \psi) S^{-1})^k = S \operatorname{diag}(\varphi, \psi)^k S^{-1} = S \operatorname{diag}(\varphi^k, \psi^k) S^{-1}.$$

Nach Beispiel 9.23 ist

$$S^{-1} = \frac{1}{\det(S)} \widetilde{S} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -\psi \\ -1 & \varphi \end{pmatrix}.$$

Insgesamt erhält man

$$A^k = S \operatorname{diag}(\varphi^k, \psi^k) S^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} \varphi^{k+1} & \psi^{k+1} \\ \varphi^k & \psi^k \end{pmatrix} \begin{pmatrix} 1 & -\psi \\ -1 & \varphi \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} * & * \\ \varphi^k - \psi^k & * \end{pmatrix}$$

und

$$\boxed{F_k = \frac{1}{\sqrt{5}}(\varphi^k - \psi^k)} \qquad \text{(BINET-}\textit{Formel}^6\text{)}$$

Wegen $|\psi^k| \approx 0.618^k \to 0$ gilt $F_k \approx \frac{1}{\sqrt{5}}\varphi^k$, d. h. F_k wächst exponentiell.

[6] Man kann die Formel auch durch Induktion beweisen, sofern man sie zuvor erraten hat.

Lemma 10.38. *Für $A \in K^{m\times n}$ und $B \in K^{n\times m}$ gilt $X^n\chi_{AB} = X^m\chi_{BA}$. Insbesondere haben AB und BA die gleichen von Null verschiedenen Eigenwerte.*

Beweis. Nach den Regeln für Blockmatrizen (Bemerkung 5.9, Lemma 9.9) und dem Determinantensatz gilt

$$\begin{aligned} X^n\chi_{AB} &= \det\left(\begin{pmatrix} 1_m & 0 \\ -B & X1_n \end{pmatrix}\begin{pmatrix} X1_m - AB & A \\ 0 & 1_n \end{pmatrix}\begin{pmatrix} 1_m & 0 \\ B & 1_n \end{pmatrix}\right) \\ &= \det\left(\begin{pmatrix} 1_m & 0 \\ -B & X1_n \end{pmatrix}\begin{pmatrix} X1_m & A \\ B & 1_n \end{pmatrix}\right) = \det\begin{pmatrix} X1_m & A \\ 0 & X1_n - BA \end{pmatrix} = X^m\chi_{BA}. \end{aligned}$$

Jeder Eigenwert $\lambda \in K$ von AB ist eine Nullstelle von $X^n\chi_{AB} = X^m\chi_{BA}$. Im Fall $\lambda \neq 0$ muss λ eine Nullstelle von χ_{BA} sein. Dann ist λ auch ein Eigenwert von BA (und umgekehrt). □

Bemerkung 10.39. Im Fall $n = m$ gilt sogar $\chi_{AB} = \chi_{BA}$ in der Situation von Lemma 10.38.

10.4 Minimalpolynome

Bemerkung 10.40. Wir haben bereits Polynome in Matrizen eingesetzt. Wir setzen nun umgekehrt Matrizen in Polynome ein. Für $\alpha = \sum_{k=0}^{d} a_kX^k \in K[X]$ und $A \in K^{n\times n}$ definieren wir

$$\alpha(A) := \sum_{k=0}^{d} a_kA^k \in K^{n\times n}.$$

Die Regeln aus Lemma 10.15 gelten auch in dieser Allgemeinheit.

Satz 10.41. *Für $A \in K^{n\times n}$ existiert genau ein normiertes Polynom $\mu_A \in K[X] \setminus \{0\}$ mit $\mu_A(A) = 0_n$ und $\deg(\mu_A)$ minimal.*

Beweis. Wegen $\dim K^{n\times n} = n^2$ (Lemma 5.4) sind die Potenzen $1_n = A^0, A, A^2, \ldots, A^{n^2}$ linear abhängig in $K^{n\times n}$. Also existieren $a_0, \ldots, a_{n^2} \in K$ (nicht alle 0) mit $\sum_{k=0}^{n^2} a_kA^k = 0$. Für $\alpha = \sum a_kX^k \in K[X]$ gilt somit $\alpha(A) = 0$. Indem man durch den führenden Koeffizienten von α teilt, kann man annehmen, dass α normiert ist. Dies zeigt, dass μ_A existiert. Sei auch $\tilde{\mu} \in K[X]$ normiert mit $\tilde{\mu}(A) = 0$ und $\deg(\tilde{\mu}) = \deg(\mu_A)$ minimal. Dann ist $(\mu_A - \tilde{\mu})(A) = \mu_A(A) - \tilde{\mu}(A) = 0$ und $\deg(\mu_A - \tilde{\mu}) < \deg(\mu_A)$. Die Minimalität von $\deg(\mu_A)$ zeigt $\mu_A - \tilde{\mu} = 0$, d. h. μ_A ist eindeutig bestimmt. □

Definition 10.42. Man nennt μ_A das *Minimalpolynom* von A.

Beispiel 10.43. Sei $A := \left(\begin{smallmatrix} 1 & -1 \\ 2 & 0 \end{smallmatrix}\right) \in \mathbb{Q}^{2\times 2}$. Da A keine Skalarmatrix ist, gilt $\deg \mu_A \geq 2$. Wir machen den Ansatz

$$A^2 + xA + y1_2 = \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix} + x\begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} + y\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

mit $x, y \in \mathbb{Q}$. Ein Vergleich der Matrixeinträge an Position $(1, 2)$ zeigt $x = -1$. Tatsächlich gilt die Gleichung nun für $y = 2$. Daher ist $\mu_A = X^2 - X + 2$.

Lemma 10.44. *Sei $A \in K^{n\times n}$ und $\alpha \in K[X]$ mit $\alpha(A) = 0$. Dann gilt $\mu_A \mid \alpha$.*

Beweis. Wir dividieren mit Rest: $\alpha = \mu_A\gamma + \delta$ mit $\gamma, \delta \in K[X]$ und $\deg(\delta) < \deg(\mu_A)$. Dann ist

$$\delta(A) = (\alpha - \mu_A\gamma)(A) = \alpha(A) - \mu_A(A)\gamma(A) = 0.$$

Aus der Minimalität von $\deg(\mu_A)$ folgt $\delta = 0$ und $\mu_A \mid \alpha$. $\square$

Lemma 10.45. *Ähnliche Matrizen haben das gleiche Minimalpolynom.*

Beweis. Sei $A \in K^{n\times n}$ mit $\mu_A = \sum a_k X^k$. Für $S \in \mathrm{GL}(n, K)$ gilt

$$\mu_A(SAS^{-1}) = \sum a_k(SAS^{-1})^k = \sum a_k SA^k S^{-1} = S\Big(\sum a_k A^k\Big)S^{-1} = S\mu_A(A)S^{-1} = 0_n.$$

Aus Lemma 10.44 folgt $\mu_{SAS^{-1}} \mid \mu_A$. Da Ähnlichkeit eine symmetrische Relation ist, gilt auch $\mu_A \mid \mu_{SAS^{-1}}$. Da beide Minimalpolynome normiert sind, müssen sie gleich sein. $\square$

Definition 10.46. Sei V ein K-Vektorraum mit Basis B. Für $f \in \mathrm{End}(V)$ sei wie üblich $\mu_f := \mu_{B[f]_B}$. Nach Lemma 10.45 hängt μ_f nicht von der Wahl von B ab. Die folgenden Sätze über Matrizen gelten sinngemäß auch für Endomorphismen.

Bemerkung 10.47. Aus dem Beweis von Satz 10.41 erhält man $\deg(\mu_A) \le n^2$. Der nächste Satz impliziert $\deg(\mu_A) \le n$.

Satz 10.48 (Cayley-Hamilton)**.** *Für $A \in K^{n\times n}$ gilt $\chi_A(A) = 0$ und $\mu_A \mid \chi_A$.*

Beweis. Sei $B := X1_n - A \in K[X]^{n\times n}$ und $\widetilde{B} \in K[X]^{n\times n}$ die zu B komplementäre Matrix. Aus jedem Eintrag von $\widetilde{B}$ extrahieren wir den Koeffizienten von X^k und bilden daraus die Matrix $B_k \in K^{n\times n}$. Es gilt nun

$$\widetilde{B} = \sum_{k=0}^{\infty} B_k X^k,$$

wobei nur endlich viele der B_k ungleich 0 sind. Sei $\chi_A = \sum a_k X^k$. Nach Satz 9.22 gilt

$$\sum_{k=0}^{\infty} a_k 1_n X^k = \chi_A 1_n = \det(B)1_n = \widetilde{B}B = \sum_{k=0}^{\infty} B_k X^k (X1_n - A) = \sum_{k=0}^{\infty}(B_{k-1} - B_k A)X^k,$$

wobei $B_{-1} := 0_n$. Ein Koeffizientenvergleich ergibt $a_k 1_n = B_{k-1} - B_k A$ für $k = 0, 1, \ldots$. Daher ist

$$\chi_A(A) = \sum_{k=0}^{\infty} a_k A^k = \sum_{k=0}^{\infty}(B_{k-1}A^k - B_k A^{k+1}) = \sum_{k=0}^{\infty} B_{k-1}A^k - \sum_{k=0}^{\infty} B_k A^{k+1} = 0.$$

Die zweite Behauptung folgt aus Lemma 10.44. $\square$

Beispiel 10.49.

(a) Für $A \in K^{2\times 2}$ gilt $A^2 - \mathrm{tr}(A)A + \det(A)1_2 = 0$ nach Lemma 10.31.

(b) Sei $A \in \mathrm{GL}(n,K)$ mit $\chi_A = \mu_A\gamma$ für ein $\gamma \in K[X]$. Nach Lemma 10.31 gilt

$$\mu_A(0)\gamma(0) = \chi_A(0) = \det(A) \neq 0.$$

Also hat μ_A die Form $\mu_A = X^d + a_{d-1}X^{d-1} + \ldots + a_0$ mit $a_0 \neq 0$. Man kann nun die Gleichung $A^d + a_{d-1}A^{d-1} + \ldots + a_0 1_n = 0$ auf beiden Seiten mit A^{-1} multiplizieren und erhält $A^{d-1} + a_{d-1}A^{d-2} + \ldots + a_1 1_n + a_0 A^{-1} = 0$. Dies liefert eine Formel für die Inverse

$$A^{-1} = -\frac{1}{a_0}\big(A^{d-1} + a_{d-1}A^{d-2} + \ldots + a_1 1_n\big).$$

Speziell für $n = 2$:

$$A^{-1} \overset{\text{(a)}}{=} \frac{1}{\det(A)}\big(\mathrm{tr}(A)1_2 - A\big) = \frac{1}{\det(A)}\widetilde{A}$$

(vgl. Beispiel 9.23).

Satz 10.50. *Die Eigenwerte von $A \in K^{n\times n}$ sind die Nullstellen von μ_A, d. h. χ_A und μ_A haben die gleichen Nullstellen (nicht unbedingt mit den gleichen Vielfachheiten).*

Beweis. Nach Cayley-Hamilton ist jede Nullstelle von μ_A auch eine Nullstelle von χ_A und damit ein Eigenwert von A (Satz 10.32). Sei umgekehrt $\lambda \in K$ ein Eigenwert von A mit Eigenvektor $v \in K^{n\times 1}$. Für $k \in \mathbb{N}_0$ ist $A^k v = A^{k-1}\lambda v = \ldots = \lambda^k v$. Sei $\mu_A = \sum a_k X^k$. Dann gilt

$$0 = \mu_A(A)v = \sum a_k A^k v = \sum a_k \lambda^k v = \mu_A(\lambda)v.$$

Wegen $v \neq 0$ ist $\mu_A(\lambda) = 0$, d. h. λ ist eine Nullstelle von μ_A. □

Bemerkung 10.51. Wegen $\deg \mu_A \leq \deg \chi_A$ vereinfacht Satz 10.50 die Bestimmung der Eigenwerte. Andererseits ist nicht klar, wie man μ_A effizient berechnet. Der nächste Satz verbessert Folgerung 8.12.

Satz 10.52. *Genau dann ist $A \in K^{n\times n}$ diagonalisierbar, wenn μ_A in paarweise verschiedene Linearfaktoren zerfällt.*

Beweis. Sei $A \in K^{n\times n}$ diagonalisierbar. Dann existiert $S \in \mathrm{GL}(n,K)$ mit $S^{-1}AS = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$. Die λ_i lassen sich sortieren, indem man die Spalten von S (d. h. die Eigenvektoren von A) entsprechend anordnet. Nach Lemma 10.45 können wir

$$A = \begin{pmatrix} \lambda_1 1_{n_1} & & 0 \\ & \ddots & \\ 0 & & \lambda_k 1_{n_k} \end{pmatrix}$$

annehmen, wobei $n = n_1 + \ldots + n_k$ und $\lambda_i \neq \lambda_j$ für $i \neq j$. Dann gilt

$$(A - \lambda_1 1_n)\ldots(A - \lambda_k 1_n) = \mathrm{diag}(0_{n_1},*,\ldots,*)\,\mathrm{diag}(*,\ldots,*,0_{n_2},*,\ldots,*)\ldots\mathrm{diag}(*,\ldots,*,0_{n_k}) = 0_n.$$

Nach Lemma 10.44 ist μ_A ein Teiler von $(X-\lambda_1)\ldots(X-\lambda_k)$. Andererseits sind $\lambda_1,\ldots,\lambda_k$ Eigenwerte und damit Nullstellen von μ_A. Dies zeigt $\mu_A = (X-\lambda_1)\ldots(X-\lambda_k)$.

Nehmen wir umgekehrt $\mu_A = (X-\lambda_1)\ldots(X-\lambda_k)$ mit paarweise verschiedenen $\lambda_1,\ldots,\lambda_k$ an. Sei P_k der k-dimensionale Vektorraum aller Polynome vom Grad kleiner k (Bemerkung 10.6). Für $i = 1,\ldots,k$ sei

$$\gamma_i := (X-\lambda_1)\ldots(X-\lambda_{i-1})(X-\lambda_{i+1})\ldots(X-\lambda_k) \in P_k.$$

Seien $a_1,\dots,a_k \in K$ mit $a_1\gamma_1 + \ldots + a_k\gamma_k = 0$. Dann gilt

$$a_i(\lambda_i - \lambda_1)\ldots(\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i+1})\ldots(\lambda_i - \lambda_n) = a_i\gamma_i(\lambda_i) = (a_1\gamma_1 + \ldots + a_k\gamma_k)(\lambda_i) = 0$$

und es folgt $a_i = 0$ für $i = 1,\dots,k$. Also sind $\gamma_1,\dots,\gamma_k$ linear unabhängig in P_k. Wegen $\dim P_k = k$ bilden sie sogar eine Basis. Insbesondere existieren $b_1,\dots,b_k \in K$ mit $\gamma := b_1\gamma_1 + \ldots + b_k\gamma_k = X^0 = 1$. Für $v \in K^{n\times 1}$ gilt

$$(A - \lambda_i 1_n)\gamma_i(A)v = \mu_A(A)v = 0,$$

d. h. $\gamma_i(A)v$ liegt in $E_{\lambda_i}(A)$. Andererseits ist

$$v = 1_n v = A^0 v = \gamma(A)(v) = b_1\gamma_1(A)(v) + \ldots + b_k\gamma_k(A)(v).$$

Dies zeigt $K^{n\times 1} = E_{\lambda_1}(A) + \ldots + E_{\lambda_k}(A)$. Nach Bemerkung 8.6 ist A diagonalisierbar. □

Beispiel 10.53. Sei $A \in K^{n\times n}$ mit genau zwei verschiedenen Eigenwerten. Angenommen wir finden einen Vektor $v \in K^{n\times 1}$, sodass v, Av, A^2v linear unabhängig sind. Dann sind auch $1_n, A, A^2$ linear unabhängig. Dies zeigt $\deg \mu_A \geq 3$. Nach Satz 10.52 ist A nicht diagonalisierbar.

Satz 10.54. *Für $A \in K^{n\times n}$ gilt $\chi_A \mid \mu_A^n$.*

Beweis. Sei $\mu_A = \sum a_i X^i$. Für $i \geq 1$ ist

$$(X^i 1_n - A^i) = (X1_n - A)(X^{i-1}1_n + X^{i-2}A + \ldots + XA^{i-2} + A^{i-1}).$$

Es folgt

$$\mu_A 1_n = \mu_A(X1_n) - \mu_A(A) = \sum_{i\geq 1} a_i(X^i 1_n - A^i) = (X1_n - A)B$$

für ein $B \in K[X]^{n\times n}$. Bildet man auf beiden Seiten die Determinante, so ergibt sich $\mu_A^n = \chi_A \det(B)$. □

Bemerkung 10.55. Die algebraische Vielfachheit eines Eigenwert von $A \in K^{n\times n}$ beträgt höchstens n. Zerfällt $\mu_A = (X - \lambda_1)\ldots(X - \lambda_k)$ in Linearfaktoren, so gilt daher $\chi_A \mid (X - \lambda_1)^n \ldots (X - \lambda_k)^n = \mu_A^n$ (der Beweis von Satz 10.54 kommt ohne diese Annahme aus).

11 Euklidische Geometrie

11.1 Skalarprodukte

Bemerkung 11.1. Wir betrachten in diesem Kapitel $K = \mathbb{R}$. Im Gegensatz zu beliebigen Körpern kann man $\mathbb{R}$ in positive und negative Zahlen unterteilen. Für $x \in \mathbb{R}$ sei wie üblich

$$|x| := \begin{cases} x & \text{falls } x \geq 0, \\ -x & \text{falls } x < 0 \end{cases}$$

der *Betrag* von x. Wir benutzen außerdem, dass jede positive reelle Zahl genau eine positive Quadratwurzel besitzt. Es gilt also $|x| = \sqrt{x^2}$ für alle $x \in \mathbb{R}$.

Definition 11.2. Sei V ein $\mathbb{R}$-Vektorraum. Eine Abbildung $V \times V \to \mathbb{R}$, $(v, w) \mapsto [v, w]$ heißt *Skalarprodukt*[1], falls folgende Bedingungen für alle $u, v, w \in V$ und $\lambda \in \mathbb{R}$ gelten:

- $[v, v] \geq 0$ mit Gleichheit genau dann, wenn $v = 0$ (*positiv definit*),
- $[v, w] = [w, v]$ (*symmetrisch*),
- $[\lambda u + v, w] = \lambda[u, w] + [v, w]$ (*bilinear*).

Zusammen mit einem Skalarprodukt wird V ein *euklidischer Raum*. Vektoren $v, w \in V$ heißen *orthogonal*, falls $[v, w] = 0$. Man nennt $|v| := \sqrt{[v, v]} \geq 0$ die *Norm* von v. Im Fall $|v| = 1$ nennt man v *normiert*.

Bemerkung 11.3.

(a) Die Symmetrie des Skalarprodukts zeigt

$$[u, \lambda v + w] = [\lambda v + w, u] = \lambda[v, u] + [w, u] = \lambda[u, v] + [u, w]$$

für alle $u, v, w \in V$ und $\lambda \in \mathbb{R}$. Für ein festes $x \in V$ sind also die Abbildungen $V \to \mathbb{R}$, $v \mapsto [v, x]$ und $V \to \mathbb{R}$, $v \mapsto [x, v]$ linear (dies erklärt den Begriff *bilinear*). Insbesondere ist $[v, 0] = 0 = [0, v]$ für alle $v \in V$. Dennoch ist die Abbildung $V \times V \to \mathbb{R}$, $(v, w) \mapsto [v, w]$ *nicht* linear, also kein Funktional, denn $[v, v] > 0 = [0, v] + [v, 0]$ für $v \neq 0$.

(b) Jeder Unterraum eines euklidischen Raums ist selbst ein euklidischer Raum mit dem eingeschränkten Skalarprodukt.

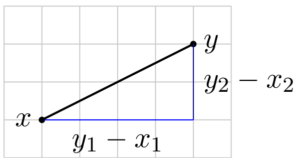
[1] Die Schreibweise $[v, w]$ ist in der Literatur nicht einheitlich. Man findet auch $\langle v, w\rangle$ (Verwechslung mit Spann), $(v \mid w)$ u. ä. Ebenso findet man $\|v\|$ anstatt $|v|$.

Beispiel 11.4.

(a) Das wichtigste Beispiel eines euklidischen Raums ist $V = \mathbb{R}^n$ mit dem *Standardskalarprodukt*

$$\boxed{[x,y] := xy^{\mathrm{t}} = \sum_{i=1}^{n} x_i y_i} \qquad (x,y \in \mathbb{R}^n).$$

Man überprüft leicht die drei Eigenschaften (positiv definit, symmetrisch und bilinear). Im Fall $n = 1$ ist $|x| = \sqrt{x_1^2}$ der gewöhnliche Betrag (dies rechtfertigt die Verwendung der Betragsstriche). Nach dem Satz des Pythagoras[2] entspricht die Norm $|y - x|$ im $\mathbb{R}^2$ dem geometrischen Abstand zwischen x und y:

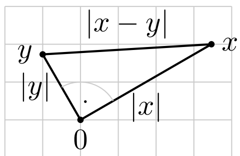


$$|y - x| = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}$$

Sind x und y orthogonal, so erhält man

$$|x - y|^2 = [x - y, x - y] = [x,x] - 2[x,y] + [y,y] = |x|^2 + |y|^2.$$

Nach der Umkehrung vom Satz des Pythagoras bilden x und y einen rechten Winkel, d. h. sie stehen senkrecht aufeinander (man schreibt $x \perp y$):



Im Allgemeinen gilt die *Parallelogrammgleichung*:

$$|x + y|^2 + |x - y|^2 = 2|x|^2 + 2|y|^2.$$

(b) Vektoren in $\mathbb{R}^n$ sind „diskrete“ Funktionen $\{1, \ldots, n\} \to \mathbb{R}$. In der Analysis betrachtet man eine „kontinuierliche“ Variante: Die stetigen Abbildungen $[0,1] \to \mathbb{R}$ auf dem abgeschlossenen Intervall $[0,1]$ bilden einen (unendlich-dimensionalen) Unterraum $V \leq \mathrm{Abb}([0,1], \mathbb{R})$ mit Skalarprodukt

$$[f,g] := \int_0^1 f(x)g(x)\,\mathrm{d}x \qquad (f,g \in V).$$

Lemma 11.5. *Sei V ein euklidischer Raum, $v, w \in V$ und $\lambda \in \mathbb{R}$. Dann gilt:*

(a) $|\lambda v| = |\lambda||v|$ (Homogenität).

(b) $|[v,w]| \leq |v||w|$ *mit Gleichheit genau dann, wenn v und w linear abhängig sind* (Cauchy-Schwarz-Ungleichung).

(c) $||v| - |w|| \leq |v + w| \leq |v| + |w|$ (Dreiecksungleichung).

Beweis.

(a) $|\lambda v| = \sqrt{[\lambda v, \lambda v]} = \sqrt{\lambda^2 [v,v]} = \sqrt{\lambda^2}\sqrt{[v,v]} = |\lambda||v|$.

[2] Man könnte auch den Abstand zwischen x und y durch $|y - x|$ *definieren* und damit den Satz des Pythagoras *beweisen.*

(b) O. B. d. A. sei $w \neq 0$. Sei $\lambda := \frac{[v,w]}{[w,w]}$. Nach den Eigenschaften des Skalarprodukts gilt

$$0 \leq |v - \lambda w|^2 = [v - \lambda w, v - \lambda w] = [v,v] - 2\lambda[v,w] + \lambda^2[w,w] = |v|^2 - \frac{[v,w]^2}{|w|^2}.$$

Es folgt $[v,w]^2 \leq |v|^2|w|^2$ und $|[v,w]| \leq |v||w|$. Gleichheit impliziert $v = \lambda w$, d. h. v und w sind linear abhängig. Sind umgekehrt v und w linear abhängig gegeben, dann existiert ein $\mu \in \mathbb{R}$ mit $v = \mu w$ und $|[v,w]| = |\mu||w|^2 \overset{(a)}{=} |\mu w||w| = |v||w|$.

(c) Zunächst ist

$$|v+w|^2 = [v+w, v+w] = [v,v] + 2[v,w] + [w,w] \overset{(b)}{\leq} |v|^2 + 2|v||w| + |w|^2 = (|v| + |w|)^2$$

und $|v+w| \leq |v| + |w|$. Daraus folgt $|v| = |v + w - w| \leq |v+w| + |w|$ und $|v| - |w| \leq |v+w|$. Vertauschen von v und w liefert $-(|v| - |w|) = |w| - |v| \leq |v+w|$, also $||v| - |w|| \leq |v+w|$. □

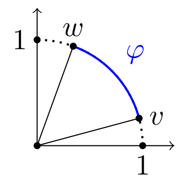
Bemerkung 11.6.

(a) Sind $v, w \in \mathbb{R}^n$ linear unabhängig, so bilden 0, v und $v + w$ ein Dreieck mit Seiten $|v|$, $|w|$ und $|v+w|$. Die Dreiecksungleichung besagt, dass die Summe von je zwei Seiten größer ist als die dritte Seite.

(b) Die Cauchy-Schwarz-Ungleichung impliziert $-1 \leq \frac{[v,w]}{|v||w|} \leq 1$ für $v, w \in V \setminus \{0\}$. Dieser Bruch verändert sich durch positive Skalierung von v und w nicht:

$$\frac{[\lambda v, \mu w]}{|\lambda v||\mu w|} = \frac{\lambda\mu[v,w]}{|\lambda||\mu||v||w|} = \frac{[v,w]}{|v||w|} \qquad (\lambda, \mu > 0)$$

Seien also v und w normiert. Dann definiert man den *Winkel* φ (im Bogenmaß) zwischen v und w als die Länge des Bogens auf dem Einheitskreis[3] zwischen v und w:



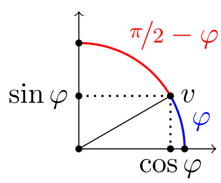
Die Länge des Halbkreisbogens nennt man π und berechnet $\pi \approx 3.14$ (Aufgabe 41). Der *Kosinus* von φ wird durch $\cos\varphi := [v,w]$ definiert.[4] Es gilt

$$\cos 0 = [e_1, e_1] = 1, \qquad \cos(\pi/2) = [e_1, e_2] = 0, \qquad \cos\pi = [e_1, -e_1] = -1.$$

Durch $\cos(\varphi + 2k\pi) = \cos\varphi$ für $k \in \mathbb{Z}$ setzt sich $\cos$ periodisch auf ganz $\mathbb{R}$ fort. Dabei gilt $\cos(-\varphi) = \cos\varphi$ und $\cos(\pi - \varphi) = -\cos\varphi$. Die „verschobene" Funktion

$$\sin\varphi := \cos(\varphi - \pi/2) = \cos(\pi/2 - \varphi)$$

für $\varphi \in \mathbb{R}$ nennt man den *Sinus* von φ. Für einen beliebigen normierten Vektor $v = (x, y)$ gilt $x = [v, e_1] = \cos\varphi$ und $y = [v, e_2] = \cos(\pi/2 - \varphi) = \sin\varphi$:



[3] in der Ebene $\langle v, w\rangle$

[4] Man kann zeigen, dass diese Definition mit der analytischen Definition als Potenzreihe übereinstimmt.

11.2 Orthonormalbasen

Definition 11.7. Sei V ein n-dimensionaler euklidischer Raum. Vektoren $b_1, \ldots, b_n$ bilden eine *Orthonormalbasis* von V, falls sie normiert und paarweise orthogonal sind, d. h. $[b_i, b_j] = \delta_{ij}$ für $1 \le i, j \le n$.

Bemerkung 11.8. Eine Orthonormalbasis $b_1, \ldots, b_n$ von V ist tatsächlich eine Basis. Dafür genügt es die lineare Unabhängigkeit zu prüfen. Seien $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ mit $\lambda_1 b_1 + \ldots + \lambda_n b_n = 0$. Dann gilt

$$\lambda_i = \sum_{j=1}^{n} \lambda_j [b_j, b_i] = \Big[\sum_{j=1}^{n} \lambda_j b_j, b_i\Big] = [0, b_i] = 0$$

für $i = 1, \ldots, n$.

Beispiel 11.9. Die Standardbasis $e_1, \ldots, e_n \in \mathbb{R}^n$ ist eine Orthonormalbasis bzgl. des Standardskalarprodukts. Jede Permutation einer Orthonormalbasis ist wieder eine Orthonormalbasis.

Satz 11.10 (Gram-Schmidt-Verfahren)**.** *Seien $v_1, \ldots, v_k \in V$ linear unabhängig. Wir definieren rekursiv:*

$$\boxed{b_s := v_s - \sum_{i=1}^{s-1} \frac{[v_s, b_i]}{[b_i, b_i]} b_i} \qquad (s = 1, \ldots, k).$$

Dann sind $b_1, \ldots, b_k$ paarweise orthogonal mit $\langle v_1, \ldots, v_k\rangle = \langle b_1, \ldots, b_k\rangle$. Folglich ist $\frac{1}{|b_1|} b_1, \ldots, \frac{1}{|b_k|} b_k$ eine Orthonormalbasis von $\langle v_1, \ldots, v_k\rangle$.

Beweis. Induktion nach k: Für $k = 1$ ist $b_1 = v_1 \ne 0$. Sei nun $k \ge 2$ und die Behauptung für $k - 1$ bereits bewiesen, d. h. $\langle v_1, \ldots, v_{k-1}\rangle = \langle b_1, \ldots, b_{k-1}\rangle$ und $[b_i, b_j] = 0$ für $1 \le i < j \le k - 1$. Wegen $\sum_{i=1}^{k-1} \frac{[v_k, b_i]}{[b_i, b_i]} b_i \in \langle b_1, \ldots, b_{k-1}\rangle$ gilt

$$\langle v_1, \ldots, v_k\rangle = \langle b_1, \ldots, b_{k-1}, v_k\rangle = \langle b_1, \ldots, b_k\rangle.$$

Für $i = 1, \ldots, k - 1$ ist außerdem

$$[b_k, b_i] = [v_k, b_i] - \sum_{j=1}^{k-1} \frac{[v_k, b_j]}{[b_j, b_j]} [b_j, b_i] = [v_k, b_i] - [v_k, b_i] = 0.$$

Damit ist die erste Behauptung bewiesen. Wegen $|\frac{1}{|b_i|} b_i| = \frac{1}{|b_i|} |b_i| = 1$ folgt die zweite Behauptung. □

Folgerung 11.11. *Jeder euklidische Raum besitzt* (*mindestens*) *eine Orthonormalbasis.*

Beweis. Man wende das Gram-Schmidt-Verfahren auf eine beliebige Basis an. □

Beispiel 11.12. Seien $v_1 := (1, 0, 1)$, $v_2 := (0, 1, 1)$ und $v_3 := (-1, 2, 0)$ linear unabhängig in $\mathbb{R}^3$. Bezüglich des Standardskalarprodukts erhält man

$$\begin{aligned}
b_1 &:= v_1 = (1, 0, 1), \\
b_2 &:= v_2 - \frac{[v_2, b_1]}{[b_1, b_1]} b_1 = (0, 1, 1) - \frac{1}{2}(1, 0, 1) = \frac{1}{2}(-1, 2, 1),
\end{aligned}$$

$$b_3 := v_3 - \frac{[v_3, b_1]}{[b_1, b_1]} b_1 - \frac{[v_3, b_2]}{[b_2, b_2]} b_2 = (-1, 2, 0) + \frac{1}{2}(1, 0, 1) - \frac{5}{6}(-1, 2, 1) = \frac{1}{3}(1, 1, -1).$$

Man beachte, dass Skalierungsfaktoren in dieser Rechnung keine Rolle spielen. Man kann b_3 also etwas bequemer mit $b_2 = (-1, 2, 1)$ ausrechnen. Nach Normierung ist $\frac{1}{\sqrt{2}}(1, 0, 1)$, $\frac{1}{\sqrt{6}}(-1, 2, 1)$, $\frac{1}{\sqrt{3}}(1, 1, -1)$ eine Orthonormalbasis von $\mathbb{R}^3$. Um die Einträge der Vektoren zu minimieren, kann es nützlich sein zuerst den Gauß-Algorithmus anzuwenden, bevor man das Gram-Schmidt-Verfahren startet. In diesem Beispiel würde man am Ende die Standardbasis von $\mathbb{R}^3$ erhalten.

Definition 11.13. Sei V ein euklidischer Raum und $S \subseteq V$. Dann nennt man

$$S^\perp := \{v \in V : \forall s \in S : [v, s] = 0\}$$

das *orthogonale Komplement* von S in V. Im Fall $S = \{v\}$ schreibt man $s^\perp := S^\perp$.

Bemerkung 11.14. Für $v, w \in S^\perp$ und $\lambda \in \mathbb{R}$ gilt $[\lambda v + w, s] = \lambda[v, s] + [w, s] = 0$ für alle $s \in S$, d. h. $\lambda v + w \in S^\perp$. Daher ist $S^\perp$ ein Unterraum, selbst wenn S nur eine Teilmenge ist. Außerdem gilt $S^\perp = \langle S \rangle^\perp$.

Lemma 11.15. *Für Unterräume U, W eines euklidischen Raums V gilt*

(a) $V = U \oplus U^\perp$ *und* $\dim V = \dim U + \dim U^\perp$.

(b) $(U^\perp)^\perp = U$.

(c) $U \subseteq W \iff W^\perp \subseteq U^\perp$.

Beweis.

(a) Für $v \in U \cap U^\perp$ gilt $|v|^2 = [v, v] = 0$ und $v = 0$. Also ist $U \cap U^\perp = \{0\}$ und $U + U^\perp = U \oplus U^\perp$. Wir können eine Basis $v_1, \ldots, v_k$ von U zu einer Basis $v_1, \ldots, v_n$ von V ergänzen. Das Gram-Schmidt-Verfahren liefert eine Orthonormalbasis $b_1, \ldots, b_n$ von V mit $\langle v_1, \ldots, v_k \rangle = \langle b_1, \ldots, b_k \rangle$. Daher ist $b_{k+1}, \ldots, b_n \in U^\perp$ und $V = U \oplus U^\perp$.

(b) Nach Definition ist $U \subseteq (U^\perp)^\perp$. Nach (a) ist $\dim(U^\perp)^\perp = \dim V - \dim U^\perp = \dim U$. Also ist $U = (U^\perp)^\perp$.

(c) Nach Definition gilt

$$U \subseteq W \implies W^\perp \subseteq U^\perp \overset{\text{(b)}}{\implies} U = (U^\perp)^\perp \subseteq (W^\perp)^\perp = W. \qquad \square$$

Beispiel 11.16.

(a) In $V = \mathbb{R}^n$ lässt sich ein orthogonales Komplement von $U \leq V$ bzgl. des Standardskalarprodukts mit dem Gauß-Algorithmus bestimmen: Man schreibt die Vektoren eines Erzeugendensystems von U als Zeilen in eine Matrix $A \in \mathbb{R}^{k \times n}$. Die Lösungsmenge L_0 des homogenen Gleichungssystems $Ax = 0$ ist $U^\perp$, denn nach Satz 6.6 gilt $\dim L_0 = n - \mathrm{rk}(A) = n - \dim U = \dim U^\perp$.

(b) Für $v = (x, y) \in \mathbb{R}^2 \setminus \{0\}$ gilt $v^\perp = \langle (y, -x) \rangle$.

(c) Seien $v, w \in \mathbb{R}^3$ linear unabhängig. Wir ergänzen zu einer Basis u, v, w von $\mathbb{R}^3$, sodass die Matrix

$$A := \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

Determinante 1 hat (das geht immer, indem man u geeignet skaliert). Es gibt genau einen Vektor $x \in \langle v, w\rangle^\perp$ mit $Ax^{\mathrm{t}} = e_1$ und zwar

$$x^{\mathrm{t}} = A^{-1}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \overset{9.22}{=} \widetilde{A}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \det(A_{11}) \\ -\det(A_{12}) \\ \det(A_{13}) \end{pmatrix} = \begin{pmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \\ v_1w_2 - v_2w_1 \end{pmatrix} =: v \times w.$$

Man nennt $v \times w$ das *Kreuzprodukt* von v und w. Nach Konstruktion gilt $\langle v, w\rangle^\perp = \langle v \times w\rangle$. Die Richtung von $v \times w$ lässt sich mit der *Rechte-Hand-Regel* bestimmen: Zeigt v in Richtung Daumen und w in Richtung Zeigefinger, so zeigt $v \times w$ in Richtung des Mittelfingers der rechten Hand.

11.3 Symmetrische und orthogonale Abbildungen

Definition 11.17. Sei V ein euklidischer Raum und $f \in \mathrm{End}(V)$. Man nennt f

- *symmetrisch*,[5] falls $[f(v), w] = [v, f(w)]$ für alle $v, w \in V$.
- *orthogonal*,[6] falls $[f(v), f(w)] = [v, w]$ für alle $v, w \in V$.

Bemerkung 11.18.

(a) Die Nullabbildung ist symmetrisch. Sind $f, g \in \mathrm{End}(V)$ symmetrisch und $\lambda \in \mathbb{R}$, dann ist offenbar auch $\lambda f + g$ symmetrisch. Die symmetrischen Abbildungen bilden also einen Unterraum von $\mathrm{End}(V)$.

(b) Orthogonale Abbildungen $f \in \mathrm{End}(V)$ sind Isomorphismen, denn aus $v \in \mathrm{Ker}(f)$ folgt $|v|^2 = [v, v] = [f(v), f(v)] = 0$, also $v = 0$. Wegen $|f(v - w)| = |v - w|$ und

$$\frac{[f(v), f(w)]}{|f(v)||f(w)|} = \frac{[v, w]}{|v||w|}$$

erhält f Abstände und Winkel (Bemerkung 11.6). Insbesondere bildet f Orthonormalbasen auf Orthonormalbasen ab. Mit f, g sind auch $f \circ g$ und f^{-1} orthogonal. Die Menge der orthogonalen Abbildungen bildet daher eine Untergruppe $\mathrm{O}(V)$ von $\mathrm{GL}(V)$. Man nennt $\mathrm{O}(V)$ die *orthogonale Gruppe* von V.

(c) Der folgende Satz zeigt, dass längen-erhaltende Abbildung automatisch orthogonal (insbesondere linear) sind.

Satz 11.19 (Mazur-Ulam). *Sei V ein euklidischer Raum und $f \in \mathrm{Abb}(V, V)$ mit $|f(v)| = |v|$ für alle $v \in V$. Dann ist $f \in \mathrm{O}(V)$.*

[5] oder *selbstadjungiert*, siehe Abschnitt 13.2
[6] oder *isometrisch*

Beweis. Für $v, w \in V$ gilt

$$[f(v), f(w)] = \frac{1}{2}\big(|f(v)|^2 + |f(w)|^2 - |f(v) - f(w)|^2\big) = \frac{1}{2}\big(|v|^2 + |w|^2 - |v - w|^2\big) = [v, w].$$

Daraus folgt

$$\begin{aligned} |f(\lambda v + w) - \lambda f(v) - f(w)|^2 &= [f(\lambda v + w) - \lambda f(v) - f(w),\ f(\lambda v + w) - \lambda f(v) - f(w)] \\ &= [\lambda v + w - \lambda v - w,\ \lambda v + w - \lambda v - w] = 0 \end{aligned}$$

für $\lambda \in \mathbb{R}$. Also ist f linear und orthogonal. □

Lemma 11.20. *Sei V euklidisch mit Orthonormalbasis B, $f \in \mathrm{End}(V)$ und $A := {}_B[f]_B$. Dann gilt:*

(a) f symmetrisch $\iff$ $A^{\mathrm{t}} = A$.

(b) f orthogonal $\iff$ $A^{\mathrm{t}} = A^{-1}$.

Beweis. Sei $B = \{b_1, \ldots, b_n\}$ und $A = (a_{ij})$. Dann ist $f(b_i) = \sum_{j=1}^n a_{ji} b_j$ für $i = 1, \ldots, n$.

(a) Ist f symmetrisch, so gilt $a_{ji} = [f(b_i), b_j] = [b_i, f(b_j)] = a_{ij}$ für $1 \le i, j \le n$. Also ist $A = A^{\mathrm{t}}$. Sei umgekehrt $A = A^{\mathrm{t}}$. Dann folgt $[f(b_i), b_j] = [b_i, f(b_j)]$ für $1 \le i, j \le n$. Für beliebige $v = \sum \lambda_i b_i$ und $w = \sum \mu_i b_i$ in V gilt

$$[f(v), w] = \sum_{i,j=1}^n \lambda_i \mu_j [f(b_i), b_j] = \sum_{i,j=1}^n \lambda_i \mu_j [b_i, f(b_j)] = [v, f(w)].$$

Also ist f symmetrisch.

(b) Ist f orthogonal, so gilt

$$\sum_{k=1}^n a_{ki} a_{kj} = \left[\sum_{k=1}^n a_{ki} b_k, \sum_{k=1}^n a_{kj} b_k\right] = [f(b_i), f(b_j)] = [b_i, b_j] = \delta_{ij}.$$

Dies zeigt $A^{\mathrm{t}} A = 1_n$ und $A^{\mathrm{t}} = A^{-1}$. Ist umgekehrt $A^{\mathrm{t}} A = 1_n$, so gilt $[f(b_i), f(b_j)] = \delta_{ij} = [b_i, b_j]$. Wie in (a) folgt $[f(v), f(w)] = [v, w]$ für alle $v, w \in V$, d. h. f ist orthogonal. □

Bemerkung 11.21.

(a) Für einen beliebigen Körper K nennt man Matrizen $A \in \mathrm{GL}(n, K)$ *orthogonal*, falls $A^{\mathrm{t}} = A^{-1}$. Man zeigt leicht, dass die orthogonalen Matrizen eine Untergruppe $\mathrm{O}(n, K)$ von $\mathrm{GL}(n, K)$ bilden. Wie üblich entspricht $\mathrm{O}(n, \mathbb{R})$ der Gruppe $\mathrm{O}(V)$ durch Basiswahl (so wie sich $\mathrm{GL}(V)$ und $\mathrm{GL}(n, K)$ entsprechen). Für jede orthogonale Matrix A gilt

$$\det(A)^2 \overset{9.12}{=} \det(A)\det(A^{\mathrm{t}}) = \det(AA^{\mathrm{t}}) = \det(1_n) = 1$$

und $\det(A) = \pm 1$. Man nennt

$$\mathrm{SO}(n, K) := \mathrm{O}(n, K) \cap \mathrm{SL}(n, K) \le \mathrm{O}(n, K) \le \mathrm{GL}(n, K).$$

die *spezielle orthogonale Gruppe* vom Grad n über K.

(b) Die Gleichung $A^t A = 1_n = AA^t$ bedeutet für orthogonale reelle Matrizen, dass die Spalten (bzw. Zeilen) von A eine Orthonormalbasis von $\mathbb{R}^n$ bzgl. des Standardskalarprodukts bilden. Sei v ein Eigenvektor von A zum Eigenwert $\lambda \in \mathbb{R}$. Dann gilt

$$|v|^2 = v^t v = v^t A^t A v = (Av)^t (Av) = \lambda^2 v^t v = \lambda^2 |v|^2$$

und $\lambda = \pm 1$.

Beispiel 11.22.

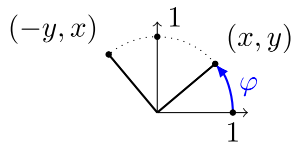
(a) Jede Permutationsmatrix ist orthogonal, denn die Zeilen bilden eine Orthonormalbasis (nämlich eine Permutation der Standardbasis).

(b) Für $A \in \mathrm{O}(2, \mathbb{R})$ gilt

$$A = \begin{pmatrix} x & \mp y \\ y & \pm x \end{pmatrix}$$

mit $x^2 + y^2 = 1 = \pm \det(A)$ (vgl. Beispiel 11.16).

- Im Fall $\det(A) = 1$ beschreibt A eine *Drehung* um den Winkel φ zwischen e_1 und (x, y):



  Es gilt dann

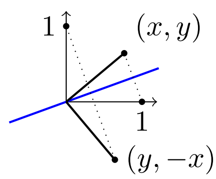
$$A = D(\varphi) := \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

  Für zwei Winkel φ und ψ erhält man $D(\varphi + \psi) = D(\varphi)D(\psi)$, woraus die bekannten *Additionstheoreme* folgen:

$$\cos(\varphi + \psi) = \cos(\varphi)\cos(\psi) - \sin(\varphi)\sin(\psi),$$
$$\sin(\varphi + \psi) = \sin(\varphi)\cos(\psi) + \sin(\psi)\cos(\varphi).$$

  Offenbar besitzt $D(\varphi)$ nur dann (reelle) Eigenwerte, wenn $\varphi \in \{0, \pi\}$, d. h. $D(0) = 1_2$ und $D(\pi) = -1_2$.

- Im Fall $\det(A) = -1$ beschreibt A eine *Spiegelung* an der Winkelhalbierenden zwischen e_1 und (x, y):



  Es gilt dann

$$A = S(\varphi) := \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}.$$

Die Spiegelachse wird von einem Eigenvektor zum Eigenwert 1 aufgespannt. Orthogonal dazu steht ein Eigenvektor zum Eigenwert -1 (beachte: $\det(A)$ ist das Produkt der Eigenwerte). Nach Folgerung 8.12 ist $S(\varphi)$ diagonalisierbar. Tatsächlich ist

$$D(\varphi/2)^{-1}S(\varphi)D(\varphi/2) = D(-\varphi/2)S(\varphi)D(\varphi/2) = S(0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Für spezielle Winkel erhält man (vgl. Aufgabe 40):

φ	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	π
$D(\varphi)$	$\frac{1}{2}\begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix}$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$	-1_2
$S(\varphi)$	$\frac{1}{2}\begin{pmatrix} \sqrt{3} & 1 \\ 1 & -\sqrt{3} \end{pmatrix}$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	$\frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Damit lässt sich auch

$$D(\pi/12) = D(\pi/3 - \pi/4) = D(\pi/3)D(\pi/4)^{-1} = D(\pi/3)D(\pi/4)^{\mathrm{t}}$$

berechnen.

Bemerkung 11.23. Um zu zeigen, dass symmetrische Endomorphismen diagonalisierbar sind, müssen wir die reellen Zahlen vorübergehend verlassen.

11.4 Komplexe Zahlen

Lemma 11.24. *Der $\mathbb{R}$-Vektorraum $\mathbb{C} := \mathbb{R}^2$ wird durch die Multiplikation*

$$(a,b)\cdot(c,d) := (ac - bd,\ ad + bc) \qquad (a,b,c,d \in \mathbb{R})$$

zu einem Körper.

Beweis. Als Vektorraum ist $(\mathbb{C}, +)$ bereits eine abelsche Gruppe. Die Multiplikation in $\mathbb{C}$ führen wir auf die Matrizenmultiplikation zurück.[7] Dafür betrachten wir die injektive Abbildung

$$\Gamma\colon \mathbb{C} \to \mathbb{R}^{2\times 2}, \quad z = (a,b) \mapsto \Gamma(z) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Für $x, y, z \in \mathbb{C}$ gilt $\Gamma(x) + \Gamma(y) = \Gamma(x+y)$ und $\Gamma(x)\Gamma(y) = \Gamma(x\cdot y)$ (nachrechnen). Aus Lemma 5.8 folgt

$$\Gamma(x(yz)) = \Gamma(x)\Gamma(yz) = \Gamma(x)(\Gamma(y)\Gamma(z)) = (\Gamma(x)\Gamma(y))\Gamma(z) = \Gamma(xy)\Gamma(z) = \Gamma((xy)z)$$

und $x(yz) = (xy)z$, da Γ injektiv ist. Analog beweist man das Kommutativ- und Distributivgesetz. Wegen $\Gamma(1,0) = 1_2$ ist $(1,0)$ das Einselement in $\mathbb{C}$. Es bleibt zu zeigen, dass $z \neq 0$ invertierbar ist. Es gilt $\det(\Gamma(z)) = a^2 + b^2 > 0$ und

$$\Gamma(z)^{-1} \overset{9.23}{=} \frac{1}{\det(\Gamma(z))}\widetilde{\Gamma(z)} = \frac{1}{a^2+b^2}\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \Gamma\Big(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\Big). \qquad \square$$

[7] Man kann die Axiome auch direkt mit der Definition nachrechnen.

Definition 11.25. Man nennt $\mathbb{C}$ den Körper der *komplexen Zahlen*.

- Mittels der Abbildung $\mathbb{R} \to \mathbb{C}$, $a \mapsto (a, 0)$ werden wir $\mathbb{R}$ als Teilmenge von $\mathbb{C}$ auffassen. Die Verknüpfungen in $\mathbb{R}$ entsprechen genau denen in $\mathbb{C}$ mit den gleichen neutralen Elementen.
- Man nennt $\mathrm{i} := (0, 1) \in \mathbb{C} \setminus \mathbb{R}$ die *imaginäre Einheit*. Es gilt $\mathrm{i}^2 = (-1, 0) = -1$. Da $1, \mathrm{i}$ eine Basis von $\mathbb{C}$ bilden, lässt sich jede komplexe Zahl eindeutig in der Form $z = a + b\mathrm{i}$ schreiben, wobei $\mathrm{Re}(z) := a \in \mathbb{R}$ der *Realteil* und $\mathrm{Im}(z) := b \in \mathbb{R}$ der *Imaginärteil* von z ist.
- Man nennt $|z| := \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2} \geq 0$ den *Betrag* von z (dies entspricht der Norm in $\mathbb{R}^2$).

Bemerkung 11.26. Im Gegensatz zu $\mathbb{R}$ ist $\mathbb{C}$ kein *angeordneter* Körper, d. h. es gibt keine Ordnungsrelation $\leq$ auf $\mathbb{C}$ mit

$$\begin{aligned} a \leq b &\implies a + c \leq b + c, \\ a, b \geq 0 &\implies ab \geq 0 \end{aligned}$$

für alle $a, b, c \in \mathbb{C}$. Denn angenommen es gilt $\mathrm{i} > 0$. Dann wäre $-1 = \mathrm{i}^2 > 0$ und $1 = (-1)^2 > 0$. Nun erhält man den Widerspruch $0 = 1 - 1 > 0 + 0 = 0$. Ist hingegen $\mathrm{i} < 0$, so wäre $0 = \mathrm{i} - \mathrm{i} < -\mathrm{i}$ und $-1 = (-\mathrm{i})^2 > 0$. Dies führt zum selben Widerspruch.

Lemma 11.27. *Für $z \in \mathbb{C} \setminus \{0\}$ und $n \in \mathbb{N}$ existieren paarweise verschiedene n-te Wurzeln $\zeta_1, \ldots, \zeta_n \in \mathbb{C}$ mit $\zeta_1^n = \ldots = \zeta_n^n = z$.*

Beweis. Sei $z_0 := \frac{1}{|z|} z \in \mathbb{C}$ und $\Gamma(z_0) = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) \in \mathbb{R}^{2 \times 2}$ wie im Beweis von Lemma 11.24. Wegen $a^2 + b^2 = |z_0|^2 = 1$ ist $\Gamma(z_0) = D(\varphi) \in \mathrm{O}(2, \mathbb{R})$ für einen Winkel φ mit $a = \cos\varphi$ und $b = \sin\varphi$. Wir definieren $\varphi_k := \frac{\varphi + 2k\pi}{n}$ für $k = 1, \ldots, n$. Dann gilt

$$D(\varphi_k)^n = D(n\varphi_k) = D(\varphi + 2k\pi) = D(\varphi) = \Gamma(z_0).$$

Die Zahlen $z_k := \cos\varphi_k + \mathrm{i}\sin\varphi_k \in \mathbb{C}$ erfüllen also $z_k^n = z_0$. Wegen $|z| > 0$ existiert $\sqrt[n]{|z|} \in \mathbb{R}_{>0}$ (Analysis). Für $\zeta_k := \sqrt[n]{|z|} z_k$ erhält man $\zeta_k^n = |z| z_0 = z$ für $k = 1, \ldots, n$. Sei nun $\zeta_k = \zeta_l$ für $1 \leq k \leq l \leq n$. Dann unterscheiden sich φ_k und φ_l nur um ein Vielfaches von 2π. Dies zeigt $2\pi(l - k) = 2\pi cn$ für ein $c \in \mathbb{N}_0$. Aus $0 \leq l - k < n$ folgt $k = l$. Daher sind die n-ten Wurzeln $\zeta_1, \ldots, \zeta_n$ paarweise verschieden. $\square$

Beispiel 11.28. Die (n-ten) Wurzeln aus 1 nennt man *Einheitswurzeln*. Sie entsprechen Drehungen um $2\pi k/n$ mit $k \in \mathbb{Z}$. Die vierten Einheitswurzeln sind 1, i, -1, $-\mathrm{i}$.

Folgerung 11.29. *Sei $A \in \mathbb{C}^{n \times n}$ und $k \in \mathbb{N}$ mit $A^k = A$. Dann ist A diagonalisierbar.*

Beweis. Das Minimalpolynom μ_A teilt $X^k - X = X(X^{k-1} - 1)$ nach Lemma 10.44. Die Nullstellen von $X^k - X$ sind die $(k-1)$-ten Einheitswurzeln und 0. Nach Lemma 10.24 zerfällt μ_A in paarweise verschiedene Linearfaktoren. Die Behauptung folgt nun aus Satz 10.52. $\square$

Beispiel 11.30. Sei $k \in \mathbb{N}$ und $A := D(2\pi/k)$. Wegen $A^{k+1} = AA^k = AD(2\pi) = A$ ist A über $\mathbb{C}$ diagonalisierbar, aber nicht unbedingt über $\mathbb{R}$.

Definition 11.31. Die Abbildung $\mathbb{C} \to \mathbb{C}$, $a + b\mathrm{i} \mapsto a - b\mathrm{i} =: \overline{a + b\mathrm{i}}$ heißt *komplexe Konjugation*.

Lemma 11.32. *Für $z, w \in \mathbb{C}$ gilt $|z|^2 = z\overline{z}$, $\overline{z+w} = \overline{z} + \overline{w}$ und $\overline{zw} = \overline{z} \cdot \overline{w}$.*

Beweis. Für $z = a + b\mathrm{i}$ und $w = c + d\mathrm{i}$ gilt

$$\begin{aligned} |z|^2 &= a^2 + b^2 = (a + b\mathrm{i})(a - b\mathrm{i}) = z\overline{z}, \\ \overline{z+w} &= a + c - (b + d)\mathrm{i} = a - b\mathrm{i} + c - d\mathrm{i} = \overline{z} + \overline{w}, \\ \overline{zw} &= ac - bd - (ad + bc)\mathrm{i} = (a - b\mathrm{i})(c - d\mathrm{i}) = \overline{z} \cdot \overline{w}. \end{aligned}$$ □

Bemerkung 11.33.

(a) Für Matrizen $A = (a_{ij}) \in \mathbb{C}^{n\times m}$ definieren wir $\overline{A} := (\overline{a_{ij}}) \in \mathbb{C}^{n\times m}$. Für $B = (b_{ij}) \in \mathbb{C}^{m\times k}$ gilt

$$\overline{AB} = \overline{\Big(\sum_{l=1}^{m} a_{il}b_{lj}\Big)_{ij}} \overset{11.32}{=} \Big(\sum_{l=1}^{m} \overline{a_{il}b_{lj}}\Big)_{ij} \overset{11.32}{=} \Big(\sum_{l=1}^{m} \overline{a_{il}}\,\overline{b_{lj}}\Big)_{ij} = \overline{A} \cdot \overline{B}.$$

(b) Nach Lemma 11.27 hat das Polynom $X^n - z$ für jedes $z \in \mathbb{C}$ eine Nullstelle (insbesondere ist i eine Nullstelle von $X^2 + 1$). Erstaunlicherweise besitzt sogar jedes nicht-konstante Polynom in $\mathbb{C}[X]$ eine Nullstelle.[8]

Satz 11.34 (Fundamentalsatz der Algebra)**.** *Jedes Polynom $\alpha \in \mathbb{C}[X] \setminus \mathbb{C}$ besitzt eine Nullstelle in $\mathbb{C}$.*

Beweis. Der Beweis benutzt Analysis (genauer die Vollständigkeit von $\mathbb{R}$) und ist für diese Vorlesung zu schwierig. Siehe Algebra-Skript. □

Folgerung 11.35. *Jedes normierte Polynom $\alpha \in \mathbb{C}[X] \setminus \mathbb{C}$ zerfällt in Linearfaktoren.*

Beweis. Induktion nach $d := \deg(\alpha) \geq 1$. Im Fall $d = 1$ ist α selbst ein Linearfaktor, da α normiert ist. Sei nun $d \geq 2$. Nach dem Fundamentalsatz besitzt α eine Nullstelle $x \in \mathbb{C}$. Nach Lemma 10.22 existiert ein $\beta \in \mathbb{C}[X]$ mit $\alpha = (X - x)\beta$ und $\deg(\beta) = d - 1$. Da α normiert ist, muss auch β normiert sein. Nach Induktion zerfällt β in Linearfaktoren und damit auch α. □

Beispiel 11.36. Sei $\alpha \in \mathbb{R}[X]$ mit ungeradem Grad. Um zu zeigen, dass α eine Nullstelle hat, können wir annehmen, dass α normiert ist. Dann gilt $\lim_{x\to\pm\infty} \alpha(x) = \pm\infty$. Nach dem Zwischenwertsatz der Analysis besitzt die stetige Funktion $\mathbb{R} \to \mathbb{R}$, $x \mapsto \alpha(x)$ eine reelle Nullstelle. In der Praxis lässt sich eine solche Nullstelle allerdings nur näherungsweise berechnen. Sei konkret $\alpha := X^5 - 4X + 2$. Anhand des Graphen von α vermuten wir eine Nullstelle in der Nähe von $x_0 := 0.5$. Sei $\alpha' = 5X^4 - 4$ die Ableitung von α (Analysis). Beim *Newton-Verfahren* berechnet man die rekursive Folge

$$x_{n+1} := x_n - \frac{\alpha(x_n)}{\alpha'(x_n)} \qquad (n \geq 0),$$

also $x_1 = 0.50847\ldots$, $x_2 = 0.50849948\ldots$ usw. Ist x_0 (so wie hier) „gut" gewählt, so konvergiert die Folge $(x_n)_n$ quadratisch gegen eine Nullstelle, d. h. mit jeder Iteration verdoppelt sich die Anzahl der korrekten Dezimalstellen. Tatsächlich sind bereits alle angegebenen Dezimalstellen von x_2 korrekt.

[8] Man sagt: $\mathbb{C}$ ist *algebraisch abgeschlossen*.

Bemerkung 11.37. Sei $x \in \mathbb{C}$ eine Nullstelle von $\alpha = \sum a_k X^k \in \mathbb{C}[X]$. Wir definieren $\overline{\alpha} := \sum \overline{a_k} X^k \in \mathbb{C}[X]$. Es gilt dann
$$\overline{\alpha}(\overline{x}) = \sum \overline{a_k x^k} = \overline{\alpha(x)} = \overline{0} = 0,$$
d. h. $\overline{x}$ ist eine Nullstelle von $\overline{\alpha}$. Im Fall $\alpha \in \mathbb{R}[X]$ gilt also: $\alpha(x) = 0 \iff \alpha(\overline{x}) = 0$. Ggf. ist
$$(X - x)(X - \overline{x}) = X^2 - (x + \overline{x})X + x\overline{x} = X^2 - 2\mathrm{Re}(x)X + |x|^2 \in \mathbb{R}[X]$$
ein Teiler von α.

11.5 Der Hauptachsensatz

Satz 11.38 (Hauptachsensatz). *Sei V ein euklidischer Raum und $f \in \mathrm{End}(V)$. Genau dann ist f symmetrisch, wenn V eine Orthonormalbasis aus Eigenvektoren von f besitzt. Insbesondere sind symmetrische Endomorphismen diagonalisierbar.*

Beweis. Sei B eine Orthonormalbasis aus Eigenvektoren von f. Dann ist ${}_B[f]_B$ eine Diagonalmatrix und daher symmetrisch. Nach Lemma 11.20 ist f symmetrisch. Sei nun umgekehrt f symmetrisch. Wir argumentieren durch Induktion nach $n := \dim V$. Im Fall $n = 1$ kann jeder normierte Vektor für eine Orthonormalbasis aus Eigenvektoren verwendet werden. Sei also $n \geq 2$ und die Behauptung für $n - 1$ bereits bewiesen. Sei zunächst B eine beliebige Orthonormalbasis von V. Dann können wir die symmetrische Matrix $A := {}_B[f]_B$ auch als komplexe Matrix auffassen. Nach dem Fundamentalsatz der Algebra besitzt $\chi_A \in \mathbb{C}[X] \setminus \mathbb{C}$ eine Nullstelle $\lambda \in \mathbb{C}$. Also ist λ ein Eigenwert von A. Sei $v = (v_1, \ldots, v_n)^{\mathrm{t}} \in \mathbb{C}^{n \times 1}$ ein entsprechender Eigenvektor. Wegen
$$\lambda \sum_{i=1}^{n} |v_i|^2 = \lambda \overline{v}^{\mathrm{t}} v = \overline{v}^{\mathrm{t}} A v = (\overline{v}^{\mathrm{t}} A v)^{\mathrm{t}} = v^{\mathrm{t}} A^{\mathrm{t}} \overline{v} = v^{\mathrm{t}} A \overline{v} = v^{\mathrm{t}} \overline{A v} = \overline{\lambda} v^{\mathrm{t}} \overline{v} = \overline{\lambda} \sum_{i=1}^{n} |v_i|^2$$
ist $\lambda = \overline{\lambda} \in \mathbb{R}$. Nun ist λ auch ein Eigenwert von f und wir können einen entsprechenden Eigenvektor $b_1 \in V$ wählen. Nach Normierung ist $|b_1| = 1$. Für $U := b_1^{\perp}$ gilt $V = \langle b_1 \rangle \oplus U$ nach Lemma 11.15. Für $u \in U$ gilt
$$[f(u), b_1] = [u, f(b_1)] = \lambda[u, b_1] = 0,$$
d. h. $f(u) \in U$. Daher liegt die Einschränkung $g := f_{|U}$ in $\mathrm{End}(U)$. Offenbar ist g auch symmetrisch. Wegen $\dim U = n-1$ besitzt U nach Induktion eine Orthonormalbasis $b_2, \ldots, b_n \in U$ aus Eigenvektoren von g. Wegen $f(b_i) = g(b_i)$ sind $b_2, \ldots, b_n$ auch Eigenvektoren von f. Insgesamt ist $b_1, \ldots, b_n$ eine Orthonormalbasis von V aus Eigenvektoren von f. □

Folgerung 11.39. *Genau dann ist $A \in \mathbb{R}^{n \times n}$ symmetrisch, wenn eine Matrix $S \in \mathrm{O}(n, \mathbb{R})$ mit $S^{\mathrm{t}} A S = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ existiert. Ggf. sind $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A.*

Beweis. Ist $D := S^{\mathrm{t}} A S = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, so gilt
$$A^{\mathrm{t}} = (SDS^{\mathrm{t}})^{\mathrm{t}} = SD^{\mathrm{t}} S^{\mathrm{t}} = SDS^{\mathrm{t}} = A,$$
d. h. A ist symmetrisch. Für die Umkehrung sei $V := \mathbb{R}^n$ und $f \in \mathrm{End}(V)$ mit $[f] = A$. Sei A symmetrisch. Nach Lemma 11.20 ist f symmetrisch. Nach dem Hauptachsensatz existiert eine Orthonormalbasis B von V bestehend aus Eigenvektoren von f. Sei E die Standardbasis von V und $S := {}_E\Delta_B \in \mathrm{GL}(n, \mathbb{R})$. Da die Spalten von S aus den Vektoren von B bestehen, gilt $S \in \mathrm{O}(n, \mathbb{R})$. Aus Folgerung 7.27 folgt
$$S^{\mathrm{t}} A S = S^{-1} A S = {}_B[f]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$
mit den Eigenwerten $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ von f. □

Bemerkung 11.40. Sei $f \in \mathrm{End}(V)$ symmetrisch. Seien $v, w \in V$ Eigenvektoren von f zu verschiedenen Eigenwerten λ bzw. μ. Dann gilt $\lambda[v,w] = [f(v),w] = [v,f(w)] = \mu[v,w]$ und es folgt $[v,w] = 0$. Merkregel: Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal. Man kann die gesuchte Orthonormalbasis von V also berechnen, indem man das Gram-Schmidt-Verfahren auf jeden Eigenraum anwendet.

Beispiel 11.41. In Beispiel 8.4 hatten wir die Abbildung $f \in \mathrm{End}(\mathbb{R}^3)$ mit symmetrischer Matrix

$$A := [f] = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

untersucht. Es gilt

$$E_1(A) = \langle \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \rangle, \qquad\qquad E_4(A) = \langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \rangle$$

(beachte: $\mathrm{tr}(A) = 6 = 1 + 1 + 4$). Das Gram-Schmidt-Verfahren für $E_1(A)$ liefert

$$b_1 := (1, 0, -1), \qquad\qquad b_2 := (0, 1, -1) - \frac{1}{2}(1, 0, -1) = \frac{1}{2}(-1, 2, -1).$$

Der Eigenvektor zum Eigenwert 4 muss nur normiert werden. Insgesamt erhält man die Orthonormalbasis $\frac{1}{\sqrt{2}}(1, 0, -1)$, $\frac{1}{\sqrt{6}}(-1, 2, -1)$, $\frac{1}{\sqrt{3}}(1, 1, 1)$ aus Eigenvektoren von f. Für die orthogonale Matrix

$$S := \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \end{pmatrix}$$

gilt $S^{\mathrm{t}} A S = S^{-1} A S = \mathrm{diag}(1, 1, 4)$. Man kann das Ergebnis nutzen, um Wurzeln von Matrizen zu ziehen. Für $\sqrt{A} := S\,\mathrm{diag}(1, 1, 2) S^{\mathrm{t}}$ gilt nämlich

$$\sqrt{A}^2 = S\,\mathrm{diag}(1, 1, 2) S^{\mathrm{t}} S\,\mathrm{diag}(1, 1, 2) S^{\mathrm{t}} = S\,\mathrm{diag}(1, 1, 2)^2 S^{\mathrm{t}} = S\,\mathrm{diag}(1, 1, 4) S^{\mathrm{t}} = A.$$

Mehr dazu in Satz 12.46 und Satz 14.42.

Satz 11.42 (Euler). *Sei V ein 3-dimensionaler euklidischer Raum und $f \in \mathrm{O}(V)$. Dann existiert eine Orthonormalbasis B von V und ein Winkel φ mit*

$$_B[f]_B = \begin{pmatrix} \pm 1 & 0 \\ 0 & D(\varphi) \end{pmatrix}.$$

Beweis. Da 3 ungerade ist, besitzt χ_f eine Nullstelle $\lambda \in \mathbb{R}$ nach Beispiel 11.36 (oder Bemerkung 11.37). Nach Bemerkung 11.21 ist $\lambda \in \{\pm 1\}$. Sei b_1 ein entsprechender normierter Eigenvektor und $U := b_1^{\perp}$. Für $u \in U$ gilt

$$[f(u), b_1] = [f(u), \lambda^2 b_1] = \lambda[f(u), f(b_1)] = \lambda[u, b_1] = 0$$

und $f(u) \in U$. Daher liegt die Einschränkung $g := f_{|U}$ in $\mathrm{O}(U)$. Für eine Orthonormalbasis $C := \{b_2, b_3\}$ von U ist $_C[g]_C \in \mathrm{O}(2, \mathbb{R})$ nach Lemma 11.20. Im Fall $\det(g) = 1$ ist $_C[g]_C = D(\varphi)$ nach Beispiel 11.22 und die Behauptung folgt mit $B := \{b_1, b_2, b_3\}$. Sei nun $\det(g) = -1$. Nach Beispiel 11.22 ist g eine Spiegelung mit Eigenwerten 1 und -1. Da dies auch Eigenwerte von f sind, können wir λ durch $-\lambda$ ersetzen und b_1 entsprechend wählen. Anschließend ist $\det(g) = 1$ und die Behauptung folgt wie zuvor. $\square$

Bemerkung 11.43. Die Matrizen in Satz 11.42 mit Determinante 1 (also von der Form $\operatorname{diag}(1, D(\varphi))$) entsprechen Drehungen um den Winkel φ, wobei die Drehachse vom ersten Basisvektor aufgespannt wird. Nach dem Determinantensatz ist die Komposition von Drehungen wieder eine Drehung (im 2-dimensionalen Raum ist dies klar wegen $D(\varphi)D(\psi) = D(\varphi + \psi)$). Achtung: Die orthogonalen Abbildungen mit Determinante -1 sind nicht unbedingt Spiegelungen. Es gibt auch sogenannte *Drehspiegelungen*, d. h. Kompositionen einer Drehung mit einer Spiegelung.

Beispiel 11.44. Während eines Fußballspiels gibt es einen Punkt auf der Oberfläche des Fußballs, der sich zu zwei verschiedenen Zeitpunkten exakt am gleichen Ort befindet. Begründung: Der Mittelpunkt des Fußballs liegt zu Beginn der ersten und zweiten Halbzeit auf dem Anstoßpunkt. Dazwischen führt der Ball mit jedem Schuss eine Drehung (und Translation) aus. Da die Komposition von Drehungen wieder eine Drehung ist, wird auch die Transformation auf dem Anstoßpunkt durch eine Drehung f beschrieben. Die Drehachse von f schneidet die Oberfläche des Balls an zwei Punkten. Diese bleiben also fest.

Definition 11.45. Sei V ein euklidischer Raum und $v \in V \setminus \{0\}$. Man nennt

$$S_v \colon V \to V, \qquad w \mapsto w - 2\frac{[w,v]}{[v,v]}v$$

die *Spiegelung* an $v^\perp$.

Bemerkung 11.46. Für $\lambda \in \mathbb{R}^\times$ ist $S_{\lambda v} = S_v$. Wir können daher annehmen, dass v normiert ist. Dann gilt

$$S_v(w) = w - 2[w,v]v$$

für alle $w \in V$. Offenbar ist S_v linear und $S_v(w) = w$ gilt genau dann, wenn $w \in v^\perp$. Außerdem ist $S_v(v) = -v$. Geometrisch entspricht S_v also einer Spiegelung an der Hyperebene $v^\perp$. Bzgl. einer geeigneten Basis hat S_v die Darstellungsmatrix $\operatorname{diag}(-1, 1, \ldots, 1)$ (vgl. Aufgabe 42). Insbesondere ist S_v orthogonal und $\det S_v = -1$. Außerdem gilt $S_v \circ S_v = \mathrm{id}_V$.

Satz 11.47 (Cartan-Dieudonné). *Sei V ein n-dimensionaler euklidischer Raum und $f \in \mathrm{O}(V)$. Dann ist f eine Komposition von höchstens n Spiegelungen.*

Beweis. Im Fall $f = \mathrm{id}_V$ ist f die Komposition von 0 Spiegelungen. Sei daher $f \neq \mathrm{id}_V$ und $w \in V$ mit $f(w) \neq w$. Im Fall $n = 1$ ist $f = -\mathrm{id}_V = S_1$. Sei nun $n \geq 2$ und die Behauptung für $n-1$ bereits beweisen. Für $v := f(w) - w \neq 0$ gilt

$$[f(w) + w, v] = [f(w) + w, f(w) - w] = [f(w), f(w)] - [w,w] = 0,$$

d. h. $f(w) + w \in v^\perp$. Also ist

$$(S_v \circ f)(w) = \frac{1}{2}\Big(S_v(f(w) + w) + S_v(v)\Big) = \frac{1}{2}\big(f(w) + w - v\big) = w.$$

Sei $U := w^\perp$. Wegen $g := S_v \circ f \in \mathrm{O}(V)$ ist $g(U) = U$. Nach Induktion ist $g_{|U}$ eine Komposition von höchstens $n-1$ Spiegelungen $S_{w_1}, \ldots, S_{w_k}$ mit $w_1, \ldots, w_k \in U$. Man kann S_{w_i} mit der gleichen Formel auch als Spiegelung von V auffassen. Wegen $w \in U^\perp$ gilt $S_{w_i}(w) = w$ für $i = 1, \ldots, k$. Damit ist $f = S_v \circ g$ ein Produkt von höchstens n Spiegelungen. □

Bemerkung 11.48. Nach Bemerkung 11.46 ist $f \in \mathrm{SO}(V)$ ein Produkt einer geraden Anzahl an Spiegelungen. Insbesondere benötigt man im Fall $\dim V = 2n + 1$ nur $2n$ Spiegelungen.

12 Bilinearformen

12.1 Gram-Matrizen

Bemerkung 12.1. Wir verallgemeinern in diesem Kapitel Teile der euklidischen Geometrie auf beliebige Körper. Stets sei V ein endlich-dimensionaler K-Vektorraum.

Definition 12.2.

- Eine *Bilinearform* auf V ist eine Abbildung $\beta\colon V \times V \to K$, die in der ersten und zweiten Komponente linear ist, d. h.

$$\beta(\lambda u + v, w) = \lambda\beta(u, w) + \beta(v, w),$$
$$\beta(u, \lambda v + w) = \lambda\beta(u, v) + \beta(u, w).$$

  für alle $u, v, w \in V$ und $\lambda \in K$.

- Man nennt β
  - *symmetrisch*, falls $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$ gilt.
  - *antisymmetrisch*, falls $\beta(v, w) = -\beta(w, v)$ für alle $v, w \in V$ gilt.
  - *alternierend*, falls $\beta(v, v) = 0$ für alle $v \in V$ gilt.
  - *ausgeartet*, falls $\exists v \in V \setminus \{0\} : \forall w \in V : \beta(v, w) = 0$.

Beispiel 12.3.

(a) Die triviale Bilinearform $\beta(v, w) = 0$ für alle $v, w \in V$ ist symmetrisch, antisymmetrisch und alternierend. Für $V \neq \{0\}$ ist sie ausgeartet. In der Regel interessieren wir uns für nicht-ausgeartete Bilinearformen.

(b) Ein Skalarprodukt $\beta(v, w) = [v, w]$ auf einem euklidischen Raum V ist eine symmetrische Bilinearform. Für $V \neq \{0\}$ ist β nicht-ausgeartet, denn $[v, v] = |v|^2 > 0$ für $v \neq 0$.

(c) Für $V = K^n$ und $A \in K^{n \times n}$ definiert $\beta_A(v, w) := vAw^{\mathrm{t}}$ eine Bilinearform. Dies folgt aus den Rechenregeln für Matrizen. Wir werden in Satz 12.9 sehen, dass jede Bilinearform diese Form hat.

(d) Die Wahl $A = 1_n$ in (c) führt zur symmetrischen Bilinearform

$$\beta(v, w) = vw^{\mathrm{t}} = \sum_{i=1}^{n} v_i w_i.$$

Im Fall $K = \mathbb{R}$ erhält man das Standardskalarprodukt.

(e) Sei $V := K^2$ und $\beta(v, w) := \det\left(\begin{smallmatrix} v \\ w \end{smallmatrix}\right)$ für $v, w \in V$. Nach Lemma 9.5 ist β eine alternierende Bilinearform. Im Allgemeinen ist det eine *Multilinearform*.

Bemerkung 12.4.

(a) Jede alternierende Bilinearform β ist antisymmetrisch, denn aus

$$0 = \beta(v+w, v+w) = \beta(v,v) + \beta(v,w) + \beta(w,v) + \beta(w,w) = \beta(v,w) + \beta(w,v)$$

folgt $\beta(v,w) = -\beta(w,v)$. Für $K \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ gilt auch die Umkehrung, denn aus $\beta(v,v) = -\beta(v,v)$ folgt $\beta(v,v) = 0$. Für $K = \mathbb{F}_2$ ist diese Schlussweise ungültig, denn $1+1=0$. Hier ist symmetrisch und antisymmetrisch sogar gleichbedeutend. Insbesondere ist die Bilinearform $\beta(v,w) = vw^{\mathrm{t}}$ auf $V = K^n$ (anti)symmetrisch, aber nicht alternierend. Um diesen Fall zu vermeiden, werden wir im Folgenden oft $1+1 \neq 0$ annehmen.[1]

(b) Jede symmetrische Bilinearform β definiert eine *quadratische Form* $q\colon V \to K$ durch $q(v) := \beta(v,v)$. Ist $1+1 \neq 0$, so kann β aus q durch *Polarisierung* zurückgewinnen:

$$\beta(v,w) = \frac{1}{2}\Big(q(v+w) - q(v) - q(w)\Big) \qquad (v, w \in V).$$

Satz 12.5. *Die Menge* $\mathrm{Bil}(V)$ *aller Bilinearformen auf V ist Unterraum von* $\mathrm{Abb}(V \times V, K)$.

Beweis. Die Nullabbildung ist die triviale Bilinearform. Sei $\beta, \gamma \in \mathrm{Bil}(V)$ und $\lambda \in K$. Wie in Satz 7.13 zeigt man, dass $\beta + \gamma$ und $\lambda\beta$ in der ersten und zweiten Komponenten linear sind (allerdings ist $\mathrm{Bil}(V) \nsubseteq \mathrm{Hom}(V \times V, K)$). Dies zeigt $\beta + \gamma, \lambda\beta \in \mathrm{Bil}(V)$. □

Bemerkung 12.6. Offenbar bilden die symmetrischen (antisymmetrischen, alternierenden) Bilinearformen jeweils einen Unterraum von $\mathrm{Bil}(V)$. Satz 12.9 gibt Aufschluss über die Dimension dieser Unterräume.

Satz 12.7. *Sei $\beta \in \mathrm{Bil}(V)$. Für $v \in V$ sei $F_v\colon V \to K$, $w \mapsto \beta(v,w)$. Dann ist $F\colon V \to V^*$, $v \mapsto F_v$ ist ein Homomorphismus. Genau dann ist β nicht-ausgeartet, wenn F ein Isomorphismus ist.*

Beweis. Wegen der Bilinearität von β ist $F_v \in V^*$ und F ein Homomorphismus. Offenbar ist β genau dann nicht-ausgeartet, wenn F injektiv ist. Wegen $\dim V^* = \dim V$ ist injektiv äquivalent zu bijektiv. □

Definition 12.8. Sei β eine Bilinearform auf V. Sei $B := \{b_1, \ldots, b_n\}$ eine Basis von V. Man nennt

$$_B[\beta]_B := \big(\beta(b_i, b_j)\big)_{i,j=1}^n \in K^{n\times n}$$

die *Gram-Matrix* von β bzgl. B. Ist $V = K^n$ und B die Standardbasis, so sei $[\beta] := {_B[\beta]_B}$.

Satz 12.9. *Sei B eine Basis von V mit $|B| = n$. Dann ist die Abbildung*

$$_B[.]_B\colon \mathrm{Bil}(V) \to K^{n\times n}, \qquad \beta \mapsto {_B[\beta]_B}$$

ein Isomorphismus. Für $A := {_B[\beta]_B}$ gilt:

(a) β symmetrisch $\iff$ A symmetrisch.

(b) β antisymmetrisch $\iff$ $A^{\mathrm{t}} = -A$.[2]

[1] Neben $\mathbb{F}_2$ gibt es viele (auch unendliche) Körper mit $1+1=0$.

[2] Man nennt A *antisymmetrisch* oder *schiefsymmetrisch*.

(c) *β nicht-ausgeartet $\iff$ A invertierbar.*

Beweis. Für $\beta, \gamma \in \mathrm{Bil}(V)$ und $\lambda \in K$ gilt

$$_B[\lambda\beta + \gamma]_B = ((\lambda\beta + \gamma)(b_i, b_j)) = \lambda(\beta(b_i, b_j)) + (\gamma(b_i, b_j)) = \lambda _B[\beta]_B + {}_B[\gamma]_B,$$

d. h. $_B[.]_B$ ist linear. Ist $_B[\beta]_B = 0$, so gilt

$$\beta\Big(\sum \lambda_i b_i, \sum \mu_j b_j\Big) = \sum_{i,j} \lambda_i \mu_j \beta(b_i, b_j) = 0$$

für alle $\lambda_i, \mu_j \in K$. Also ist $\beta = 0$ und $_B[.]_B$ ist injektiv. Für die Surjektivität sei $A \in K^{n\times n}$ gegeben. Aus der Linearität der Koordinatendarstellung $v \mapsto {}_B[v]$ (siehe Beweis von Satz 7.10) folgt, dass

$$\beta(v, w) := {}_B[v] A _B[v]^{\mathrm{t}}$$

eine Bilinearform definiert. Wegen $\beta(b_i, b_j) = e_i A e_j^{\mathrm{t}} = a_{ij}$ ist $_B[\beta]_B = A$. Also ist $_B[.]_B$ ein Isomorphismus.

(a) Ist β symmetrisch, so gilt $a_{ij} = \beta(b_i, b_j) = \beta(b_j, b_i) = a_{ji}$ für alle $1 \le i, j \le n$. Daher ist A symmetrisch. Sei umgekehrt A symmetrisch. Dann ist $\beta(b_i, b_j) = a_{ij} = a_{ji} = \beta(b_j, b_i)$. Für $_B[v] = (v_1, \ldots, v_n)$ und $_B[w] = (w_1, \ldots, w_n)$ folgt

$$\beta(v, w) = \sum_{1\le i,j\le n} v_i w_j \beta(b_i, b_j) = \sum_{1\le i,j\le n} v_i w_j \beta(b_j, b_i) = \beta(w, v),$$

d. h. β ist symmetrisch.

(b) Ist β antisymmetrisch, so gilt $a_{ij} = \beta(b_i, b_j) = -\beta(b_j, b_i) = -a_{ji}$ und $A^{\mathrm{t}} = -A$. Die Umkehrung zeigt man wie in (a).

(c) Sei β ausgeartet und $v \ne 0$ mit $\beta(v, w) = 0$ für alle $w \in W$. Für $x := {}_B[v]^{\mathrm{t}}$ und $i = 1, \ldots, n$ gilt dann

$$e_i A x = \sum_{j=1}^n \beta(b_i, b_j) x_j = \beta(b_i, v) = 0.$$

Dies zeigt $Ax = 0$. Nach Satz 6.6 ist A nicht invertierbar. Ist umgekehrt A nicht invertierbar, so existiert $x \in K^{n\times 1} \setminus \{0\}$ mit $Ax = 0$. Für $v = \sum_{i=1}^n x_i b_i$ gilt dann $\beta(v, w) = 0$ für alle $w \in V$. Daher ist β ausgeartet. □

Beispiel 12.10. Sei $1 + 1 \ne 0$ in K und $n := \dim V$ ungerade. Sei $\beta \in \mathrm{Bil}(V)$ antisymmetrisch mit Gram-Matrix A bzgl. einer beliebigen Basis. Nach Satz 12.9 gilt

$$\det(A) = \det(A^{\mathrm{t}}) = \det(-A) \overset{9.8}{=} (-1)^n \det(A) = -\det(A)$$

und $\det(A) = 0$. Daher muss β ausgeartet sein.

Satz 12.11. *Seien B und C Basen von V. Für alle $\beta \in \mathrm{Bil}(V)$ gilt*

$$\boxed{_C[\beta]_C = {}_B\Delta_C^{\mathrm{t}} {}_B[\beta]_B {}_B\Delta_C.}$$

Beweis. Sei $B = \{b_1, \ldots, b_n\}$, $C = \{c_1, \ldots, c_n\}$ und $S = (s_{ij}) = {}_B\Delta_C$. Für $1 \le i, j \le n$ gilt

$$\beta(c_i, c_j) = \sum_{k,l=1}^{n} s_{ki}\beta(b_k, b_l)s_{lj} = \sum_{k=1}^{n} s_{ki}({}_B[\beta]_B S)_{kj} = (S^{\mathrm{t}}{}_B[\beta]_B S)_{ij}. \qquad \square$$

Definition 12.12. Matrizen $A, B \in K^{n\times n}$ heißen *kongruent*, falls ein $S \in \mathrm{GL}(n, K)$ mit $B = SAS^{\mathrm{t}}$ existiert.[3]

Bemerkung 12.13.

(a) Wie in Bemerkung 7.30 zeigt man, dass die Kongruenz von Matrizen eine Äquivalenzrelation ist. Nach Satz 12.9 und Satz 12.11 bestimmt jede Bilinearform auf V eine Kongruenzklasse von Gram-Matrizen. Wir zeigen im nächsten Abschnitt, dass man Bilinearformen ähnlich wie Endomorphismen diagonalisieren kann.

(b) Nach Lemma 5.15 ist

$$\mathrm{rk}(B) = \mathrm{rk}(SAS^{\mathrm{t}}) \le \min\{\mathrm{rk}(S), \mathrm{rk}(AS^{\mathrm{t}})\} = \mathrm{rk}(AS^{\mathrm{t}}) \le \min\{\mathrm{rk}(A), \mathrm{rk}(S^{\mathrm{t}})\} = \mathrm{rk}(A).$$

Aus Symmetriegründen folgt $\mathrm{rk}(B) = \mathrm{rk}(A)$, d. h. kongruente Matrizen haben den gleichen Rang. Andererseits haben A und B nicht unbedingt die gleiche Determinante, denn $\det(B) = \det(SAS^{\mathrm{t}}) = \det(S)^2 \det(A)$. Wegen Bemerkung 10.35 haben A und B in der Regel auch nicht die gleichen Eigenwerte. Bereits im Fall $n = 1$ sieht man, dass A und B ebenso nicht die gleiche Spur haben.

(c) Nach dem Hauptachsensatz ist jede symmetrische reelle Matrix kongruent und ähnlich zu einer Diagonalmatrix.

12.2 Sylvesters Trägheitssatz

Definition 12.14. Sei $\beta \in \mathrm{Bil}(V)$ und $S \subseteq V$. Wie in euklidischen Räumen definieren wir das *orthogonale Komplement* von S durch

$$S^{\perp} := \{v \in V : \forall s \in S : \beta(v, s) = 0\}.$$

Bemerkung 12.15.

(a) In der Definition von $S^{\perp}$ müsste man streng genommen zwischen „links-orthogonal“ und „rechts-orthogonal“ ($\forall s \in S : \beta(s, v) = 0$) unterscheiden. In der Regel werden wir annehmen, dass β (anti)symmetrisch ist, sodass beide Versionen äquivalent sind.

(b) Für $U \le V$ ist $U^{\perp} \le V$. Genau dann ist β ausgeartet, falls $V^{\perp} \ne \{0\}$ gilt.

Satz 12.16. *Sei $\beta \in \mathrm{Bil}(V)$ nicht-ausgeartet. Für $U \le V$ gilt $\dim V = \dim U + \dim U^{\perp}$.*

Beweis. Sei $F\colon V \to V^*$, $v \mapsto F_v$ der Isomorphismus aus Satz 12.7. Dann gilt

$$F_v \in F(U^{\perp}) \iff \forall u \in U : F_v(u) = \beta(v, u) = 0 \iff F_v \in U^0,$$

d. h. $F(U^{\perp}) = U^0$. Die Behauptung folgt aus Lemma 7.43. $\square$

[3] Im Gegensatz zu Ähnlichkeit führen wir kein Symbol für die Kongruenz von Matrizen ein.

Bemerkung 12.17. Achtung: Im Gegensatz zu euklidischen Räumen gilt im Allgemeinen weder $U \cap U^\perp = \{0\}$ noch $V = U \oplus U^\perp$. Sei zum Beispiel $V = \mathbb{F}_2^2$ und $[\beta] = 1_2$. Für $U = \langle (1, 1) \rangle$ gilt $U = U^\perp$.

Definition 12.18. Sei $\beta \in \mathrm{Bil}(V)$ symmetrisch. Eine Basis $B = \{b_1, \ldots, b_n\}$ von V heißt *Orthogonalbasis* bzgl. β, falls $\beta(b_i, b_j) = 0$ für $i \neq j$ gilt. Ist zusätzlich $\beta(b_i, b_i) = 1$, so nennt man B eine *Orthonormalbasis* bzgl. β.

Bemerkung 12.19.

(a) Offenbar ist B genau dann *Orthogonalbasis* (bzw. *Orthonormalbasis*) bzgl. β, falls ${}_B[\beta]_B$ eine Diagonalmatrix (bzw. ${}_B[\beta]_B = 1_n$) ist.

(b) Nach Folgerung 11.11 besitzt jedes Skalarprodukt auf einem euklidischen Raum eine Orthonormalbasis.

Satz 12.20. *Sei $1+1 \neq 0$ in K. Dann besitzt jede symmetrische Bilinearform auf V eine Orthogonalbasis.*

Beweis. Sei $\beta \in \mathrm{Bil}(V)$ symmetrisch. Wir argumentieren durch Induktion nach $n := \dim V$. Im Fall $n \leq 1$ oder $\beta = 0$ ist jede Basis eine Orthogonalbasis. Sei also $n \geq 2$ und $\beta(v, w) \neq 0$ für gewisse $v, w \in V$. Im Fall $\beta(v, v) = \beta(w, w) = 0$ ist

$$\beta(v + w, v + w) = \beta(v, v) + \beta(v, w) + \beta(w, v) + \beta(w, w) = 2\beta(v, w) \neq 0$$

(beachte $1 + 1 \neq 0$). In jedem Fall existiert also ein $b_1 \in V \setminus \{0\}$ mit $\beta(b_1, b_1) \neq 0$. Sei $U := \langle b_1 \rangle$. Wegen $U \cap U^\perp = \{0\}$ ist $V = U \oplus U^\perp$. Die Einschränkung von β auf $U \times U$ ist eine symmetrische Bilinearform. Nach Induktion besitzt $U^\perp$ eine Orthogonalbasis $b_2, \ldots, b_n$. Nun ist $b_1, \ldots, b_n$ eine Orthogonalbasis von V. $\square$

Bemerkung 12.21.

(a) Die Matrix-Version von Satz 12.20 lautet: Jede symmetrische Matrix ist zu einer Diagonalmatrix kongruent (falls $1 + 1 \neq 0$). Wir werden sehen, dass man die Diagonaleinträge speziell wählen kann.

(b) Ist $1 + 1 = 0$ in K, so wird Satz 12.20 falsch: Die Bilinearform β mit $[\beta] = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ besitzt keine Orthogonalbasis, denn $\beta(v, v) = 0$ für alle $v \in K^2$.

(c) Das Gram-Schmidt-Verfahren zur Berechnung von Orthogonalbasen funktioniert in dieser Allgemeinheit nicht immer, denn man muss durch $\beta(b_i, b_i)$ teilen, aber dieser Wert kann 0 sein. Wir modifizieren stattdessen den Gauß-Algorithmus wie folgt:

   (1) Sei $A := [\beta]$. Nehmen wir induktiv an, dass die Zeilen und Spalten $1, \ldots, k-1$ von A bereits die gewünschte Diagonalgestalt haben (zu Beginn sei $k = 1$).

   (2) Nehmen wir zunächst $a_{kk} \neq 0$ an. Dann kann man $a_{ik} = 0$ für $i = k{+}1, \ldots, n$ erreichen, indem man wie gewohnt ein Vielfaches der k-ten Zeile auf die darunterliegenden Zeilen addiert. Dies entspricht der Multiplikation mit Elementarmatrizen $S_1, \ldots, S_{n-k}$ von links. Anschließend setze man $a_{ki} = 0$ für $i = k + 1, \ldots, n$. Dies entspricht der Multiplikation von $S_1^{\mathrm{t}}, \ldots, S_{n-k}^{\mathrm{t}}$ von rechts (in beliebiger Reihenfolge). Insgesamt wird A zu SAS^{t}, wobei $S = S_1 \ldots S_n$. Wegen $(SAS^{\mathrm{t}})^{\mathrm{t}} = (S^{\mathrm{t}})^{\mathrm{t}} A^{\mathrm{t}} S^{\mathrm{t}} = SAS^{\mathrm{t}}$ bleibt A durch dieses Vorgehen symmetrisch.

(3) Sei nun $a_{kk} = 0$. Existiert ein $l > k$ mit $a_{ll} \neq 0$, so tausche man die Zeilen k und l und anschließend die Spalten k und l. Dadurch werden a_{kk} und a_{ll} vertauscht und A bleibt symmetrisch. Man ist nun in Situation (2).

(4) Sei schließlich $a_{ii} = 0$ für $i = k, \ldots, n$. Ist $a_{ij} = 0$ für alle $k \leq i, j \leq n$, so sind wir fertig. Sei daher $a_{ij} \neq 0$ für gewisse $k \leq i < j \leq n$. Wir addieren die i-te Zeile zur j-ten Zeile und anschließend die i-te Spalte zur j-ten Spalte. Der Eintrag an Position (j, j) wird dadurch zu $2a_{ij} \neq 0$ (beachte $1 + 1 \neq 0$). Also ist man in Situation (3).

(5) Zur Bestimmung der Orthogonalbasis B führe man alle Zeilenoperationen (aber nicht die Spaltenoperationen) an der Einheitsmatrix aus. Dadurch entsteht die Matrix $S \in \mathrm{GL}(n, K)$, sodass SAS^{t} eine Diagonalmatrix ist (vgl. Satz 6.17). Nach Satz 12.11 sind die Zeilen von S die Vektoren von B.

Beispiel 12.22.

$$
\begin{aligned}
(A|1_3) &= \left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \\
&\sim \left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 1 & 0 \\ 0 & -1/2 & 0 & -1/2 & 1/2 & 0 \\ 0 & 0 & -1 & -1 & -1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1/2 & 0 & -1/2 & 1/2 & 0 \\ 0 & 0 & -1 & -1 & -1 & 1 \end{array}\right) \\
S &= \begin{pmatrix} 1 & 1 & 0 \\ -1/2 & 1/2 & 0 \\ -1 & -1 & 1 \end{pmatrix}
\end{aligned}
$$

Satz 12.23 (Sylvesters Trägheitssatz). *Sei β eine symmetrische Bilinearform auf $V = \mathbb{R}^n$. Dann existiert eine Basis B von V mit*

$$_B[\beta]_B = \mathrm{diag}(1_r, -1_s, 0_t).$$

Die Zahlen r, s, t hängen nicht von B ab.

Beweis. Nach Satz 12.20 existiert eine Orthogonalbasis $B = \{b_1, \ldots, b_n\}$ von V. Nach Umsortierung können wir

$$\beta(b_i, b_i) \begin{cases} > 0 & \text{für } i = 1, \ldots, r, \\ < 0 & \text{für } i = r+1, \ldots, r+s, \\ = 0 & \text{für } i = r+s+1, \ldots, r+s+t = n \end{cases}$$

annehmen. Für $i = 1, \ldots, r+s$ ersetzen wir b_i durch $\frac{1}{\sqrt{|\beta(b_i,b_i)|}} b_i$. Dann gilt $\beta(b_i, b_i) = 1$ für $i = 1, \ldots, r$ und $\beta(b_i, b_i) = -1$ für $i = r+1, \ldots, r+s$. Damit ist die Existenz von B gezeigt.

Wegen $V^\perp = \langle b_{r+s+1}, \ldots, b_n \rangle$ hängt $t = \dim V^\perp$ nicht von B ab. Sei $V_+ := \langle b_1, \ldots, b_r \rangle$. Für $v = \sum_{i=1}^r \lambda_i b_i \in V_+$ mit $\lambda_i \in \mathbb{R}$ gilt $\beta(v, v) = \sum_{i=1}^n \lambda_i^2 \geq 0$ mit Gleichheit genau dann, wenn $v = 0$. Sei $B' = \{b'_1, \ldots, b'_n\}$ eine weitere Basis von V mit

$$_{B'}[\beta]_{B'} = \mathrm{diag}(1_{r'}, -1_{s'}, 0_t).$$

Sei $V'_{\leq 0} := \langle b'_{r'+1}, \ldots, b'_n \rangle$. Für $v \in V_+ \cap V'_{\leq 0}$ gilt einerseits $\beta(v, v) \geq 0$ und andererseits $\beta(v, v) \leq 0$. Dies zeigt $\beta(v, v) = 0$ und $v = 0$. Also ist $V_+ \cap V'_{\leq 0} = \{0\}$ und

$$r + s' + t = \dim(V_+ \oplus V'_{\leq 0}) \leq \dim V = n = r + s + t.$$

Es folgt $s' \leq s$. Aus Symmetriegründen ist $s' = s$ und $r' = r$. □

Definition 12.24. In der Situation von Satz 12.23 nennt man $\operatorname{ind}(\beta) := (r, s, t)$ den *Index* von β.[4] Für eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ definiert man $\operatorname{ind}(A) := \operatorname{ind}(\beta)$ durch der Bilinearform β mit $[\beta] = A$.

Beispiel 12.25.

(a) Jedes Skalarprodukt auf $\mathbb{R}^n$ hat Index $(n, 0, 0)$.

(b) Ist $\operatorname{ind}(\beta) = (r, s, t)$, so ist β genau dann ausgeartet, wenn $t > 0$ gilt.

(c) In der Relativitätstheorie betrachtet man den *Minkowski-Raum* $\mathbb{R}^4$ bzgl. einer Bilinearform β mit Index $(3, 1, 0)$. Die vierte Dimension beschreibt dabei die Zeit.

Bemerkung 12.26. Zur Berechnung des Index einer Matrix A kann man zunächst eine Orthogonalbasis $\tilde{B}$ mit dem modifizierten Gauß-Algorithmus aus Bemerkung 12.21 bestimmen (beachte: $1 + 1 \neq 0$ in $\mathbb{R}$). Anschließend zählt man die positiven und negativen Einträge auf der Hauptdiagonale. Dividiert man die Vektoren in $\tilde{B}$ jeweils durch die Wurzel des entsprechenden Diagonaleintrags, so erhält man die Basis B wie im Beweis von Satz 12.23. Die Matrix aus Beispiel 12.22 hat zum Beispiel Index $(1, -2, 0)$ bzgl.

$$B = \Big\{ \frac{1}{\sqrt{2}}(1, 1, 0),\ \frac{1}{\sqrt{2}}(-1, 1, 0),\ (-1 - 1, 1) \Big\}.$$

Satz 12.27. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch mit Index (r, s, t). Dann ist r die Anzahl der positiven Eigenwerte und s die Anzahl der negativen Eigenwerte von A jeweils gezählt mit (algebraischen) Vielfachheiten.*

Beweis. Nach der Matrix-Version des Hauptachsensatzes ist A zu $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ kongruent, wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A sind (die algebraische Vielfachheit jedes Eigenwerts stimmt mit der geometrischen Vielfachheit überein). Man kann das Argument aus dem Beweis von Satz 12.23 auf diese Matrix anwenden. Die positiven λ_i werden dabei zu 1 transformiert und die negativen zu -1. □

Folgerung 12.28. *Für symmetrische Matrizen $A, B \in \mathbb{R}^{n\times n}$ sind die folgenden Aussagen äquivalent:*

(1) A und B sind kongruent.

(2) $\operatorname{ind}(A) = \operatorname{ind}(B)$.

(3) A und B haben die gleiche Anzahl an positiven Eigenwerten und die gleiche Anzahl an negativen Eigenwerten.[5]

[4] Dieser Begriff ist in der Literatur nicht einheitlich! Manche Autoren nennen $r - s$ die *Signatur* von β. Aus Dimension, Rang und Signatur lässt sich der Index bestimmen.

[5] Das beschreibt die *Trägheit* in Sylvesters Satz.

Beweis. Da kongruente Matrizen die gleiche Bilinearform beschreiben, gilt (1)⇒(2). Aus Satz 12.27 folgt (2)⇒(3). Gilt (3), so haben A und B den gleichen Index nach Satz 12.27. Also sind A und B kongruent, d. h. (1) gilt. □

Bemerkung 12.29.

(a) Für jedes $m \le n$ gibt es genau $m+1$ Kongruenzklassen von symmetrischen $n \times n$-Matrizen mit Rang m (nämlich mit Index $(i, m-i, n-m)$ für $i = 0, \dots, m$). Daher gibt es

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

symmetrische Bilinearformen auf $\mathbb{R}^n$ bis auf Basiswahl (vgl. Beispiel 1.16).

(b) Arbeitet man über $\mathbb{C}$ anstatt $\mathbb{R}$, so kann man stets ${}_B[\beta]_B = \operatorname{diag}(1_r, 0_t)$ erreichen. Man kann nämlich jeden Basisvektor b mit $\beta(b,b) < 0$ durch $\frac{\mathrm{i}}{\sqrt{|\beta(b,b)|}}b$ ersetzen. Daher gibt es nur $n+1$ symmetrische Bilinearformen auf $\mathbb{C}^n$ bis auf Basiswahl. Wir beweisen in Satz 13.25 Sylvesters Trägheitssatz über $\mathbb{C}$ mit einem anderen Kongruenzbegriff.

Satz 12.30. *Sei V ein beliebiger K-Vektorraum. Sei $\beta \in \operatorname{Bil}(V)$ alternierend und nicht-ausgeartet. Dann existiert eine Basis B von V mit* ${}_B[\beta]_B = \left(\begin{smallmatrix} 0_n & 1_n \\ -1_n & 0_n \end{smallmatrix}\right)$.

Beweis. Induktion nach $\dim V$.[6] Sei $b_1 \in V \setminus \{0\}$. Da β nicht-ausgeartet ist, existiert $c_1 \in V$ mit $\beta(b_1, c_1) \neq 0$. Indem man c_1 durch $\beta(b_1,c_1)^{-1}c_1$ ersetzt, erreicht man $\beta(b_1, c_1) = 1$. Sei $U := \langle b_1, c_1 \rangle$. Für $v = \lambda b_1 + \mu c_1 \in U$ gilt $\beta(v, b_1) = \mu\beta(c_1, b_1) = -\mu$ und $\beta(v, c_1) = \lambda$. Dies zeigt $U \cap U^\perp = 0$. Nach Satz 12.16 gilt $V = U \oplus U^\perp$. Für $u \in U^\perp$ existiert ein $v \in V$ mit $\beta(u, v) \neq 0$. Schreibt man $v = v_1 + v_2$ mit $v_1 \in U$ und $v_2 \in U^\perp$, so folgt $\beta(u, v_2) = \beta(u, v) \neq 0$. Daher ist auch die Einschränkung β' von β auf $U^\perp \times U^\perp$ nicht-ausgeartet und alternierend. Nach Induktion besitzt $U^\perp$ eine Basis $B' = \{b_2, \dots, b_n, c_2, \dots, c_n\}$ mit

$${}_{B'}[\beta']_{B'} = \begin{pmatrix} 0_{n-1} & 1_{n-1} \\ -1_{n-1} & 0_{n-1} \end{pmatrix}.$$

Nun ist $B = \{b_1, \dots, b_n, c_1, \dots, c_n\}$ eine Basis mit der gewünschten Eigenschaft. □

Bemerkung 12.31. Ein Vektorraum V mit einer alternierenden, nicht-ausgearteten Bilinearform nennt man einem *symplektischen Raum*. Darauf gehen wir nicht weiter ein.

12.3 Positiv definite Matrizen

Bemerkung 12.32. In diesem Abschnitt verallgemeinern wir die positive Definitheit von Skalarprodukten in euklidischen Räumen. Stets sei V ein $\mathbb{R}$-Vektorraum.

Definition 12.33. Eine symmetrische Bilinearform $\beta \in \operatorname{Bil}(V)$ nennt man

- *positiv (semi)definit*, falls $\beta(v, v) > 0$ (bzw. $\beta(v, v) \ge 0$) für alle $v \in V \setminus \{0\}$,
- *negativ (semi)definit*, falls $\beta(v, v) < 0$ (bzw. $\beta(v, v) \le 0$) für alle $v \in V \setminus \{0\}$,

[6] Nach Beispiel 12.10 ist $\dim V$ gerade, aber das wird hier nicht benutzt.

- *indefinit*, falls $\exists v, w \in V : \beta(v,v) > 0 > \beta(w,w)$.

Diese Begriffe übertragen sich auf symmetrische Matrizen A, indem man ein β mit $[\beta] = A$ wählt.

Bemerkung 12.34.

(a) Offensichtlich ist jede positiv definite Bilinearform auch positiv semidefinit. Ist β positiv (semi)definit, so ist $-\beta$ negativ (semi)definit. Daher kann man sich in Regel auf die positive Eigenschaft beschränken.

(b) Ist β positiv (semi)definit, so ist die dazu gehörige quadratische Form aus Bemerkung 12.4 positiv (bzw. nicht-negativ).

(c) Die Summe von positiv (semi)definiten Bilinearformen und Matrizen ist wieder positiv (semi)definit. Das Produkt von positiv (semi)definiten Matrizen ist in der Regel nicht symmetrisch und kann nach unserer Definition nicht positiv (semi)definit sein.

Beispiel 12.35.

(a) Sei

$$A = \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Für $x \in \mathbb{R}^n \setminus \{0\}$ gilt

$$xAx^{\mathrm{t}} = 2\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n-1} x_i x_{i+1} = x_1^2 + \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + x_n^2.$$

Im Fall $x_1 \neq 0$ oder $x_n \neq 0$ folgt $xAx^{\mathrm{t}} \geq x_1^2 + x_n^2 > 0$. Anderenfalls existiert ein $i \geq 0$ mit $x_i = 0 \neq x_{i+1}$. Dann ist ebenfalls $xAx^{\mathrm{t}} \geq (x_i - x_{i+1})^2 > 0$. Also ist A positiv definit.

(b) In der Analysis bildet man aus den zweiten partiellen Ableitungen einer Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ die *Hesse-Matrix* $(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}) \in \mathbb{R}^{n\times n}$, welche an einem Punkt $x \in \mathbb{R}^n$ nur dann positiv definit sein kann, wenn x ein lokales Minimum der Funktion ist.

Satz 12.36. *Für jede symmetrische Bilinearform $\beta \in \mathrm{Bil}(V)$ mit Index (r,s,t) gilt*

(a) β positiv semidefinit $\iff$ $s = 0$.

(b) β positiv definit $\iff$ $s = t = 0$.

(c) β indefinit $\iff$ $s, t > 0$.

Beweis. Sei B die Basis aus Sylvesters Trägheitssatz. Ist $s > 0$, so existiert ein $b \in B$ mit $\beta(b,b) = -1$. Dann kann β nicht positiv semidefinit sein. Ist $t > 0$, so existiert $b \in B$ mit $\beta(b,b) = 0$. Dann kann β nicht positiv definit sein. Ist $s = 0$ (bzw. $s = t = 0$) so gilt $\beta(b,b) \geq 0$ (bzw. $\beta(b,b) > 0$) für alle $b \in B$. Da B eine Orthogonalbasis ist, folgt leicht, dass β positiv (semi)definit ist. Genau dann ist β indefinit, falls $b, c \in B$ mit $\beta(b,b) = 1 = -\beta(c,c)$ existieren. Das bedeutet $s, t > 0$. $\square$

Folgerung 12.37. *Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ ist genau dann positiv (semi)definit, wenn alle Eigenwerte von A positiv (bzw. nicht-negativ) sind.*

Beweis. Bekanntlich sind alle Eigenwerte von A reell. Die Behauptung folgt aus Satz 12.27. $\square$

Beispiel 12.38. Für

$$A = (1+\delta_{ij}) = \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}$$

funktioniert der Trick aus Beispiel 12.35 nicht. Wegen $\operatorname{rk}(A - 1_n) = 1$ ist 1 ein Eigenwert von A mit (geometrischer) Vielfachheit $n-1$ (vgl. Aufgabe 26). Der fehlende Eigenwert muss $\operatorname{tr}(A) - (n-1) = 2n - n + 1 = n + 1$ sein. Daher ist A positiv definit.

Bemerkung 12.39. Da sich die Eigenwerte in der Praxis nur schwer berechnen lassen, ist es nützlich andere Kriterien für die positive Definitheit zu kennen. Eine notwendige (aber nicht hinreichende) Bedingung ist, dass alle Diagonaleinträge positiv sind, denn $a_{ii} = e_i A e_i^{\mathrm{t}}$.

Lemma 12.40. *Genau dann ist $A \in \mathbb{R}^{n\times n}$ positiv semidefinit, wenn eine Matrix $S \in \mathbb{R}^{n\times n}$ mit $A = SS^{\mathrm{t}}$ existiert. Ist A positiv definit, so ist S invertierbar.*

Beweis. Sei $A = SS^{\mathrm{t}}$ für ein $S \in \mathbb{R}^{n\times n}$. Für $v \in \mathbb{R}^n$ gilt $vAv^{\mathrm{t}} = vS(vS)^{\mathrm{t}} = |vS|^2 \ge 0$. Daher ist A positiv semidefinit. Ist S invertierbar, so gilt $|vS| > 0$ für alle $v \neq 0$. Dann ist A positiv definit. Sei umgekehrt A positiv semidefinit. Nach Sylvesters Trägheitssatz und Satz 12.36 existiert ein $U \in \mathrm{GL}(n,\mathbb{R})$ mit $A = UDU^{\mathrm{t}}$, wobei $D = \operatorname{diag}(1_r, 0_t)$. Für $S = UD$ gilt $SS^{\mathrm{t}} = A$ wegen $D^2 = D$. Ist A positiv definit, so ist $S = U$ invertierbar. $\square$

Satz 12.41 (Sylvester-Kriterium)**.** *Sei $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ symmetrisch und $A_k := (a_{ij})_{i,j=1}^k$ für $1 \le k \le n$. Genau dann ist A positiv definit, wenn $\det(A_k) > 0$ für $k = 1,\ldots,n$ gilt.*

Beweis. Induktion nach n: Im Fall $n = 1$ ist A genau dann positiv definit, falls $\det(A) = \det(A_1) = a_{11} > 0$. Sei nun $n \ge 2$ und $1 \le k \le n$. Sei A positiv definit und $v \in \mathbb{R}^k \setminus \{0\}$. Für $w := (v_1,\ldots,v_k,0,\ldots,0) \in \mathbb{R}^n$ gilt $vA_kv^{\mathrm{t}} = wAw^{\mathrm{t}} > 0$. Also ist A_k positiv definit. Nach Lemma 12.40 existiert ein $S \in \mathrm{GL}(k,\mathbb{R})$ mit $A_k = SS^{\mathrm{t}}$. Es folgt $\det(A_k) = \det(S)^2 > 0$.

Nehmen wir umgekehrt $\det(A_k) > 0$ für $k = 1,\ldots,n$ an. Nach Induktion ist A_{n-1} positiv definit. Sei $\beta \in \operatorname{Bil}(\mathbb{R}^n)$ mit $[\beta] = A$. Dann ist die Einschränkung β_1 von β auf $\mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ positiv definit, denn $[\beta_1] = A_{n-1}$. Sei $b_1',\ldots,b_{n-1}' \in \mathbb{R}^{n-1}$ eine Orthonormalbasis von β_1. Durch Anfügen einer 0 erhält man die Vektoren $b_i := (b_i', 0) \in \mathbb{R}^n$ mit

$$\beta(b_i, b_j) = b_i A b_j^{\mathrm{t}} = b_i' A_{n-1} b_j' = \delta_{ij}$$

für $i = 1,\ldots,n-1$. Sei $V_1 := \langle b_1,\ldots,b_{n-1}\rangle$. Da β auf V_1 positiv definit ist, gilt $V_1 \cap V_1^{\perp} = \{0\}$. Mit $b_n \in V_1^{\perp} \setminus \{0\}$ ist $B = \{b_1,\ldots,b_n\}$ eine Orthogonalbasis von $\mathbb{R}^n$ und

$$D := {}_B[\beta]_B = \operatorname{diag}(1,\ldots,1,\lambda)$$

für ein $\lambda \in \mathbb{R}$. Da A und D kongruent sind, existiert ein $S \in \mathrm{GL}(n,\mathbb{R})$ mit $D = SAS^{\mathrm{t}}$. Es folgt $\lambda = \det(D) = \det(S)^2 \det(A) > 0$. Mit D ist nun auch A positiv definit. $\square$

Beispiel 12.42.

(a) Die Matrix $A = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ ist genau dann positiv definit, wenn $a > 0$ und $ac > b^2$.

(b) Sei

$$A := \begin{pmatrix} 2 & -1 & . & . & -1 \\ -1 & 2 & -1 & . & . \\ . & -1 & 2 & -1 & . \\ . & . & -1 & 2 & -1 \\ -1 & . & . & -1 & 2 \end{pmatrix}$$

Nach Beispiel 12.35 gilt $\det(A_k) > 0$ für $k = 1, \ldots, 4$. Da die Zeilensummen von A verschwinden, ist $(1, \ldots, 1)^{\mathrm{t}}$ ein Eigenvektor zum Eigenwert 0. Dies zeigt $\det(A) = 0$. Der Beweis von Satz 12.41 zeigt, dass A positiv semidefinit ist, aber nicht positiv definit.

Bemerkung 12.43.

(a) Achtung: Aus $\det(A_k) \geq 0$ für $k = 1, \ldots, n$ folgt nicht, dass A positiv semidefinit ist (betrachte zum Beispiel $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & -1 \end{smallmatrix}\right)$).

(b) Ebenso ist die negative Definitheit nicht zu $\det(A_k) < 0$ äquivalent. Bekanntlich ist A genau dann negativ definit, wenn $-A$ positiv definit ist. Dies ist äquivalent zu $(-1)^k \det(A_k) > 0$ für $k = 1, \ldots, n$.

Satz 12.44. *Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ mit charakteristischem Polynom*

$$\chi_A = X^n + a_1 X^{n-1} + \ldots + a_n \in \mathbb{R}[X]$$

ist genau dann positiv definit, falls die Koeffizienten von χ_A alternierende Vorzeichen haben, d. h. $a_i(-1)^i > 0$ für $i = 1, \ldots, n$.

Beweis. Sei $a_i(-1)^i > 0$ für $i = 1, \ldots, n$. Für $\lambda < 0$ gilt dann

$$\chi_A(\lambda) = (-1)^n \big(|\lambda|^n + |a_1||\lambda|^{n-1} + \ldots + |a_n|\big) \neq 0.$$

Also besitzt A nur positive Eigenwerte. Nach Folgerung 12.37 ist A positiv definit.

Sei umgekehrt A positiv definit mit Eigenwerten $\lambda_1, \ldots, \lambda_n > 0$. Dann gilt

$$\chi_A = \prod_{i=1}^n (X - \lambda_i) = X^n - \Big(\sum_{i=1}^n \lambda_i\Big) X^{n-1} + \Big(\sum_{i<j} \lambda_i \lambda_j\Big) X^{n-2} + \ldots + (-1)^n \lambda_1 \ldots \lambda_n$$

und

$$a_k(-1)^k = \sum_{i_1 < \ldots < i_k} \lambda_{i_1} \ldots \lambda_{i_n} > 0$$

für $k = 1, \ldots, n$. $\square$

Beispiel 12.45. Für

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

gilt

$$\chi_A = (X-2)^2(X-1) + 2 - 2(X-2) - (X-1) = X^3 - 5X^2 + 5X + 3.$$

Da die letzten beiden Koeffizienten das gleiche Vorzeichen haben, kann A nicht positiv definit sein.

Satz 12.46. *Sei $A \in \mathbb{R}^{n\times n}$ positiv (semi)definit und $k \in \mathbb{N}$. Dann besitzt A genau eine positiv (semi)definite k-te Wurzel $W \in \mathbb{R}^{n\times n}$, d. h. $W^k = A$.*

Beweis. Nach dem Hauptachsensatz existiert ein $S \in \mathrm{O}(n,\mathbb{R})$ mit $S^{\mathrm{t}}AS = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$. Da A positiv semidefinit ist, gilt $\lambda_1,\ldots,\lambda_n \ge 0$ nach Folgerung 12.37. Nun ist $W := S\operatorname{diag}(\sqrt[k]{\lambda_1},\ldots,\sqrt[k]{\lambda_n})S^{\mathrm{t}}$ positiv (semi)definit mit $W^k = A$. Sei $\alpha \in \mathbb{R}[X]$ mit $\alpha(\lambda_i) = \sqrt[k]{\lambda_i}$ für $i = 1,\ldots,n$ (Interpolation). Wegen $(S^{\mathrm{t}}AS)^i = S^{\mathrm{t}}A^iS$ für $i \in \mathbb{N}$ gilt

$$\alpha(A) = S\alpha(S^{\mathrm{t}}AS)S^{\mathrm{t}} = W.$$

Sei auch $B \in \mathbb{R}^{n\times n}$ positiv (semi)definit mit $B^k = A$. Der Hauptachsensatz liefert ein $T \in \mathrm{O}(n,\mathbb{R})$ mit $T^{\mathrm{t}}BT = \operatorname{diag}(\mu_1,\ldots,\mu_n)$ und $\mu_1,\ldots,\mu_n \ge 0$. Dabei sind $\mu_1^k,\ldots,\mu_n^k$ die Eigenwerte von $B^k = A$. Also existiert eine Permutationsmatrix P mit $P^{\mathrm{t}}T^{\mathrm{t}}BTP = S^{\mathrm{t}}WS$ und $P^{\mathrm{t}}T^{\mathrm{t}}ATP = S^{\mathrm{t}}AS$. Daher ist TPS^{t} mit A und mit $\alpha(A) = W$ vertauschbar. Dies zeigt $B = TPS^{\mathrm{t}}WSP^{\mathrm{t}}T^{\mathrm{t}} = W$. $\square$

13 Unitäre Räume

13.1 Sesquilinearformen

Bemerkung 13.1. Wir entwickeln in diesem Kapitel eine Geometrie für die komplexen Zahlen. Anstelle des euklidischen Skalarprodukts tritt eine „schiefsymmetrische“ Abbildung, die nur noch in der ersten Koordinate linear ist. Das Gegenstück des Hauptachsensatzes ist der Spektralsatz.

Definition 13.2. Ein *Skalarprodukt* auf einem $\mathbb{C}$-Vektorraum V ist eine Abbildung $V \times V \to \mathbb{C}$ mit folgenden Eigenschaften ($u, v, w \in V$, $\lambda \in \mathbb{C}$):

- $[v, v] \in \mathbb{R}$ und $[v, v] \geq 0$ mit Gleichheit genau dann, wenn $v = 0$ (*positiv definit*),
- $[v, w] = \overline{[w, v]}$ (*schiefsymmetrisch*),
- $[\lambda u + v, w] = \lambda[u, w] + [v, w]$.

Zusammen mit einem Skalarprodukt wird V ein *unitärer Raum*. Vektoren $v, w \in V$ heißen *orthogonal*, falls $[v, w] = 0$. Man nennt $|v| := \sqrt{[v, v]} \geq 0$ die *Norm* von v. Im Fall $|v| = 1$ nennt man v *normiert*.

Beispiel 13.3. Das *Standardskalarprodukt* auf $V = \mathbb{C}^n$ ist definiert durch

$$[v, w] := v\overline{w}^{\mathrm{t}} = \sum_{i=1}^{n} v_i \overline{w_i}$$

für $v, w \in V$. Man prüft leicht, dass die Eigenschaften des Skalarprodukts erfüllt sind.

Bemerkung 13.4.

(a) Für $v, w \in V$ und $\lambda \in \mathbb{C}$ gilt

$$[u, \lambda v + w] = \overline{[\lambda v + w, u]} = \overline{\lambda[v, u] + [w, u]} = \overline{\lambda}[u, v] + [u, w],$$

d. h. das Skalarprodukt ist in der zweiten Komponente nicht linear. Man spricht daher von einer *Sesquilinearform.*[1]

(b) Die meisten der in Abschnitt 11.1 bewiesenen Sätze hängen nicht wesentlich von der Bilinearität das Skalarprodukts ab und können daher problemlos übertragen werden.

Lemma 13.5. *Sei V ein unitärer Raum, $v, w \in V$ und $\lambda \in \mathbb{C}$. Dann gilt*

(a) $|\lambda v| = |\lambda||v|$ (Homogenität).

(b) $|[v, w]| \leq |v||w|$ *mit Gleichheit genau dann, wenn v und w linear abhängig sind* (Cauchy-Schwarz-Ungleichung).

[1] *sesqui* ist lateinisch für eineinhalb.

(c) $||v|-|w|| \le |v+w| \le |v|+|w|$ (Dreiecksungleichung).

Beweis.

(a) $|\lambda v| = \sqrt{[\lambda v, \lambda v]} = \sqrt{\lambda\overline{\lambda}[v,v]} = \sqrt{|\lambda|^2}\sqrt{[v,v]} = |\lambda||v|$.

(b) O. B. d. A. sei $w \neq 0$ und $\lambda := \frac{[v,w]}{[w,w]}$. Wegen $\overline{\lambda} = \frac{[w,v]}{[w,w]}$ gilt

$$0 \le |v-\lambda w|^2 = [v-\lambda w, v-\lambda w] = [v,v] - \lambda[w,v] - \overline{\lambda}[v,w] + |\lambda|^2[w,w] = |v|^2 - \frac{|[v,w]|^2}{|w|^2}.$$

Es folgt $|[v,w]|^2 \le |v|^2|w|^2$ und $|[v,w]| \le |v||w|$. Gleichheit impliziert $v = \lambda w$, d. h. v und w sind linear abhängig. Sind umgekehrt v und w linear abhängig gegeben, dann existiert ein $\mu \in \mathbb{C}$ mit $v = \mu w$ und $|[v,w]| = |\mu||w|^2 \overset{\text{(a)}}{=} |\mu w||w| = |v||w|$.

(c) Sei $[v,w] = a + b\mathrm{i}$. Dann gilt

$$[v,w] + [w,v] = [v,w] + \overline{[v,w]} = 2a \le 2\sqrt{a^2+b^2} = 2|[v,w]|.$$

Es folgt

$$|v+w|^2 = [v+w, v+w] \le [v,v] + 2|[v,w]| + [w,w] \overset{\text{(b)}}{\le} |v|^2 + 2|v||w| + |w|^2 = (|v|+|w|)^2$$

und $|v+w| \le |v|+|w|$. Also ist $|v| = |v+w-w| \le |v+w| + |w|$ und $|v|-|w| \le |v+w|$. Vertauschen von v und w liefert $-(|v|-|w|) = |w|-|v| \le |v+w|$, d. h. $||v|-|w|| \le |v+w|$. □

Bemerkung 13.6. Sei V ein unitärer Raum.

(a) Eine Basis $b_1, \ldots, b_n$ von V heißt *Orthonormalbasis* falls $[b_i, b_j] = \delta_{ij}$ für $1 \le i,j \le n$ gilt. Das Gram-Schmidt-Verfahren zur Berechnung einer Orthonormalbasis funktioniert unverändert für unitäre Räume. Insbesondere besitzt jeder unitäre Raum eine Orthonormalbasis.

(b) Für $U \le V$ definiert man wie üblich $U^\perp := \{v \in V : \forall u \in U : [v,u] = 0\} \le V$. Die Regeln aus Lemma 11.15

- $V = U \oplus U^\perp$,
- $(U^\perp)^\perp = U$,
- $U \subseteq W \iff W^\perp \subseteq U^\perp$,

gelten auch für unitäre Räume.

13.2 Adjungierte Abbildungen

Satz 13.7. *Sei V ein unitärer Raum und $f \in \operatorname{End}(V)$. Dann existiert genau ein $f^* \in \operatorname{End}(V)$ mit $[f(v), w] = [v, f^*(w)]$ für alle $v, w \in V$.*

Beweis. Sei $b_1, \ldots, b_n$ eine Orthonormalbasis von V. Wir definieren $f^* \in \operatorname{End}(V)$ durch

$$f^*(b_j) := \sum_{i=1}^{n} [f(b_i), b_j] b_i$$

für $j = 1, \ldots, n$. Für $v = \sum \lambda_i b_i$ und $w = \sum \mu_j b_j$ gilt

$$[f(v), w] = \sum_{i,j=1}^{n} \lambda_i \overline{\mu_j} [f(b_i), b_j] = \sum_{i,j=1}^{n} \lambda_i \overline{\mu_j} [b_i, f^*(b_j)] = [v, f^*(w)].$$

Sei auch $f_1 \in \operatorname{End}(V)$ mit $[f(v), w] = [v, f_1(w)]$ für alle $v, w \in V$. Dann ist $[v, f^*(w) - f_1(w)] = 0$. Für $v := f^*(w) - f_1(w)$ ergibt sich $f^*(w) = f_1(w)$ für alle $w \in V$. Dies zeigt $f_1 = f^*$. □

Definition 13.8. In der Situation von Satz 13.7 nennt man f^* die zu f *adjungierte* Abbildung.

Bemerkung 13.9.

(a) Für $v, w \in V$ gilt $[v, f(w)] = \overline{[f(w), v]} = \overline{[w, f^*(v)]} = [f^*(v), w]$. Insbesondere ist $(f^*)^* = f$.

(b) Für $f, g \in \operatorname{End}(V)$ und $\lambda \in \mathbb{C}$ gilt $(\lambda f + g)^* = \overline{\lambda} f^* + g^*$.

(c) Achtung: Wir benutzen für die adjungierte Abbildung von f die gleiche Bezeichnung wie für die zu f duale Abbildung. Die duale Abbildung liegt jedoch in $\operatorname{End}(V^*)$.

Definition 13.10. Sei V ein unitärer Raum. Man nennt $f \in \operatorname{End}(V)$

- *hermitesch*, falls $f = f^*$, d. h. $[f(v), w] = [v, f(w)]$ für alle $v, w \in V$.
- *unitär*, falls $[f(v), f(w)] = [v, w]$ für alle $v, w \in V$.
- *normal*, falls $f \circ f^* = f^* \circ f$.

Bemerkung 13.11.

(a) Sei $f \in \operatorname{End}(V)$ hermitesch. Sei $\lambda \in \mathbb{C}$ ein Eigenwert von f mit Eigenvektor $v \in V$. Dann gilt

$$\lambda |v|^2 = [\lambda v, v] = [f(v), v] = [v, f(v)] = [v, \lambda v] = \overline{\lambda} |v|^2$$

und $\lambda = \overline{\lambda} \in \mathbb{R}$.

(b) Sei $f \in \operatorname{End}(V)$ unitär. Für $v \in \operatorname{Ker}(f)$ ist $|v|^2 = [v, v] = [f(v), f(v)] = 0$, d. h. $v = 0$. Daher ist f ein Isomorphismus. Wegen $[f(v), w] = [v, f^{-1}(w)]$ für alle $v, w \in V$ folgt außerdem $f^* = f^{-1}$. Sind $f, g \in \operatorname{End}(V)$ unitär, so auch $f \circ g$ und f^{-1}. Daher bilden die unitären Abbildungen eine Untergruppe $\mathrm{U}(V)$ von $\mathrm{GL}(V)$. Man nennt $\mathrm{U}(V)$ die *unitäre Gruppe* vom Grad n.

(c) Sei λ ein Eigenwert einer unitären Abbildung f mit Eigenvektor $v \in V$. Dann gilt $|\lambda|^2 |v|^2 = [\lambda v, \lambda v] = [f(v), f(v)] = [v, v] = |v|^2$ und $|\lambda| = 1$.

(d) Hermitesche und unitäre Abbildungen sind offenbar normal.

Lemma 13.12. *Sei V unitär mit Orthonormalbasis B und $f \in \operatorname{End}(V)$. Dann gilt ${}_B[f^*]_B = \overline{{}_B[f]_B}^{\mathrm{t}}$.*

Beweis. Sei $B = \{b_1, \dots, b_n\}$. Sei $f(b_i) = \sum_{j=1}^n a_{ji}b_j$ und $f^*(b_i) = \sum_{j=1}^n a^*_{ji}b_j$ für $i = 1, \dots, n$. Die Behauptung folgt aus
$$\overline{a^*_{ji}} = [b_j, f^*(b_i)] = [f(b_j), b_i] = a_{ij}.$$ □

Definition 13.13. Für $A \in \mathbb{C}^{n\times n}$ sei $A^* := \overline{A}^{\mathrm{t}} = \overline{A^{\mathrm{t}}}$ die zu A *adjungierte* Matrix.

Folgerung 13.14. *Sei V unitär mit Orthonormalbasis B, $f \in \mathrm{End}(V)$ und $A := {}_B[f]_B$. Dann gilt*

(a) f hermitesch $\iff A^* = A \iff A^{\mathrm{t}} = \overline{A}$.

(b) f unitär $\iff A^* = A^{-1} \iff A^{\mathrm{t}} = \overline{A}^{-1}$.

(c) f normal $\iff A^*A = AA^* \iff A^{\mathrm{t}}\overline{A} = \overline{A}A^{\mathrm{t}}$.

Beweis. Es gilt
$$f \text{ hermitesch} \iff f^* = f \overset{13.12}{\iff} A^* = A \iff \overline{A^*} = \overline{A} \iff A^{\mathrm{t}} = \overline{A}.$$
Die anderen Aussagen beweist man analog. □

Definition 13.15. Eine Matrix $A \in \mathbb{C}^{n\times n}$ heißt

- *hermitesch*, falls $A^* = A$.
- *unitär*, falls $A^* = A^{-1}$.
- *normal*, falls $AA^* = A^*A$.

Beispiel 13.16.

(a) Hermitesche und unitäre Matrizen sind normal. Diagonalmatrizen sind normal, aber nicht unbedingt hermitesch oder unitär.

(b) Reelle Matrizen sind genau dann hermitesch (bzw. unitär), wenn sie symmetrisch (bzw. orthogonal) sind.

(c) Invertieren, Transponieren und Komplex-Konjugieren sind miteinander vertauschbare Operatoren auf $\mathbb{C}^{n\times n}$ (vgl. Bemerkung 5.9). Ist A hermitesch (bzw. unitär, normal), so auch $\overline{A}$, A^{t} und A^* (nachrechnen).

Bemerkung 13.17. Die unitären Matrizen bilden eine Untergruppe $\mathrm{U}(n, \mathbb{C})$ von $\mathrm{GL}(n, \mathbb{C})$, die wie üblich $\mathrm{U}(V)$ entspricht. Für $A \in \mathrm{U}(n, \mathbb{C})$ gilt $|\det(A)|^2 = \det(A)\overline{\det(A)} = \det(A^{\mathrm{t}}\overline{A}) = 1$. Nach Bemerkung 13.11 haben auch alle Eigenwerte von A den Betrag 1. Man nennt
$$\mathrm{SU}(n, \mathbb{C}) := \mathrm{U}(n, \mathbb{C}) \cap \mathrm{SL}(n, \mathbb{C}) \leq \mathrm{U}(n, \mathbb{C}).$$
die *spezielle unitäre Gruppe* vom Grad n.

13.3 Der Spektralsatz

Satz 13.18 (Spektralsatz[2])**.** *Sei V unitär und $f \in \mathrm{End}(V)$. Genau dann ist f normal, wenn V eine Orthonormalbasis aus Eigenvektoren von f besitzt. Insbesondere sind normale Endomorphismen diagonalisierbar.*

Beweis. Sei B eine Orthonormalbasis aus Eigenvektoren von f. Dann ist $A := {}_B[f]_B$ eine Diagonalmatrix. Sicher ist A^* auch eine Diagonalmatrix. Insbesondere sind A und A^* vertauschbar. Nach Folgerung 13.14 ist f normal. Sei umgekehrt f normal. Wir argumentieren durch Induktion nach $n := \dim V$. Im Fall $n = 1$ liefert jeder normierte Vektor eine Orthonormalbasis aus Eigenvektoren. Sei $n \geq 2$. Nach dem Fundamentalsatz der Algebra besitzt f einen Eigenwert $\lambda \in \mathbb{C}$ mit Eigenvektor $b_1 \in V$. O. B. d. A. sei $|b_1| = 1$. Wegen

$$\begin{aligned}
|f^*(b_1) - \overline{\lambda} b_1|^2 &= [f^*(b_1), f^*(b_1)] - \lambda[f^*(b_1), b_1] - \overline{\lambda}[b_1, f^*(b_1)] + |\lambda|^2[b_1, b_1] \\
&= [f(f^*(b_1)), b_1] - \lambda[b_1, f(b_1)] - \overline{\lambda}[f(b_1), b_1] + |\lambda|^2 = [f^*(f(b_1)), b_1] - |\lambda|^2 \\
&= \lambda[f^*(b_1), b_1] - |\lambda|^2 = \lambda[b_1, f(b_1)] - |\lambda|^2 = 0
\end{aligned}$$

ist $\overline{\lambda}$ ein Eigenwert von f^* zum Eigenvektor b_1. Für $U := \langle b_1 \rangle$ gilt $V = U \oplus U^\perp$. Für $u \in U^\perp$ ist $[f(u), b_1] = [u, f^*(b_1)] = \lambda[u, b_1] = 0$ und $f(u) \in U^\perp$. Daher ist die Einschränkung von f auf $U^\perp$ ein normaler Endomorphismus. Nach Induktion existiert eine Orthonormalbasis $b_2, \ldots, b_n$ von $U^\perp$ aus Eigenvektoren von f. Nun ist $b_1, \ldots, b_n$ eine Orthonormalbasis von V aus Eigenvektoren von f. $\square$

Bemerkung 13.19. Ist $A \in \mathbb{R}^{n\times n}$ symmetrisch, so liefert der Spektralsatz ein $S \in \mathrm{U}(n, \mathbb{C})$ mit $S^*AS = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Im Gegensatz zum Hauptachsensatz erhält man nicht, dass S orthogonal (also reell) ist.

Folgerung 13.20. *Sei V unitär und $f \in \mathrm{End}(V)$ normal mit Eigenwerten $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. Dann gilt*

(a) f hermitesch $\iff \lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

(b) f unitär $\iff |\lambda_1| = \ldots = |\lambda_n| = 1$.

Beweis. Nach dem Spektralsatz existiert eine Orthonormalbasis B von V mit $D := {}_B[f]_B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Nach Folgerung 13.14 ist f genau dann hermitesch (bzw. unitär), wenn $\overline{D} = D^{\mathrm{t}} = D$ (bzw. $1_n = D\overline{D} = \mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2)$) gilt. Dies zeigt die Behauptung. $\square$

Bemerkung 13.21. Sei $f \in \mathrm{End}\, V$ normal. Seien $v, w \in V$ Eigenvektoren zu verschiedenen Eigenwerten λ bzw. μ. Wie im Beweis des Spektralsatz ist w ein Eigenvektor von f^* zum Eigenwert $\overline{\mu}$. Aus

$$\lambda[v, w] = [f(v), w] = [v, f^*(w)] = [v, \overline{\mu} w] = \mu[v, w]$$

folgt $[v, w] = 0$. Daher sind Eigenvektoren zu verschiedenen Eigenwerten orthogonal (wie bei symmetrischen Matrizen, siehe Bemerkung 11.40). Die Orthonormalbasis im Spektralsatz lässt sich also wie beim Hauptachsensatz mit dem Gram-Schmidt-Verfahren bestimmen.

[2] Die Menge der Eigenwerte von $f \in \mathrm{End}(V)$ nennt man das *Spektrum* von f.

Beispiel 13.22. Nehmen wir an, dass

$$A := \begin{pmatrix} 5\mathrm{i} & -4 & 2 \\ 4 & 5\mathrm{i} & 2\mathrm{i} \\ -2 & 2\mathrm{i} & 8\mathrm{i} \end{pmatrix}$$

normal ist (anderenfalls wird sich ein Widerspruch ergeben). Man berechnet

$$\chi_A = (X-5\mathrm{i})^2(X-8\mathrm{i}) - 32\mathrm{i} + 8(X-5\mathrm{i}) + 16(X-8\mathrm{i}) = X^3 - 18\mathrm{i}X^2 - 81X = X(X-9\mathrm{i})^2.$$

Wegen

$$A \sim \begin{pmatrix} 5\mathrm{i} & -4 & 2 \\ 0 & 9\mathrm{i}/5 & 18\mathrm{i}/5 \\ 0 & 18\mathrm{i}/5 & 36\mathrm{i}/5 \end{pmatrix} \sim \begin{pmatrix} 5\mathrm{i} & -4 & 2 \\ 0 & 9\mathrm{i}/5 & 18\mathrm{i}/5 \\ 0 & 0 & 0 \end{pmatrix}$$

ist $(2\mathrm{i}, -2, 1)$ ein Eigenvektor zum Eigenwert 0. Nach Normierung sei $b_1 := \frac{1}{3}(2\mathrm{i}, -2, 1)$. Der Eigenraum zum Eigenwert 9i wird durch $c_2 := (0, 1, 2)$ und $c_3 := (\mathrm{i}, 1, 0)$ aufgespannt. Wir stellen fest, dass diese Vektoren in der Tat orthogonal zu b_1 sind. Daher muss A normal sein. Das Gram-Schmidt-Verfahren liefert:

$$\begin{aligned} b_2 &:= \frac{1}{\sqrt{5}}(0,1,2), \\ c_3' &:= c_2 - [c_2, b_2]b_2 = (\mathrm{i}, 4/5, -2/5), \\ b_3 &:= \frac{1}{3\sqrt{5}}(5\mathrm{i}, 4, -2). \end{aligned}$$

Nun ist $B = \{b_1, b_2, b_3\}$ eine Orthonormalbasis aus Eigenvektoren von A.

Satz 13.23 (Schur-Zerlegung)**.** *Für $A \in \mathbb{C}^{n\times n}$ existiert ein $S \in \mathrm{U}(n, \mathbb{C})$, sodass S^*AS eine Dreiecksmatrix ist. Genau dann ist A normal, wenn S^*AS eine Diagonalmatrix ist.*

Beweis. O. B. d. A. sei $n \geq 2$. Sei $\lambda \in \mathbb{C}$ ein Eigenwert von A und $v \in \mathbb{C}^n$ ein normierter Eigenvektor zu λ. Wir ergänzen v mit Gram-Schmidt zu einer Orthonormalbasis von $\mathbb{C}^n$. Nach Wechsel zu dieser Basis gilt $A = \left(\begin{smallmatrix} \lambda & * \\ 0 & A_1 \end{smallmatrix}\right)$ mit $A_1 \in \mathbb{C}^{(n-1)\times(n-1)}$. Induktiv existiert ein $S_1 \in \mathrm{U}(n-1, \mathbb{C})$, sodass $S_1^* A_1 S_1$ eine obere Dreiecksmatrix ist. Jetzt ist $S := \mathrm{diag}(1_1, S_1) \in \mathrm{U}(n, \mathbb{C})$ und S^*AS ist eine obere Dreiecksmatrix. Eine untere Dreiecksmatrix erhält man, indem man A^{t} in eine obere Dreiecksmatrix transformiert und anschließend transponiert.

Ist A normal, so ist $D := (d_{ij}) = S^*AS$ eine normale obere Dreiecksmatrix. Wegen

$$|d_{ii}|^2 = (DD^*)_{ii} = (D^*D)_{ii} = |d_{1i}|^2 + \ldots + |d_{ii}|^2$$

ist $a_{ji} = 0$ für $j = 1, \ldots, i-1$ und $i = 1, \ldots, n$. Dies zeigt, dass D eine Diagonalmatrix ist. Ist umgekehrt D eine Diagonalmatrix, so ist A nach dem Spektralsatz normal. □

Beispiel 13.24. Die Matrix

$$A := \begin{pmatrix} -1 & -1 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}.$$

hat offenbar den Eigenvektor e_3 zum Eigenwert 1. Übergang zur Orthonormalbasis $\{e_3, e_2, e_1\}$ ergibt

$$A \approx \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & -1 \end{pmatrix}.$$

Die Matrix $A_1 = \left(\begin{smallmatrix} 1 & 2 \\ -1 & -1 \end{smallmatrix}\right)$ im Beweis von Satz 13.23 hat den Eigenvektor $(1+\mathrm{i}, -1)$ zum Eigenwert i. Orthogonal dazu steht $(1, 1-\mathrm{i})$. Nach Normierung erhält man

$$A \approx \frac{1}{3}\begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1-\mathrm{i} & -1 \\ 0 & 1 & 1+\mathrm{i} \end{pmatrix}\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & -1 \end{pmatrix}\begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1+\mathrm{i} & 1 \\ 0 & -1 & 1-\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 & (2+2\mathrm{i})/\sqrt{3} & 2/\sqrt{3} \\ 0 & \mathrm{i} & 1-2\mathrm{i} \\ 0 & 0 & -\mathrm{i} \end{pmatrix}.$$

Satz 13.25 (Trägheitssatz für hermitesche Matrizen)**.** *Für jede hermitesche Matrix $A \in \mathbb{C}^{n\times n}$ existiert ein $S \in \mathrm{GL}(n,\mathbb{C})$ mit*

$$S^*AS = \mathrm{diag}(1_r, -1_s, 0_t)$$

Die Zahlen r, s, t sind eindeutig bestimmt.

Beweis. Nach dem Spektralsatz können wir $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ annehmen. Nach Folgerung 13.20 sind die Eigenwerte $\lambda_1, \ldots, \lambda_n$ von A reell. Sei o. B. d. A. $\lambda_1, \ldots, \lambda_r > 0$, $\lambda_{r+1}, \ldots, \lambda_{r+s} < 0$ und $\lambda_{r+s+1} = \ldots = \lambda_n = 0$. Die Gleichung gilt nun mit

$$S := \mathrm{diag}(\sqrt{|\lambda_1|}^{-1}, \ldots, \sqrt{|\lambda_{r+s}|}^{-1}, 1_t) \in \mathrm{GL}(n,\mathbb{C}).$$

Mit $\mathrm{rk}(A) = r+s$ ist t eindeutig bestimmt. Für alle $v \in \langle e_1, \ldots, e_r\rangle \setminus \{0\}$ gilt $\overline{v}Av^{\mathrm{t}} > 0$. Sei umgekehrt $U \le \mathbb{C}^n$ mit $\overline{u}Au^{\mathrm{t}} > 0$ für alle $u \in U \setminus \{0\}$. Für $u \in U \cap \langle e_{r+1}, \ldots, e_n\rangle$ gilt einerseits $\overline{u}Au^{\mathrm{t}} \ge 0$ und andererseits $\overline{u}Au^{\mathrm{t}} \le 0$. Dies zeigt $u = 0$ und $U \cap \langle e_{r+1}, \ldots, e_n\rangle = \{0\}$. Es folgt $\dim U \le r$. Somit ist r die maximale Dimension eines Unterraums $U \le \mathbb{C}^n$ mit $\overline{u}Au^{\mathrm{t}} > 0$ für alle $u \in U \setminus \{0\}$. Auf diese Weise ist r eindeutig durch A bestimmt. Nun ist auch $s = n - r - t$ eindeutig bestimmt. □

Beispiel 13.26. Die Zahlen r, s, t in Satz 13.25 kann man mit dem modifizierten Gauß-Verfahren aus Bemerkung 12.21 bestimmen (beachte: $1 + 1 \ne 0$ in $\mathbb{C}$). Zum Beispiel:

$$\begin{aligned}
(A|1_3) &= \left(\begin{array}{ccc|ccc} 1 & \mathrm{i} & 2 & 1 & 0 & 0 \\ -\mathrm{i} & 1 & -1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & \mathrm{i} & 2 & 1 & 0 & 0 \\ 0 & 0 & 2\mathrm{i}-1 & \mathrm{i} & 1 & 0 \\ 0 & -2\mathrm{i}-1 & -1 & -2 & 0 & 1 \end{array}\right) \\
&\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & -2\mathrm{i}-1 & -2 & 0 & 1 \\ 0 & 2\mathrm{i}-1 & 0 & \mathrm{i} & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -2 & 0 & 1 \\ 0 & 0 & 5 & 2-3\mathrm{i} & 1 & 2\mathrm{i}-1 \end{array}\right) \\
&\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & (2-3\mathrm{i})/\sqrt{5} & 1/\sqrt{5} & (2\mathrm{i}-1)/\sqrt{5} \\ 0 & 0 & -1 & -2 & 0 & 1 \end{array}\right) \\
S^* &= \frac{1}{\sqrt{5}}\begin{pmatrix} \sqrt{5} & 0 & 0 \\ 2-3\mathrm{i} & 1 & 2\mathrm{i}-1 \\ -2\sqrt{5} & 0 & \sqrt{5} \end{pmatrix}
\end{aligned}$$

Bemerkung 13.27. Der Satz von Mirsky gibt eine notwendige und hinreichende Bedingung für die Existenz von Matrizen mit vorgegebenen Eigenwerten und Hauptdiagonalelementen (deren Summe muss gleich sein). Für hermitesche Matrizen gelten stärkere Einschränkungen an diese Zahlen.

Satz 13.28 (Schur-Horn).

(a) *Ist $A \in \mathbb{C}^{n\times n}$ hermitesch mit Hauptdiagonale $d_1 \ge \ldots \ge d_n$ und Eigenwerten $\lambda_1 \ge \ldots \ge \lambda_n$, so gilt*
$$\sum_{i=1}^k d_i \le \sum_{i=1}^k \lambda_i$$
für $k = 1, \ldots, n$ mit Gleichheit im Fall $k = n$.

(b) *Sind reelle Zahlen $d_1 \ge \ldots \ge d_n$ und $\lambda_1 \ge \ldots \ge \lambda_n$ mit $\sum_{i=1}^k d_i \le \sum_{i=1}^k \lambda_i$ für $k = 1, \ldots, n$ und $\sum_{i=1}^n d_i = \sum_{i=1}^n \lambda_i$ gegeben, so existiert eine reelle symmetrische Matrix mit Hauptdiagonale $d_1, \ldots, d_n$ und Eigenwerten $\lambda_1, \ldots, \lambda_n$.*

Beweis (Chan-Li).

(a) Nach dem Spektralsatz existiert eine unitäre Matrix $S := (s_{ij})$ mit
$$A = (a_{ij}) = S^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n) S.$$
Für $1 \le k \le n$ gilt
$$\sum_{i=1}^k d_i = \sum_{i=1}^k a_{ii} = \sum_{i=1}^k \sum_{j=1}^n \overline{s_{ji}} s_{ji} \lambda_j = \sum_{j=1}^n \lambda_j \sum_{i=1}^k |s_{ji}|^2. \tag{13.1}$$
Da S unitär ist, gilt $t_j := \sum_{i=1}^k |s_{ji}|^2 \le 1$ mit Gleichheit falls $k = n$. Es folgt
$$\begin{aligned}\sum_{j=1}^n t_j &= \sum_{i=1}^k \sum_{j=1}^n |s_{ji}|^2 = k,\\ \sum_{i=1}^k (d_i - \lambda_i) &= \sum_{j=1}^n \lambda_j t_j - \sum_{j=1}^k \lambda_j + \lambda_k \overbrace{\Bigl(k - \sum_{i=1}^n t_i\Bigr)}^{=0}\\ &= \sum_{j=1}^k \underbrace{(\lambda_j - \lambda_k)}_{\ge 0} \underbrace{(t_j - 1)}_{\le 0} + \sum_{i=k+1}^n t_i \underbrace{(\lambda_i - \lambda_k)}_{\le 0} \le 0.\end{aligned}$$
Also ist $\sum_{i=1}^k d_i \le \sum_{i=1}^k \lambda_i$ und $\sum_{i=1}^n d_i = \operatorname{tr}(A) = \sum_{i=1}^n \lambda_i$.

(b) Induktion nach n: Im Fall $n = 1$ erfüllt $A := (d_1) = (\lambda_1)$ die Behauptung. Sei $n = 2$. Dann gilt $\lambda_1 \ge d_1 \ge d_2 = \lambda_1 + \lambda_2 - d_1 \ge \lambda_2$. Im Fall $\lambda_1 = \lambda_2$ können wir $A := d_1 1_2$ wählen. Im Fall $\lambda_1 > \lambda_2$ ist
$$S := \frac{1}{\sqrt{\lambda_1 - \lambda_2}} \begin{pmatrix} \sqrt{d_1 - \lambda_2} & -\sqrt{\lambda_1 - d_1} \\ \sqrt{\lambda_1 - d_1} & \sqrt{d_1 - \lambda_2} \end{pmatrix} \in \mathrm{O}(2, \mathbb{R}).$$
Für $A = (a_{ij}) = S^{\mathrm{t}} \operatorname{diag}(\lambda_1, \lambda_2) S$ gilt
$$a_{11} = \frac{1}{\lambda_1 - \lambda_2}\bigl((d_1 - \lambda_2)\lambda_1 + (\lambda_1 - d_1)\lambda_2\bigr) = d_1$$
nach (13.1). Es folgt $a_{22} = \lambda_1 + \lambda_2 - d_1 = d_2$. Sei nun $n \ge 3$ und
$$\lambda_2' := \lambda_1 + \lambda_2 - d_1 \ge \lambda_2.$$
Nach Induktion existiert $S \in \mathrm{O}(2, \mathbb{R})$, sodass $S^{\mathrm{t}} \operatorname{diag}(\lambda_1, \lambda_2) S$ Hauptdiagonale (d_1, λ_2') hat. Da die Folgen $(\lambda_2', \lambda_3, \ldots, \lambda_n)$ und $(d_2, \ldots, d_n)$ die gleichen Voraussetzungen erfüllen, existiert nach

Induktion ein $T \in \mathrm{O}(n-1,\mathbb{R})$, sodass $T^{\mathrm{t}} \operatorname{diag}(\lambda_2', \ldots, \lambda_n) T$ Hauptdiagonale $d_2, \ldots, d_n$ hat. Für $U := \operatorname{diag}(S, 1_{n-2}) \operatorname{diag}(1_1, T) \in \mathrm{O}(n,\mathbb{R})$ hat

$$A := U^{\mathrm{t}} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U = \begin{pmatrix} 1 & 0 \\ 0 & T^{\mathrm{t}} \end{pmatrix} \begin{pmatrix} d_1 & * & & & 0 \\ * & \lambda_2' & & & \\ & & \lambda_3 & & \\ & & & \ddots & \\ 0 & & & & \lambda_n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & T \end{pmatrix}$$

Hauptdiagonale $(d_1, \ldots, d_n)$ und Eigenwerte $\lambda_1, \ldots, \lambda_n$. $\square$

Beispiel 13.29. Wir konstruieren eine symmetrische Matrix mit Eigenwerten $3, 2, 1$ und Hauptdiagonale $2, 2, 2$. Mit der Bezeichnung aus dem obigen Beweis ist $\lambda_2' = 3$. Man erhält

$$\begin{aligned} A &= \begin{pmatrix} 1 & . & . \\ . & 1/\sqrt{2} & 1/\sqrt{2} \\ . & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} . & 1 & . \\ -1 & . & . \\ . & . & 1 \end{pmatrix} \begin{pmatrix} 3 & . & . \\ . & 2 & . \\ . & . & 1 \end{pmatrix} \begin{pmatrix} . & -1 & . \\ 1 & . & . \\ . & . & 1 \end{pmatrix} \begin{pmatrix} 1 & . & . \\ . & 1/\sqrt{2} & -1/\sqrt{2} \\ . & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \sqrt{2} & . & . \\ . & 1 & 1 \\ . & -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & . & . \\ . & 3 & . \\ . & . & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & . & . \\ . & 1 & -1 \\ . & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & . & . \\ . & 2 & -1 \\ . & -1 & 2 \end{pmatrix}. \end{aligned}$$

14 Die Jordan-Normalform

14.1 Haupträume

Bemerkung 14.1. Nach Satz 10.34 gibt es zwei Gründe, warum ein Endomorphismus nicht diagonalisierbar ist:

- Das charakteristische Polynom zerfällt nicht in Linearfaktoren.
- Die Eigenräume sind zu „klein“, d. h. die geometrische Vielfachheit eines Eigenwerts ist kleiner als die entsprechende algebraische Vielfachheit.

Über $\mathbb{C}$ tritt das erste Problem nach dem Fundamentalsatz der Algebra nicht auf. Um das zweite Problem zu umgehen, ersetzen wir Eigenräume durch größere Unterräume. Im Folgenden sei V ein endlich-dimensionaler K-Vektorraum.

Definition 14.2. Sei $f \in \mathrm{End}(V)$. Ein Unterraum $U \leq V$ heißt *f-invariant*, falls $f(U) \subseteq U$ gilt.

Beispiel 14.3. Für $f \in \mathrm{End}(V)$ sind $\mathrm{Ker}(f)$ und $f(V)$ stets f-invariante Unterräume, denn $f(\mathrm{Ker}(f)) = \{0\} \subseteq \mathrm{Ker}(f)$ und $f(f(V)) \subseteq f(V)$.

Bemerkung 14.4. Sei $U \leq V$ ein f-invarianter Unterraum.

(a) Für $v, w \in V$ gilt

$$v + U = w + U \implies v - w \in U \implies f(v) - f(w) = f(v - w) \in U \implies f(v) + U = f(w) + U.$$

Daher ist $\overline{f}\colon V/U \to V/U$, $v + U \mapsto f(v) + U$ ein wohldefinierter Endomorphismus.

(b) Sei B_1 eine Basis von U und $B = B_1 \dot\cup B_2$ eine Basis von V. Aus Dimensionsgründen ist dann $\{b + U : b \in B_2\}$ eine Basis von V/U. Außerdem ist ${}_B[f]_B = \left(\begin{smallmatrix} A & C \\ 0 & D \end{smallmatrix}\right)$, wobei $A = {}_{B_1}[f_{|U}]_{B_1}$ und $D = {}_{B_2}[\overline{f}]_{B_2}$. Besitzt U ein f-invariantes Komplement W (d. h. $V = U \oplus W$), so kann man $C = 0$ durch geeignete Basiswahl erreichen. Wir versuchen daher V in möglichst kleine f-invariante Unterräume zu zerlegen.

Definition 14.5. Man nennt $f \in \mathrm{End}(V)$ *trigonalisierbar*, falls eine Basis B von V existiert, sodass ${}_B[f]_B$ eine Dreiecksmatrix ist. Analog heißt $A \in K^{n \times n}$ *trigonalisierbar*, falls A zu einer Dreiecksmatrix ähnlich ist.

Bemerkung 14.6. Sei $B = \{b_1, \ldots, b_n\}$ eine Basis von V, sodass ${}_B[f]_B$ eine obere Dreiecksmatrix ist. Dann ist ${}_{B'}[f]_{B'}$ mit $B' = \{b_n, b_{n-1}, \ldots, b_1\}$ eine untere Dreiecksmatrix. Man muss in Definition 14.5 also nicht spezifizieren, ob man obere oder untere Dreiecksmatrizen betrachtet (vgl. Bemerkung 15.26).

Beispiel 14.7. Nach der Schur-Zerlegung ist jede komplexe quadratische Matrix trigonalisierbar. Der nächste Satz erweitert Satz 10.52.

Satz 14.8. *Eine Abbildung $f \in \mathrm{End}(V)$ ist genau dann trigonalisierbar, wenn μ_f (oder χ_f) in Linearfaktoren zerfällt.*

Beweis. Ist ${}_B[f]_B$ (für eine Basis B von V) eine Dreiecksmatrix mit Diagonale $\lambda_1, \ldots, \lambda_n$, so zerfällt $\chi_f = (X - \lambda_1) \ldots (X - \lambda_n)$ in Linearfaktoren. Nach Cayley-Hamilton und Lemma 10.24 zerfällt auch μ_f in Linearfaktoren.

Nehmen wir umgekehrt an, dass μ_f in Linearfaktoren zerfällt. Nach Satz 10.54 zerfällt auch χ_f in Linearfaktoren. O. B. d. A. sei $V \neq \{0\}$. Dann existiert ein Eigenvektor $v \in V$ von f. Offenbar ist $U := \langle v \rangle$ f-invariant. Sei $\overline{V} := V/U$ und $\overline{f}$ wie in Bemerkung 14.4. Dann ist $\mu_{\overline{f}} \mid \mu_f$. Nach Lemma 10.24 zerfällt auch $\mu_{\overline{f}}$ in Linearfaktoren. Durch Induktion nach $\dim V$ enthalten wir eine Basis $\overline{B} = \{\overline{b_1}, \ldots, \overline{b_{n-1}}\}$ von $\overline{V}$, sodass ${}_{\overline{B}}[\overline{f}]_{\overline{B}}$ eine untere Dreiecksmatrix ist. Wir wählen $b_i \in \overline{b_i}$ für $i = 1, \ldots, n-1$. Dann gilt $f(b_i) \in \langle b_i, \ldots, b_{n-1}, v \rangle$ für $i = 1, \ldots, n-1$. Daher ist $B := \{b_1, \ldots, b_{n-1}, v\}$ eine Basis von V, sodass ${}_B[f]_B$ eine untere Dreiecksmatrix ist. □

Lemma 14.9. *Sei $f \in \mathrm{End}(V)$.*

(a) Ist $U \leq V$ f-invariant, so ist $U^0 \leq V^*$ f^*-invariant.

(b) Ist $U \leq V^*$ f^*-invariant, so ist $U_0 \leq V$ f-invariant.

Beweis. Die dualen Komplemente U^0, U_0 und die duale Abbildung f^* wurden in Abschnitt 7.3 definiert.

(a) Sei $\gamma \in U^0$. Für alle $u \in U$ gilt $f^*(\gamma)(u) = \gamma(f(u)) = 0$, d. h. $f^*(\gamma) \in U^0$.

(b) Sei $v \in U_0$. Für alle $\gamma \in U$ gilt $\gamma(f(v)) = f^*(\gamma)(v) = 0$, d. h. $f(v) \in U_0$. □

Lemma 14.10. *Sei $f \in \mathrm{End}(V)$ diagonalisierbar und $U \leq V$ ein f-invarianter Unterraum. Dann ist auch die Einschränkung $f_{|U}$ diagonalisierbar.*

Beweis. Für das Minimalpolynom μ_1 von $f_{|U}$ gilt $\mu_1 \mid \mu_f$ nach Lemma 10.44. Nach Satz 10.52 zerfällt μ_f in paarweise verschiedene Linearfaktoren. Nach Lemma 10.24 gilt dies auch für μ_1. □

Lemma 14.11 (Simultane Diagonalisierung)**.** *Seien $f, g \in \mathrm{End}(V)$ diagonalisierbar. Genau dann gilt $f \circ g = g \circ f$, wenn f und g* simultan *diagonalisierbar sind, d. h. es existiert eine Basis B von V, sodass ${}_B[f]_B$ und ${}_B[g]_B$ Diagonalmatrizen sind.*

Beweis. Sei B eine Basis von V, sodass $D_f := {}_B[f]_B$ und $D_g := {}_B[g]_B$ Diagonalmatrizen sind. Aus $D_f D_g = D_g D_f$ folgt dann $f \circ g = g \circ f$. Sei umgekehrt $f \circ g = g \circ f$. Für die Eigenwerte $\lambda_1, \ldots, \lambda_k \in K$ von f gilt $V = E_{\lambda_1}(f) \oplus \ldots \oplus E_{\lambda_k}(f)$, da f diagonalisierbar ist. Für $v \in U := E_{\lambda_i}(f)$ gilt

$$f(g(v)) = g(f(v)) = \lambda_i g(v)$$

und $g(v) \in U$. Daher ist U ein g-invarianter Unterraum. Nach Lemma 14.10 ist $g_{|U}$ diagonalisierbar. Man kann also eine Basis von V wählen, die aus gemeinsamen Eigenvektoren von f und g besteht. □

Lemma 14.12 (Simultane Trigonalisierung)**.** *Seien $f, g \in \mathrm{End}(V)$ vertauschbar und trigonalisierbar. Dann sind f und g* simultan *trigonalisierbar, d. h. es existiert eine Basis B von V, sodass ${}_B[f]_B$ und ${}_B[g]_B$ untere Dreiecksmatrizen sind.*

Beweis. O. B. d. A. sei $V \neq \{0\}$. Nach Satz 14.8 besitzt f einen Eigenwert $\lambda \in K$. Wie im Beweis von Lemma 14.11 ist $U := E_\lambda(f)$ ein g-invarianter Unterraum. Das Minimalpolynom von $g_{|U}$ teilt μ_g und zerfällt daher in Linearfaktoren. Daher existiert ein Eigenvektor $u \in U$ von g. Nun ist $W := \langle u \rangle$ sowohl f-invariant und g-invariant. Sei $\overline{V} := V/W$, $\overline{f} \in \operatorname{End}(\overline{V})$ und $\overline{g} \in \operatorname{End}(\overline{V})$ wie im Beweis von Satz 14.8. Wegen $\mu_{\overline{f}} \mid \mu_f$ und $\mu_{\overline{g}} \mid \mu_g$ kann man induktiv annehmen, dass $\overline{f}$ und $\overline{g}$ simultan trigonalisierbar sind. Die Behauptung folgt nun wie im Beweis von Satz 14.8. $\square$

Beispiel 14.13. Im Gegensatz zur simultanen Diagonalisierung ist die Umkehrung von Lemma 14.12 falsch: $\operatorname{diag}(1, 0)$ und $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ sind trivialerweise trigonalisierbar, aber nicht vertauschbar.

Definition 14.14. Sei $\dim V = n$ und $f \in \operatorname{End}(V)$ mit Eigenwert $\lambda \in K$. Man nennt

$$H_\lambda(f) := \operatorname{Ker}\big((f - \lambda \operatorname{id}_V)^n\big) \le V$$

den *Hauptraum* von f zu λ. Ist λ Eigenwert einer Matrix $A \in K^{n\times n}$, so definiert man analog $H_\lambda(A) := \operatorname{Ker}\big((A - \lambda 1_n)^n\big)$.

Beispiel 14.15. Sei $f \in \operatorname{End}(K^2)$ mit $A := [f] = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Dann ist $E_1(f) = \langle e_1 \rangle$. Wegen $(A - 1_2)^2 = 0$ ist $H_1(f) = K^2$, d. h. der Hauptraum ist größer als der Eigenraum.

Bemerkung 14.16. Für $v \in E_\lambda(f)$ gilt

$$(f - \lambda \operatorname{id}_V)^n(v) = (f - \lambda \operatorname{id}_V)^{n-1}\big((f - \lambda \operatorname{id}_V)(v)\big) = (f - \lambda \operatorname{id}_V)^{n-1}(0) = 0.$$

Dies zeigt $E_\lambda(f) \subseteq H_\lambda(f)$. Offenbar ist f mit $f - \lambda \operatorname{id}$ und $(f - \lambda \operatorname{id})^n$ vertauschbar. Für $v \in H_\lambda(f)$ gilt daher

$$(f - \lambda \operatorname{id}_V)^n(f(v)) = \big((f - \lambda \operatorname{id}_V)^n \circ f\big)(v) = f\big((f - \lambda \operatorname{id}_V)^n(v)\big) = f(0) = 0.$$

Also ist $f(v) \in H_\lambda(f)$ und $H_\lambda(f)$ ist f-invariant. Wir zeigen als Nächstes, dass $H_\lambda(f)$ ein f-invariantes Komplement besitzt.

Lemma 14.17 (Fitting)**.** *Sei* $\dim V = n$ *und* $f \in \operatorname{End}(V)$. *Dann ist*

$$\boxed{V = f^n(V) \oplus \operatorname{Ker}(f^n)}$$

eine Zerlegung in f-invariante Unterräume.

Beweis. Für $v \in \operatorname{Ker}(f^k)$ gilt $f^{k+1}(v) = f(f^k(v)) = f(0) = 0$. Dies zeigt $\operatorname{Ker}(f) \le \operatorname{Ker}(f^2) \le \ldots$. Da die Dimension dieser Unterräume durch n beschränkt ist, existiert ein $k \le n$ mit $\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1})$. Für $v \in \operatorname{Ker}(f^{k+2})$ gilt $f^{k+1}(f(v)) = f^{k+2}(v) = 0$, also $f(v) \in \operatorname{Ker}(f^{k+1}) = \operatorname{Ker}(f^k)$. Dies zeigt $f^{k+1}(v) = 0$ und $v \in \operatorname{Ker}(f^{k+1})$. Induktiv erhält man

$$\operatorname{Ker}(f^k) = \operatorname{Ker}(f^{k+1}) = \ldots = \operatorname{Ker}(f^n) = \operatorname{Ker}(f^{n+1}) = \ldots.$$

Für $v \in \operatorname{Ker}(f^n) \cap f^n(V)$ existiert ein $w \in V$ mit $v = f^n(w)$. Wegen $f^{2n}(w) = f^n(v) = 0$ ist $w \in \operatorname{Ker}(f^{2n}) = \operatorname{Ker}(f^n)$ und es folgt $v = f^n(w) = 0$. Also ist $\operatorname{Ker}(f^n) \cap f^n(V) = \{0\}$. Der Homomorphiesatz zeigt $V = \operatorname{Ker}(f^n) \oplus f^n(V)$. Wegen $f(\operatorname{Ker}(f^n)) \subseteq \operatorname{Ker}(f^{n-1}) \subseteq \operatorname{Ker}(f^n)$ und $f(f^n(V)) = f^n(f(V)) \subseteq f^n(V)$ sind beide Unterräume f-invariant. $\square$

Beispiel 14.18. Seien $\lambda, \lambda' \in K$ verschiedene Eigenwerte von $f \in \operatorname{End}(V)$. Für $U := H_\lambda(f) \cap H_{\lambda'}(f)$ gilt
$$(f - \lambda \operatorname{id})^n(U) = \{0\} = (f - \lambda' \operatorname{id})^n(U).$$
Das Minimalpolynom der Einschränkung $g := f_{|U} \in \operatorname{End}(U)$ teilt daher $(X - \lambda)^n$ und $(X - \lambda')^n$ nach Lemma 10.44. Aus Lemma 10.24 folgt $\mu_g = 1$ und $U = \{0\}$.

Satz 14.19 (Hauptraumzerlegung). *Sei $f \in \operatorname{End}(V)$, sodass μ_f in Linearfaktoren zerfällt. Dann gilt*
$$V = H_{\lambda_1}(f) \oplus \ldots \oplus H_{\lambda_k}(f)$$
für die verschiedenen Eigenwerte $\lambda_1, \ldots, \lambda_k \in K$ von f.

Beweis. Wir argumentieren durch Induktion nach $n := \dim V$. Der Fall $n = 1$ ist trivial. Sei also $n \geq 2$ und die Behauptung für $n - 1$ bereits bewiesen. Da μ_f in Linearfaktoren zerfällt, besitzt f einen Eigenwert $\lambda = \lambda_1 \in K$ nach Satz 10.50. Für die Abbildung $g := f - \lambda \operatorname{id}_V$ gilt
$$V = g^n(V) \oplus \operatorname{Ker}(g^n) = g^n(V) \oplus H_\lambda(f)$$
nach Fitting. Für $U := g^n(V)$ gilt $f(U) = (g + \lambda \operatorname{id})(U) \subseteq g(U) + U \subseteq U$, d. h. U ist f-invariant. Wir betrachten nun die Einschränkung $h := f_{|U} \in \operatorname{End}(U)$. Nach Lemma 10.44 ist $\mu_h \mid \mu_f$ und μ_h zerfällt in Linearfaktoren (Lemma 10.24). Wegen $\dim H_\lambda(f) \geq \dim E_\lambda(f) \geq 1$ ist $\dim U < n$. Nach Induktion gilt daher
$$U = H_{\lambda_2}(h) \oplus \ldots \oplus H_{\lambda_k}(h)$$
für die verschiedenen Eigenwerte $\lambda_2, \ldots, \lambda_k$ von h. Wir zeigen $H_{\lambda_i}(h) = H_{\lambda_i}(f)$ für $i = 2, \ldots, k$. Wegen $E_{\lambda_i}(h) \subseteq E_{\lambda_i}(f)$ sind $\lambda_2, \ldots, \lambda_k$ auch Eigenwerte von f. Wegen $E_{\lambda_i}(h) \cap H_\lambda(f) \leq U \cap H_\lambda(f) = \{0\}$ ist $\lambda \neq \lambda_i$ für $i = 2, \ldots, k$. Sicher ist $H_{\lambda_i}(h) \leq H_{\lambda_i}(f)$. Sei umgekehrt $v \in H_{\lambda_i}(f)$ für ein $i \geq 2$. Dann existieren $u \in U$ und $w \in H_\lambda(f)$ mit $v = u + w$. Es folgt
$$0 = (f - \lambda_i \operatorname{id})^n(v) = (f - \lambda_i \operatorname{id})^n(u) + (f - \lambda_i \operatorname{id})^n(w) \in U \oplus H_\lambda(f).$$
Aus Lemma 8.9 erhält man $(f - \lambda_i \operatorname{id})^n(u) = 0 = (f - \lambda_i \operatorname{id})^n(w)$. Nach Beispiel 14.18 ist $w \in H_\lambda(f) \cap H_{\lambda_i}(f) = \{0\}$ und $v = u \in H_{\lambda_i}(f) \cap U = H_{\lambda_i}(h)$. Dies zeigt
$$V = H_\lambda(f) \oplus U = H_\lambda(f) \oplus H_{\lambda_2}(h) \oplus \ldots \oplus H_{\lambda_k}(h) = H_{\lambda_1}(f) \oplus \ldots \oplus H_{\lambda_k}(f). \qquad \square$$

Bemerkung 14.20. Zerfällt μ_f in paarweise verschiedene Linearfaktoren, so ist f diagonalisierbar (Satz 10.52). Die Hauptraumzerlegung ist dann genau die Zerlegung in Eigenräume, denn $E_\lambda(f) \subseteq H_\lambda(f)$.

Beispiel 14.21. Sei
$$A := \begin{pmatrix} . & 1 & 1 & 1 \\ 1 & 1 & 1 & . \\ 1 & 1 & . & 1 \\ . & . & 1 & 1 \end{pmatrix} \in \mathbb{F}_2^{4\times 4}.$$
Da 0 und 1 die einzig möglichen Eigenwerte sind, berechnen wir probeweise
$$A^2 = \begin{pmatrix} . & . & . & . \\ . & 1 & . & . \\ 1 & . & 1 & . \\ 1 & 1 & 1 & . \end{pmatrix}, \qquad (A - 1_4)^2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & . & 1 & . \\ 1 & 1 & 1 & 1 \\ . & . & 1 & . \end{pmatrix}^2 = \begin{pmatrix} 1 & . & . & . \\ . & . & . & . \\ 1 & . & . & . \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$
Man sieht $(0,0,0,1), (1,0,1,0) \in H_0(A)$ und $(0,1,1,0), (0,1,0,1) \in H_1(A)$. Aus Dimensionsgründen folgt
$$\mathbb{F}_2^4 = H_0(A) \oplus H_1(A) = \langle (0,0,0,1), (1,0,1,0) \rangle \oplus \langle (0,1,1,0), (0,1,0,1) \rangle.$$

14.2 Jordanblöcke

Bemerkung 14.22. Nach der Hauptraumzerlegung interessieren wir uns für Basen von Haupträumen.

Definition 14.23. Für $\lambda \in K$ und $n \geq 1$ nennt man

$$J_n(\lambda) := \begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix} \in K^{n\times n}$$

einen *Jordanblock* zum Eigenwert λ.[1]

Bemerkung 14.24. Offenbar hat $A := J_n(\lambda)$ das charakteristisches Polynom $\chi_A = (X-\lambda)^n$, d. h. λ hat algebraische Vielfachheit n. Andererseits hat λ geometrische Vielfachheit $\dim E_\lambda(A) = 1$. Somit ist A besonders weit davon entfernt diagonalisierbar zu sein (Satz 10.34). Für den Standardbasisvektor $e_1 \in K^n$ gilt $Ae_1 = (*, 1, 0, \ldots, 0)^{\mathrm{t}}$ und induktiv

$$A^k e_1 = A(A^{k-1}e_1) = A(\underbrace{*, \ldots, *}_{k-1}, 1, 0, \ldots, 0)^{\mathrm{t}} = (\underbrace{*, \ldots, *}_{k}, 1, 0, \ldots, 0)^{\mathrm{t}}.$$

Also sind $e_1, Ae_1, \ldots, A^{n-1}e_1$ linear unabhängig. Daher müssen auch die Matrizen $1_n, A, \ldots, A^{n-1}$ linear unabhängig sein. Dies zeigt $\deg \mu_A \geq n$. Aus Cayley-Hamilton folgt $\mu_A = \chi_A = (X-\lambda)^n$.

Definition 14.25. Man nennt $f \in \mathrm{End}(V)$ (bzw. $A \in K^{n\times n}$) *nilpotent*, falls $f^m = f \circ \ldots \circ f = 0$ (bzw. $A^m = 0_n$) für ein $m \in \mathbb{N}$ gilt.

Beispiel 14.26. Sei A eine strikte (obere oder untere) Dreiecksmatrix. Dann ist $\chi_A = X^n$. Nach Cayley-Hamilton ist $\mu_A \mid X^n$ und daher $A^n = 0_n$. Also ist A nilpotent. Insbesondere sind Jordanblöcke zum Eigenwert 0 nilpotent.

Bemerkung 14.27. Ist $A \in K^{n\times n}$ nilpotent, so ist $\mu_A \mid X^m$ für ein $m \in \mathbb{N}$. Wegen $\deg \mu_A \leq n$ ist $\mu_A \mid X^n$ und $A^n = 0$. Der folgende Satz liefert ein kanonisches Repräsentantensystem für die Ähnlichkeitsklassen von nilpotenten Matrizen. Die Anzahl der Ähnlichkeitsklassen ist die Anzahl $p(n)$ der *Partitionen* von n, d. h. Zerlegungen der Form $n = n_1 + \ldots + n_k$ mit $n_1 \geq \ldots \geq n_k$. Zum Beispiel gilt $p(5) = 7$, denn

$$5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1.$$

Man kennt keine einfache Formel für $p(n)$.

Satz 14.28. *Sei $f \in \mathrm{End}(V)$ nilpotent. Dann existiert eine Basis B von V, sodass*

$${}_B[f]_B = \mathrm{diag}(J_{n_1}(0), \ldots, J_{n_s}(0)).$$

Die Zahlen $n_1 \geq \ldots \geq n_s \geq 1$ sind durch die Gleichungen

$$\boxed{\big|\{1 \leq i \leq s : n_i \geq k\}\big| = \mathrm{rk}(f^{k-1}) - \mathrm{rk}(f^k)} \qquad (k = 1, \ldots, m) \tag{14.1}$$

eindeutig bestimmt. Insbesondere ist $s = \dim \mathrm{Ker}(f)$.

[1] In manchen Büchern benutzt man $J_n(\lambda)^{\mathrm{t}}$. Das macht keinen wesentlichen Unterschied (Satz 14.44).

Beweis. Sei $m \in \mathbb{N}$ mit $f^m = 0 \neq f^{m-1}$. Für $k \in \mathbb{N}$ gilt $f(\operatorname{Ker}(f^k)) \subseteq \operatorname{Ker}(f^{k-1})$. Nach Folgerung 4.16 existieren Unterräume $U_1, \ldots, U_m$ mit

$$\begin{aligned} V = \operatorname{Ker}(f^m) &= \operatorname{Ker}(f^{m-1}) \oplus U_1, \\ \operatorname{Ker}(f^{m-1}) &= \big(\operatorname{Ker}(f^{m-2}) + f(U_1)\big) \oplus U_2, \\ &\vdots \\ \operatorname{Ker}(f) &= \big(f^{m-1}(U_1) + \ldots + f(U_{m-1})\big) \oplus U_m. \end{aligned}$$

Wir zeigen, dass alle Summen direkt sind. Dies ist für die erste Summe gegeben. Sei induktiv bereits gezeigt:

$$\operatorname{Ker}(f^{m-k+1}) = \operatorname{Ker}(f^{m-k}) \oplus f^{k-1}(U_1) \oplus f^{k-2}(U_2) \oplus \ldots \oplus f(U_{k-1}) \oplus U_k. \tag{14.2}$$

Sei $w + f^k(u_1) + \ldots + f(u_k) = 0$ mit $w \in \operatorname{Ker}(f^{m-k-1})$ und $u_i \in U_i$ für $i = 1, \ldots, k$. Dann folgt

$$\begin{aligned} f^{m-k}\big(f^{k-1}(u_1) + \ldots + u_k\big) &= f^{m-k-1}\big(f^k(u_1) + \ldots + f(u_k)\big) \\ &= f^{m-k-1}\big(w + f^k(u_1) + \ldots + f(u_k)\big) = f^{m-k-1}(0) = 0 \end{aligned}$$

und

$$f^{k-1}(u_1) + \ldots + u_k \in \operatorname{Ker}(f^{m-k}) \cap \big(f^{k-1}(U_1) \oplus \ldots \oplus f(U_{k-1}) \oplus U_k\big) \overset{(14.2)}{=} \{0\}.$$

Dies zeigt $f^{k-1}(u_1) = \ldots = u_k = 0$ (Lemma 8.9) und es folgt $w = 0$. Also ist

$$\operatorname{Ker}(f^{m-k}) = \operatorname{Ker}(f^{m-k-1}) \oplus f^k(U_1) \oplus f^{k-1}(U_2) \oplus \ldots \oplus f(U_k) \oplus U_{k+1}$$

wie gewünscht.

Insgesamt ist

$$V = U_1 \oplus f(U_1) \oplus \ldots \oplus f^{m-1}(U_1) \oplus U_2 \oplus f(U_2) \oplus \ldots \oplus f^{m-2}(U_2) \oplus \ldots \oplus U_m.$$

Sei $b_{i1}, \ldots, b_{ik_i}$ eine Basis von U_i für $i = 1, \ldots, m$ (der Fall $U_i = \{0\}$ mit $k_i = 0$ ist zugelassen). Wegen $\operatorname{Ker}(f^j) \cap U_i = \{0\}$ ist die Einschränkung $f^j_{|U_i}$ für $j = 1, \ldots, m-i$ injektiv (Lemma 7.7). Insbesondere ist $f^j(b_{i1}), \ldots, f^j(b_{ik_i})$ eine Basis von $f^j(U_i)$. Also ist

$$B := \bigcup_{i=1}^{m} \bigcup_{j=1}^{k_i} \{b_{ij}, f(b_{ij}), \ldots, f^{m-i}(b_{ij})\}$$

eine Basis von V. Wegen $U_i \subseteq \operatorname{Ker}(f^{m-i+1})$ ist $f^{m-i+1}(b_{ij}) = 0$. Somit entsprechen die Elemente $b_{ij}, f(b_{ij}), \ldots, f^{m-i}(b_{ij})$ dem Jordanblock $J_{m-i+1}(0)$ in ${}_B[f]_B$. Insgesamt hat ${}_B[f]_B$ nun die gewünschte Form. Außerdem ist

$$\begin{aligned} k_1 + \ldots + k_l &= \dim\big(f^{l-1}(U_1) \oplus f^{l-2}(U_2) \oplus \ldots \oplus U_l\big) = \dim \operatorname{Ker}(f^{m-l+1}) - \dim \operatorname{Ker}(f^{m-l}) \\ &\overset{7.12}{=} \operatorname{rk}(f^{m-l}) - \operatorname{rk}(f^{m-l+1}) \end{aligned}$$

die Anzahl der Jordanblöcke $J_{n_i}(0)$ mit $n_i \geq m - l + 1$. Indem man $m - l + 1$ durch k ersetzt, folgt (14.1). Die letzte Behauptung erhält man mit $k = 1$ in (14.1). □

Bemerkung 14.29. In der Situation von Satz 14.28 gilt

$$\begin{aligned} \big|\{1 \leq i \leq s : n_i = k\}\big| &= \big|\{1 \leq i \leq s : n_i \geq k\}\big| - \big|\{1 \leq i \leq s : n_i \geq k+1\}\big| \\ &= \operatorname{rk}(f^{k-1}) + \operatorname{rk}(f^{k+1}) - 2\operatorname{rk}(f^k) \end{aligned}$$

für $k = 1, \ldots, n$. Insbesondere ist $2\operatorname{rk}(f^k) \leq \operatorname{rk}(f^{k+1}) + \operatorname{rk}(f^{k-1})$, d. h. die Folge $\operatorname{rk}(f)$, $\operatorname{rk}(f^2), \ldots$ kann nicht zu „schnell“ fallen.

Beispiel 14.30. Sei $f \in \mathrm{End}(\mathbb{R}^5)$ mit

$$A := {}_B[f]_B = \begin{pmatrix} 2 & -3 & -1 & -1 & 2 \\ 1 & -2 & 0 & 0 & 1 \\ 0 & 2 & -1 & -1 & 0 \\ 1 & -4 & 1 & 1 & 1 \\ 0 & -1 & 1 & 1 & 0 \end{pmatrix}.$$

Man berechnet

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 & 1 \\ -1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad A^3 = 0.$$

Also ist $\mathrm{rk}(f^2) = 1$. Wie im Beweis von Satz 14.28 können wir $b_1 := e_1 \notin \mathrm{Ker}(f^2)$ und $U_1 := \langle b_1 \rangle$ mit $\mathbb{R}^5 = \mathrm{Ker}(f^2) \oplus \langle b_1 \rangle$ wählen. Weiter ist

$$A \sim \begin{pmatrix} 1 & -2 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 2 & -1 & -1 & 0 \\ 0 & -2 & 1 & 1 & 0 \\ 0 & -1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -2 & -2 & 1 \\ 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathrm{Ker}(f) \oplus f(U_1) = \langle \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \rangle.$$

Mit $b_2 := e_3 \in \mathrm{Ker}(f^2) \setminus \big(\mathrm{Ker}(f) \oplus f(U_1)\big)$ und $U_2 := \langle b_2 \rangle$ gilt $\mathrm{Ker}(f^2) = \mathrm{Ker}(f) \oplus f(U_1) \oplus U_2$. Schließlich ist $\mathrm{Ker}(f) = f^2(U_1) \oplus f(U_2)$ (also $U_3 := \{0\}$). Bezüglich der Basis $B := \{b_1, f(b_1), f^2(b_1), b_2, f(b_2)\}$ hat f die Darstellungsmatrix $\mathrm{diag}\big(J_3(0), J_2(0)\big)$.

Satz 14.31. *Sei $f \in \mathrm{End}(V)$, sodass μ_f in Linearfaktoren zerfällt. Dann existiert eine Basis B von V mit*

$${}_B[f]_B = \mathrm{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_s}(\lambda_s)), \qquad \text{(Jordan-Normalform)}$$

wobei $\lambda_1, \ldots, \lambda_s \in K$ die Eigenwerte von f sind (möglicherweise mit Vielfachheiten). Die Jordanblöcke $J_{n_i}(\lambda_i)$ sind bis auf die Reihenfolge eindeutig bestimmt.

Beweis. Wir fügen alle Puzzleteile zusammen: Seien $\lambda_1, \ldots, \lambda_k \in K$ die verschiedenen Eigenwerte von f und $H_i := H_{\lambda_i}(f)$ für $i = 1, \ldots, k$. Nach der Hauptraumzerlegung gilt $V = H_1 \oplus \ldots \oplus H_k$. Für $g_i := (f - \lambda_i \mathrm{id}_V)_{|H_i}$ gilt $g_i^n = 0$. Nach Satz 14.28 existiert eine Basis B_i von H_i mit

$${}_{B_i}[g_i]_{B_i} = \mathrm{diag}\big(J_{m_1}(0), \ldots, J_{m_t}(0)\big).$$

Nach Satz 7.18 ist

$$\begin{aligned} {}_{B_i}[f_{|H_i}]_{B_i} &= {}_{B_i}[g_i + \lambda_i \mathrm{id}_{H_i}]_{B_i} = {}_{B_i}[g_i]_{B_i} + \lambda_i {}_{B_i}[\mathrm{id}_{H_i}]_{B_i} = \mathrm{diag}\big(J_{m_1}(0), \ldots, J_{m_t}(0)\big) + \lambda_i 1 \\ &= \mathrm{diag}\big(J_{m_1}(\lambda_i), \ldots, J_{m_t}(\lambda_i)\big). \end{aligned}$$

Also ist $B := B_1 \cup \ldots \cup B_k$ eine geeignete Basis. Für ein fest gewähltes Paar (n_i, λ_i) ergibt sich die Anzahl der Blöcke $J_{n_i}(\lambda_i)$ aus Bemerkung 14.29 angewendet auf g_i. □

Folgerung 14.32. *Jede Matrix $A \in \mathbb{C}^{n\times n}$ besitzt eine Jordan-Normalform (genauer: A ist zu einer Jordan-Normalform ähnlich).*

Beweis. Nach Folgerung 11.35 zerfällt μ_A in Linearfaktoren. $\square$

Bemerkung 14.33.

(a) Da man die Eigenwerte $\lambda_1, \ldots, \lambda_n$ in der Regel nicht sinnvoll ordnen kann (vgl. Bemerkung 11.26), gibt es im Gegensatz zu Satz 14.28 keine kanonische Reihenfolge der Jordanblöcke. Man kann jedoch die *lexikografische* Ordnung

$$x <_l y \;:\Longleftrightarrow\; \mathrm{Re}(x) < \mathrm{Re}(y) \vee \big(\mathrm{Re}(x) = \mathrm{Re}(y) \wedge \mathrm{Im}(x) < \mathrm{Im}(y)\big)$$

auf $\mathbb{C}$ definieren und die Jordanblöcke zunächst nach Größe und dann nach Eigenwert bzgl. $<_l$ sortieren. Im Folgenden sprechen wir etwas ungenau von *der* Jordan-Normalform von f.

(b) Sei $A \in K^{n\times n}$. In Satz 15.39 konstruieren wir einen Körper $L \supseteq K$, sodass μ_A in $L[X]$ in Linearfaktoren zerfällt. Man kann dann A als Matrix in $L^{n\times n}$ auffassen und dort die Jordan-Normalform bestimmen.

Beispiel 14.34. Sei $f \in \mathrm{End}(\mathbb{C}^3)$ mit

$$A := [f] = \begin{pmatrix} 5 & 0 & 1 \\ -5-\mathrm{i} & -\mathrm{i} & -1 \\ -9 & 0 & -1 \end{pmatrix}.$$

Durch Laplace-Entwicklung nach der zweiten Spalte erhält man

$$\chi_f = \chi_A = (X+\mathrm{i})\big((X-5)(X+1)+9\big) = (X+\mathrm{i})(X^2-4X+4) = (X+\mathrm{i})(X-2)^2.$$

Also sind $\lambda_1 = 2$ und $\lambda_2 = -\mathrm{i}$ die Eigenwerte von f. Offensichtlich ist $b_3 := e_2$ ein Eigenvektor zu λ_2. Wegen

$$A - 2\cdot 1_3 = \begin{pmatrix} 3 & 0 & 1 \\ -5-\mathrm{i} & -2-\mathrm{i} & -1 \\ -9 & 0 & -3 \end{pmatrix} \sim \begin{pmatrix} 3 & 0 & 1 \\ -5-\mathrm{i} & -2-\mathrm{i} & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

ist $\mathrm{rk}(f - 2\,\mathrm{id}) = 2$, d. h. $\lambda_1 = 2$ hat geometrische Vielfachheit 1. Also ist f nicht diagonalisierbar und die Jordan-Normalform von f muss $\mathrm{diag}\big(J_2(2), J_1(-\mathrm{i})\big)$ sein. Wegen

$$(A - 2\cdot 1_3)^2 = \begin{pmatrix} 0 & 0 & 0 \\ 3+4\mathrm{i} & 3+4\mathrm{i} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

können wir $b_1 := e_3 \in \mathrm{Ker}\big((f-2\,\mathrm{id})^2\big) \setminus \mathrm{Ker}(f - 2\,\mathrm{id})$ wie in Beispiel 14.30 wählen. Für

$$b_2 := (f - 2\,\mathrm{id})(b_1) = (1, -1, -3)$$

gilt $f(b_1) = 2b_1 + b_2$ und $f(b_2) = (f - 2\,\mathrm{id})(b_2) + 2b_2 = 2b_2$. Für die Basis $B := \{b_1, b_2, b_3\}$ erhält man

$${}_B[f]_B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -\mathrm{i} \end{pmatrix} = \mathrm{diag}\big(J_2(2), J_1(-\mathrm{i})\big).$$

14.3 Anwendungen

Satz 14.35. *Sei* $J := \operatorname{diag}(J_{n_1}(\lambda_1), \ldots, J_{n_s}(\lambda_s))$ *die Jordan-Normalform von* $A \in K^{n\times n}$. *Seien* $\rho_1, \ldots, \rho_k$ *die verschiedenen Eigenwerte von* A. *Für* $i = 1, \ldots, k$ *sei* a_i, m_i *und* s_i *die Anzahl, das Maximum und die Summe der* n_j *mit* $\lambda_j = \rho_i$. *Dann ist* a_i *die geometrische Vielfachheit von* ρ_i *und*

$$\chi_A = (X - \rho_1)^{s_1} \ldots (X - \rho_k)^{s_k} \qquad\qquad \mu_A = (X - \rho_1)^{m_1} \ldots (X - \rho_k)^{m_k}.$$

Insbesondere ist A *genau dann diagonalisierbar, wenn* J *eine Diagonalmatrix ist.*

Beweis. Man sieht leicht $\operatorname{rk}(J - \mu_i 1_n) = n - a_i$, d. h. a_i ist die geometrische Vielfachheit von ρ_i (vgl. Satz 14.28). Da J eine untere Dreiecksmatrix ist, ist s_i die algebraische Vielfachheit von ρ_i. Da μ_A nicht von der Basiswahl abhängt, gilt

$$0 = \mu_A(J) = \operatorname{diag}\bigl(\mu_A(J_{n_1}(\lambda_1)), \ldots, \mu_A(J_{n_s}(\lambda_s))\bigr).$$

Aus Bemerkung 14.24 folgt $\mu_A = (X - \rho_1)^{m_1} \ldots (X - \rho_k)^{m_k}$. Die zweite Behauptung gilt nach Satz 10.52. $\square$

Beispiel 14.36. Für $A = \operatorname{diag}(J_3(1), J_2(1), J_4(\mathrm{i}))$ gilt $\chi_A = (X-1)^5(X-\mathrm{i})^4$ und $\mu_A = (X-1)^3(X-\mathrm{i})^4$.

Folgerung 14.37. *Sei* $A, B \in \mathbb{C}^{n\times n}$. *Im Fall* $n \le 3$ *gilt*

$$A \approx B \iff \chi_A = \chi_B,\ \mu_A = \mu_B.$$

Im Fall $n \le 6$ *sind folgende Aussagen äquivalent:*

(1) $A \approx B$.

(2) $\chi_A = \chi_B$, $\mu_A = \mu_B$ *und die geometrischen Vielfachheiten der Eigenwerte von* A *stimmen mit denen von* B *überein.*

Beweis. Bekanntlich impliziert $A \approx B$ die jeweils andere Aussage. Sei nun $n \le 3$, $\chi_A = \chi_B$ und $\mu_A = \mu_B$. Dann haben A und B die gleichen Eigenwerte mit den gleichen algebraischen Vielfachheiten. Sei λ ein Eigenwert mit algebraischer Vielfachheit $r \le 3$. Für die entsprechenden Größen der Jordanblöcke $n_1 \ge \ldots \ge n_k$ von A gilt $r = n_1 + \ldots + n_k$. Da n_1 durch μ_A festgelegt ist, sind auch k und $n_2, \ldots, n_k$ durch r eindeutig bestimmt. Also haben A und B die gleiche Jordan-Normalform. Es folgt $A \approx B$.

Sei jetzt $n \le 6$ und (2) erfüllt. Neben r und n_1 ist nun auch die geometrische Vielfachheit k von λ eindeutig bestimmt. Es gibt folgende Möglichkeiten:

- $k = 1$: Hier ist $n_1 = r$.
- $k = 2$: Hier ist $n_2 = r - n_1$.
- $k = 3$: Wegen $r \le 6$ ist $(n_1, n_2, n_3) \in \{(1,1,1), (2,1,1), (2,2,1), (2,2,2), (3,1,1), (3,2,1), (4,1,1)\}$. Diese Tripel sind durch r und n_1 eindeutig bestimmt.
- $k = 4$: Hier ist $(n_1, n_2, n_3, n_4) \in \{(1,1,1,1), (2,1,1,1), (2,2,1,1), (3,1,1,1)\}$.
- $k = 5$: Hier ist $n_1 = r - 4$ und $n_2 = \ldots = n_5 = 1$.
- $k = 6$: Hier ist $n_1 = \ldots = n_6 = 1$.

In allen Fällen folgt $A \approx B$. $\square$

Beispiel 14.38. Die Matrizen

$$\begin{aligned}\operatorname{diag}(J_2(0), J_2(0)) &\not\approx \operatorname{diag}(J_2(0), J_1(0), J_1(0)),\\ \operatorname{diag}(J_3(0), J_2(0), J_2(0)) &\not\approx \operatorname{diag}(J_3(0), J_3(0), J_1(0))\end{aligned}$$

zeigen, dass man Folgerung 14.37 nicht auf $n = 4$ bzw. $n = 7$ erweitern kann.

Satz 14.39. *Eine Matrix $A \in \mathbb{C}^{n\times n}$ ist genau dann nilpotent, wenn $\operatorname{tr}(A^k) = 0$ für $k = 1, \ldots, n$ gilt.*

Beweis. O. B. d. A. sei A in Jordan-Normalform mit paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_s \in \mathbb{C}$. Sei m_i die algebraische Vielfachheit von λ_i. Die Eigenwerte von A^k sind dann $\lambda_1^k, \ldots, \lambda_s^k$ mit den entsprechenden Vielfachheiten. Nach Bemerkung 10.35 gilt

$$\operatorname{tr}(A^k) = m_1\lambda_1^k + \ldots + m_s\lambda_s^k = 0.$$

Ist A nilpotent, so gilt $s = 1$, $\lambda_1 = 0$ und $\operatorname{tr}(A^k) = 0$ für $k = 1, \ldots, n$.

Sei umgekehrt $\operatorname{tr}(A^k) = m_1\lambda_1^k + \ldots + m_s\lambda_s^k = 0$ für $k = 1, \ldots, n$. Nehmen wir indirekt an, dass A nicht nilpotent ist. O. B. d. A. sei $\lambda_i \ne 0$ für $i = 1, \ldots, s$. Dann ist $(m_1, \ldots, m_s)^{\mathrm{t}}$ eine Lösung des homogenen linearen Gleichungssystem mit Koeffizientenmatrix $V = (\lambda_j^i) \in K^{s\times s}$. Nach Vandermonde ist allerdings $\det(V) = \lambda_1 \ldots \lambda_s \det(\lambda_j^{i-1}) \ne 0$. Widerspruch. □

Beispiel 14.40. Satz 14.39 gilt nicht über beliebigen Körpern. Zum Beispiel ist $\operatorname{tr}(1_2^k) = \operatorname{tr}(1_2) = 0$ in $\mathbb{F}_2$ für alle $k \in \mathbb{N}$.

Bemerkung 14.41. Nach Satz 12.46 besitzt jede positiv (semi)definite reelle Matrix eine k-te Wurzel. Für beliebige Matrizen ist das falsch: $J_2(0)$ besitzt keine Quadratwurzel, denn für $A \in K^{2\times 2}$ mit $A^2 = J_2(0)$ wäre $A^4 = 0$ und damit $A^2 = 0$ (Bemerkung 14.27).

Satz 14.42. *Für $k \in \mathbb{N}$ und $A \in \mathrm{GL}(n, \mathbb{C})$ existiert ein $W \in \mathbb{C}^{n\times n}$ mit $W^k = A$.*

Beweis. Wegen $(SWS^{-1})^k = SW^kS^{-1}$ für $S \in \mathrm{GL}(n, \mathbb{C})$ können wir annehmen, dass A eine Jordan-Normalform ist. Es genügt eine k-te Wurzel für jeden Jordanblock zu konstruieren. Sei also $A = J_n(\lambda)$ mit $\lambda \in \mathbb{C}$. Da A invertierbar ist, gilt $\lambda \ne 0$. Nach Lemma 11.27 existiert ein $\mu \in \mathbb{C}$ mit $\mu^k = \lambda$. Sei $J := J_n(\mu)$. Mit Induktion nach k zeigt man:

$$J^k = \begin{pmatrix} \mu^k & & & 0\\ k\mu^{k-1} & \ddots & & \\ & \ddots & \ddots & \\ * & & k\mu^{k-1} & \mu^k \end{pmatrix}.$$

Also ist $\mu^k = \lambda$ der einzige Eigenwert von J^k und die geometrische Vielfachheit beträgt $n - 1$. Nach Satz 14.35 ist A die Jordan-Normalform von J^k. Insbesondere ist $A \approx J^k$. □

Beispiel 14.43.

(a) Der Beweis zeigt $J_n(1)^k \approx J_n(1)$ für alle $k, n \in \mathbb{N}$. Wir suchen eine dritte Wurzel von $J_3(1)$. Für

$$S := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 3 & 9 \end{pmatrix}$$

gilt

$$S^{-1}J_3(1)^3 S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & -1/9 & 1/9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 3 & 9 \end{pmatrix} = J_3(1).$$

Für

$$W := S^{-1}J_3(1)S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & -1/9 & 1/9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 6 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ -1/9 & 1/3 & 1 \end{pmatrix}$$

gilt $W^3 = J_3(1)$. In Satz 15.35 bestimmen wir die Anzahl der k-ten Wurzeln von $J_n(\lambda)$ für alle $\lambda \in \mathbb{C}$ und $n, k \in \mathbb{N}$.

(b) Man rechnet leicht nach, dass $J_2(1) \in \mathrm{GL}(2, \mathbb{F}_2)$ *keine* Quadratwurzel besitzt.

Satz 14.44. *Für alle $A \in \mathbb{C}^{n\times n}$ gilt $A \approx A^{\mathrm{t}}$.*

Beweis. Wie üblich können wir $A = J_n(\lambda)$ annehmen. Da λ der einzige Eigenwert von A^{t} ist und die geometrische Vielfachheit 1 beträgt, ist A die Jordan-Normalform von A^{t}. $\square$

Bemerkung 14.45. Man kann Satz 14.44 auch direkt nachrechnen:

$$\begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}^{(-1)} J_n(\lambda) \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix} = \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix} \begin{pmatrix} 0 & & & \lambda \\ & & \iddots & 1 \\ & \iddots & \iddots & \\ \lambda & 1 & & 0 \end{pmatrix} = J_n(\lambda)^{\mathrm{t}}.$$

In Satz 15.25 beweisen wir Satz 14.44 über beliebigen Körpern.

15 Die Frobenius-Normalform

15.1 Irreduzible Polynome

Bemerkung 15.1. Wir konstruieren in diesem Kapitel eine Normalform für Endomorphismen, die im Gegensatz zur Jordan-Normalform über jedem Körper existiert. Da im Allgemeinen nicht jedes Polynom in Linearfaktoren zerfällt, kann man nicht erwarten, dass die Normalform eine Dreiecksmatrix ist (Satz 14.8). Dennoch erzielen wir die gleiche Anzahl an Nulleinträgen wie in der Jordan-Normalform. Die Grundidee ist, das Minimalpolynom in möglichst „kleine“ Polynome zu faktorisieren. Wir immer sei V ein endlich-dimensionaler K-Vektorraum.

Definition 15.2.

- Ein normiertes Polynom $\alpha \in K[X] \setminus K$ heißt *irreduzibel*, falls es keine Faktorisierung $\alpha = \beta\gamma$ mit $\beta, \gamma \in K[X] \setminus K$ gibt (vgl. Primzahl).
- Wir nennen $\alpha, \beta \in K[X]$ *teilerfremd*, falls jeder gemeinsame Teiler von α und β konstant ist, also in K liegt.

Beispiel 15.3.

(a) Normierte Polynome vom Grad 1 sind irreduzibel, denn $\deg(\alpha\beta) = \deg(\alpha) + \deg(\beta)$ für $\alpha, \beta \in K[X]$. In $\mathbb{C}[X]$ hat umgekehrt jedes irreduzible Polynom Grad 1 nach dem Fundamentalsatz der Algebra und Lemma 10.22. Im Allgemeinen sind normierte Polynome vom Grad ≤ 3 genau dann irreduzibel, wenn sie keine Nullstelle besitzen. Zum Beispiel ist $X^2 + X + 1 \in \mathbb{F}_2[X]$ irreduzibel.

(b) Das Polynom $X^2 - 2$ ist irreduzibel in $\mathbb{Q}[X]$, aber nicht in $\mathbb{R}[X]$, denn $X^2 - 2 = (X + \sqrt{2})(X - \sqrt{2})$. Analog ist $X^2 + 1$ irreduzibel in $\mathbb{R}[X]$, aber nicht in $\mathbb{C}[X]$ wegen $X^2 + 1 = (X + \mathrm{i})(X - \mathrm{i})$.

(c) Zwei verschiedene irreduzible Polynome sind teilerfremd. Ist $\gamma \in K[X] \setminus K$ ein gemeinsamer Teiler von α und β, so ist auch jeder irreduzible Teiler von γ ein gemeinsamer Teiler von α und β. Also sind α und β genau dann teilerfremd, wenn sie keine irreduziblen gemeinsamen Teiler haben.

Satz 15.4 (Euklidischer Algorithmus)**.** *Seien $\alpha, \beta \in K[X]$ mit $\deg \alpha \geq \deg \beta$. Der folgende Algorithmus bestimmt, ob α und β teilerfremd sind:*

(1) Setze $\alpha_0 := \alpha$, $\alpha_1 := \beta$ und $i := 1$.

(2) Wiederhole:

- *Division mit Rest: $\alpha_{i-1} = \alpha_i\gamma_i + \alpha_{i+1}$ mit $\deg(\alpha_{i+1}) < \deg(\alpha_i)$.*
- *Gilt $\alpha_{i+1} = 0$, so breche ab.*
- *Erhöhe i um 1.*

(3) Genau dann sind α und β teilerfremd, wenn $\alpha_i \in K$ gilt.

Beweis. Jeder gemeinsame Teiler von α_{i-1} und α_i teilt auch $\alpha_{i-1} - \alpha_i\gamma_i = \alpha_{i+1}$ für $i = 1, 2, \ldots$. Da $\deg(\alpha_i)$ mit jeder Division mit Rest kleiner wird, muss der Algorithmus nach endlich vielen Schritten abbrechen. Am Ende gilt $\alpha_{i+1} = 0 \neq \alpha_i$, d. h. $\alpha_{i-1} = \alpha_i\gamma_i$. Daher sind α und β genau dann teilerfremd, wenn α_i konstant ist. □

Beispiel 15.5. Sei $\alpha = X^5 + X^4 + X^3 + 1 \in \mathbb{F}_2[X]$ und $\beta = X^4 + X^3 + X + 1 \in \mathbb{F}_2[X]$. Es gilt

$$\begin{aligned}
X^5 + X^4 + X^3 + 1 &= (X^4 + X^3 + X + 1)X + X^3 + X^2 + X + 1 &&\Longrightarrow \alpha_2 = X^3 + X^2 + X + 1,\\
X^4 + X^3 + X + 1 &= (X^3 + X^2 + X + 1)X + X^2 + 1 &&\Longrightarrow \alpha_3 = X^2 + 1 \notin \mathbb{F}_2,\\
X^3 + X^2 + X + 1 &= (X^2 + 1)(X + 1) &&\Longrightarrow \alpha_4 = 0.
\end{aligned}$$

Also ist $X^2 + 1$ ein gemeinsamer Teiler von α und β.

Lemma 15.6 (Bézout)**.** *Sind $\alpha, \beta \in K[X]$ teilerfremd, so existieren $\tilde{\alpha}, \tilde{\beta} \in K[X]$ mit $\alpha\tilde{\alpha} + \beta\tilde{\beta} = 1$.*

Beweis. Nach Voraussetzung sind α und β nicht beide 0. Also existieren $\tilde{\alpha}, \tilde{\beta} \in K[X]$, sodass $\rho := \alpha\tilde{\alpha} + \beta\tilde{\beta} \neq 0$ minimalen Grad hat. Division mit Rest liefert $\gamma, \delta \in K[X]$ mit $\alpha = \gamma\rho + \delta$ und $\deg\delta < \deg\rho$. Also ist

$$\delta = \alpha - \gamma\rho = \alpha - \gamma\alpha\tilde{\alpha} - \gamma\beta\tilde{\beta} = \alpha(1 - \gamma\tilde{\alpha}) - \beta(\gamma\tilde{\beta}).$$

Die Wahl von ρ zeigt $\delta = 0$. Es folgt $\rho \mid \alpha$. Analog ergibt sich $\rho \mid \beta$. Da α und β teilerfremd sind, ist $\rho \in K^\times$. Nach Normierung kann man $\rho = 1$ annehmen. □

Bemerkung 15.7. Die Polynome $\tilde{\alpha}$ und $\tilde{\beta}$ in Lemma 15.6 lassen sich berechnen, indem man den euklidischen Algorithmus rückwärts liest

$$\alpha_i = \alpha_{i-2} - \alpha_{i-1}\gamma_{i-1} = \alpha_{i-2} - (\alpha_{i-3} - \alpha_{i-2}\gamma_{i-2})\gamma_{i-1} = \alpha_{i-3}\gamma_{i-2} + \alpha_{i-2}(1 - \gamma_{i-2}\gamma_{i-1}) = \ldots.$$

und durch α_i teilt.

Beispiel 15.8. Sei $\alpha := X^3 + X^2 + 1$ und $\beta := X^2 - X$. Der euklidische Algorithmus liefert

$$\begin{aligned}
X^3 + X^2 + 1 &= (X^2 - X)(X + 2) + 2X + 1\\
X^2 - X &= (2X + 1)\frac{1}{4}(2X - 3) + \frac{3}{4}.
\end{aligned}$$

Also ist

$$\begin{aligned}
\frac{3}{4} &= \beta - \frac{1}{4}(2X + 1)(2X - 3) = \beta - \frac{1}{4}\Big(\alpha - \beta(X + 2)\Big)(2X - 3)\\
&= -\frac{1}{4}(2X - 3)\alpha + \frac{1}{4}\Big(4 + (X + 2)(2X - 3)\Big)\beta
\end{aligned}$$

und

$$-\frac{1}{3}(2X - 3)\alpha + \frac{1}{3}(2X^2 + X - 2)\beta = 1.$$

Satz 15.9 (Primfaktorzerlegung in $K[X]$)**.** *Jedes normierte Polynom in $K[X] \setminus K$ ist ein Produkt von irreduziblen Faktoren, die bis auf die Reihenfolge eindeutig bestimmt sind.*

Beweis. Sei $\alpha \in K[X]$ normiert. Induktion nach $d := \deg(\alpha)$. Ist α irreduzibel (zum Beispiel $d = 1$), so sind wir fertig. Anderenfalls ist $\alpha = \beta\gamma$ mit $\deg(\beta), \deg(\gamma) < d$. Wir können annehmen, dass β und γ normiert sind. Nach Induktion sind β und γ Produkte von irreduziblen Faktoren und daher auch α.

Sei nun $\alpha = \sigma_1 \dots \sigma_n = \tau_1 \dots \tau_m$ mit irreduziblen Polynomen $\sigma_1, \dots, \sigma_n, \tau_1, \dots, \tau_m \in K[X]$. Induktion nach m. Im Fall $m = 1$ ist $n = 1$ und $\sigma_1 = \tau_1$. Sei nun $m \geq 2$. Im Fall $\sigma_1 \neq \tau_1$ sind σ_1 und τ_1 teilerfremd. Nach Bézout existieren $\tilde{\sigma}, \tilde{\tau} \in K[X]$ mit $\sigma_1\tilde{\sigma} + \tau_1\tilde{\tau} = 1$. Es folgt

$$\sigma_1(\tilde{\sigma}\tau_2 \dots \tau_m + \sigma_2 \dots \sigma_n\tilde{\tau}) = \tau_2 \dots \tau_m.$$

Induktiv erhält man $\sigma_1 = \tau_i$ für ein $i \in \{1, \dots, m\}$. Dann ist $\sigma_2 \dots \sigma_n = \tau_1 \dots \tau_{i-1}\tau_{i+1} \dots \tau_m$ und Induktion liefert $\{\sigma_1, \dots, \sigma_n\} = \{\tau_1, \dots, \tau_m\}$. $\square$

Beispiel 15.10. Wie bei natürlichen Zahlen werden wir in der Primfaktorzerlegung von Polynomen gleiche irreduzible Faktoren in Potenzen zusammenfassen. Zum Beispiel ist

$$X^6 + X^5 - X^4 - X^3 = X^3(X + 1)^2(X - 1).$$

Aus algorithmischer Sicht ist die Bestimmung der Primfaktorzerlegung (von Zahlen wie Polynomen) ein sehr schwieriges Problem.

Folgerung 15.11. *Seien $\alpha, \beta \in K[X]$ und γ ein irreduzibler Teiler von $\alpha\beta$. Dann gilt $\gamma \mid \alpha$ oder $\gamma \mid \beta$.*

Beweis. Wegen $\gamma \mid \alpha\beta$ muss γ in der Primfaktorzerlegung von $\alpha\beta$ auftauchen. Diese Primfaktorzerlegung erhält man, indem man die Primfaktorzerlegungen von α und β zusammenfasst. Also muss γ in der Zerlegung von α oder β auftreten. $\square$

Satz 15.12 (Primärzerlegung)**.** *Sei $f \in \mathrm{End}(V)$ und $\mu_f = \gamma_1^{a_1} \dots \gamma_k^{a_k}$ die Primfaktorzerlegung von μ_f in $K[X]$. Dann ist*

$$V = \mathrm{Ker}\big(\gamma_1^{a_1}(f)\big) \oplus \dots \oplus \mathrm{Ker}\big(\gamma_k^{a_k}(f)\big)$$

eine Zerlegung in f-invariante Unterräume. Außerdem ist $\gamma_i^{a_i}$ das Minimalpolynom der Einschränkung von f auf $\mathrm{Ker}\big(\gamma_i^{a_i}(f)\big)$ für $i = 1, \dots, k$.

Beweis. Für $1 \leq i \leq k$ sei $V_i := \mathrm{Ker}(\gamma_i^{a_i}(f))$. Für $v \in V_i$ gilt

$$\gamma_i^{a_i}(f)(f(v)) = f(\gamma_i^{a_i}(f)(v)) = f(0) = 0,$$

d. h. $f(v) \in V_i$. Also ist V_i f-invariant. Zum Nachweis der direkten Zerlegung argumentieren wir durch Induktion nach k. Für $k = 1$ ist $V_1 = \mathrm{Ker}(\mu_f(f)) = \mathrm{Ker}(0) = V$. Sei also $k \geq 2$. Die Polynome $\alpha := \gamma_1^{a_1}$ und $\beta := \gamma_2^{a_2} \dots \gamma_k^{a_k}$ sind teilerfremd, da sie keinen irreduziblen gemeinsamen Teiler haben. Nach Bézout existieren $\tilde{\alpha}, \tilde{\beta} \in K[X]$ mit $\alpha\tilde{\alpha} + \beta\tilde{\beta} = 1$. Sei $V_\beta := \mathrm{Ker}(\beta(f))$. Wegen $(\alpha\beta)(f) = 0 = (\beta\alpha)(f)$ ist $\beta(f)(V) \subseteq V_1$ und $\alpha(f)(V) \subseteq V_\beta$. Es folgt

$$V = \mathrm{id}(V) = (\alpha\tilde{\alpha} + \beta\tilde{\beta})(f)(V) \subseteq \alpha(f)(V) + \beta(f)(V) \subseteq V_\beta + V_1.$$

Für $v \in V_1 \cap V_\beta$ ist andererseits

$$v = (\alpha\tilde{\alpha} + \beta\tilde{\beta})(f)(v) = \tilde{\alpha}(f)\big(\alpha(f)(v)\big) + \tilde{\beta}(f)\big(\beta(f)(v)\big) = 0.$$

Dies zeigt $V = V_1 \oplus V_\beta$.

Das Minimalpolynom α_1 der Einschränkung $f_{|V_1}$ teilt α, denn $\alpha(f)(V_1) = 0$. Insbesondere gilt $\deg(\alpha_1) \le \deg(\alpha)$. Genauso ist β durch das Minimalpolynom β_1 von $f_\beta := f_{|V_\beta}$ teilbar. Wegen $V = V_1 \oplus V_\beta$ gilt andererseits $(\alpha_1\beta_1)(f) = 0$ und $\mu_f \mid \alpha_1\beta_1$. Aus

$$\deg(\alpha) + \deg(\beta) = \deg(\mu_f) \le \deg(\alpha_1\beta_1) = \deg(\alpha_1) + \deg(\beta_1)$$

folgt $\alpha_1 = \alpha$ sowie $\beta_1 = \beta$. Wir können nun die Induktionsvoraussetzung auf $f_\beta \in \mathrm{End}(V_\beta)$ anwenden. Für $i = 2, \ldots, k$ gilt $\mathrm{Ker}(\gamma_i^{a_i}(f_\beta)) = V_i \cap V_\beta = V_i$. Dies zeigt

$$V_\beta = V_2 \oplus \ldots \oplus V_k.$$

Die Behauptung folgt aus Lemma 8.9. $\square$

Bemerkung 15.13. Nehmen wir an, dass μ_f in Linearfaktoren zerfällt, d. h. es gilt $\gamma_i = X - \lambda_i$ für $i = 1, \ldots, k$, wobei $\lambda_1, \ldots, \lambda_k$ die verschiedenen Eigenwerte von f sind. Wegen

$$\mathrm{Ker}(\gamma_i^{a_i}(f)) = \mathrm{Ker}\big((f - \lambda_i \,\mathrm{id})^{a_i}\big) \subseteq \mathrm{Ker}\big((f - \lambda_i \,\mathrm{id})^n\big) = H_{\lambda_i}(f)$$

stimmt die Primärzerlegung mit der Hauptraumzerlegung aus Satz 14.19 überein.

Definition 15.14. Für $v \in V$ und $f \in \mathrm{End}(V)$ nennt man $U := \langle f^i(v) : i \in \mathbb{N}_0 \rangle \le V$ einen *zyklischen* Unterraum. Offenbar ist U f-invariant. Wir bezeichnen das Minimalpolynom von $f_{|U}$ mit μ_v.

Bemerkung 15.15. Sei $f \in \mathrm{End}(V)$ und $v \in V \setminus \{0\}$. Sei $d \in \mathbb{N}$ minimal, sodass $v, f(v), \ldots, f^d(v)$ linear abhängig sind. Dann existieren $a_0, \ldots, a_{d-1} \in K$ mit $f^d(v) + a_{d-1}f^{d-1}(v) + \ldots + a_1 f(v) + a_0 v = 0$. Sei $\alpha := X^d + a_{d-1}X^{d-1} + \ldots + a_0 \in K[X]$. Dann gilt

$$\alpha(f)(f^i(v)) = f^i(\alpha(f)(v)) = f^i(0) = 0$$

für alle $i \in \mathbb{N}_0$. Also ist $\mu_v \mid \alpha$. Da $v, f(v), \ldots, f^{d-1}(v)$ linear unabhängig, gilt andererseits $\deg(\mu_v) \ge d$. Dies zeigt $\mu_v = \alpha$.

Lemma 15.16. *Für $f \in \mathrm{End}(V)$ existiert ein $v \in V$ mit $\mu_v = \mu_f$.*

Beweis. Sei $V_i := \mathrm{Ker}(\gamma_i^{a_i}(f))$ für $i = 1, \ldots, k$ wie in Satz 15.12. Für $v \in V_i$ ist $\langle f^j(v) : j \in \mathbb{N}_0 \rangle \subseteq V_i$. Daher ist μ_v ein Teiler von $\gamma_i^{a_i}$. Nach der eindeutigen Primfaktorzerlegung gilt $\mu_v = \gamma_i^{b_i}$ für ein $b_i \le a_i$. Da $\gamma_i^{a_i}$ das Minimalpolynom von $f_{|V_i}$ ist, muss ein $v_i \in V_i$ mit $\mu_{v_i} = \gamma_i^{a_i}$ existieren. Wir setzen $v := v_1 + \ldots + v_k$. Da V_i f-invariant ist, gilt $\mu_v(f)(v_i) = 0$ für $i = 1, \ldots, k$. Dies zeigt $\gamma_i^{a_i} = \mu_{v_i} \mid \mu_v$ und es folgt $\mu_v = \mu_f$. $\square$

15.2 Begleitmatrizen

Bemerkung 15.17. Wir definieren das Gegenstück zu den Jordanblöcke über beliebigen Körpern.

Definition 15.18. Für $\alpha := X^n + a_{n-1}X^{n-1} + \ldots + a_0 \in K[X] \setminus K$ nennt man

$$B(\alpha) := \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -a_{n-2} \\ 0 & & 1 & -a_{n-1} \end{pmatrix} \in K^{n \times n}$$

die *Begleitmatrix* von α.

Beispiel 15.19. Offenbar ist $B_{X^n} = J_n(0)$ ein Jordanblock und $B_{X^n-1} = P_\sigma$ die Permutationsmatrix des n-Zyklus $\sigma = (1, \ldots, n)$.

Lemma 15.20. *Sei $\alpha \in K[X] \setminus K$ normiert und $k \in \mathbb{N}$. Dann gilt:*

(a) $\chi_{B(\alpha)} = \mu_{B(\alpha)} = \alpha$.

(b) $\alpha(B(\alpha^k)) \approx \operatorname{diag}\big(J_k(0), \ldots, J_k(0)\big)$.

Beweis. Sei $\alpha = X^n + a_{n-1}X^{n-1} + \ldots + a_0$.

(a) Sei $B := B(\alpha)$. Für $i = 1 \ldots, n-1$ gilt $Be_i = e_{i+1}$. Daher ist $e_1, Be_1, \ldots, B^{n-1}e_1$ die Standardbasis von K^n ist und $\deg(\mu_B) \ge \deg(\mu_{e_1}) \ge n$. Andererseits gilt

$$\alpha(B)e_i = B^{i-1}\alpha(B)e_1 = B^{i-1}(Be_n + a_{n-1}e_n + \ldots + a_1e_2 + a_0e_1) = 0$$

für $i = 1, \ldots, n$. Dies zeigt $\mu_B \mid \alpha$. Insgesamt ist $\mu_B = \alpha$. Nach Cayley-Hamilton ist $\chi_B = \mu_B$ wegen $\deg(\chi_B) = n$.

(b) Sei $d := \deg(\alpha^k) = k\deg(\alpha) = kn$. Für $A := B(\alpha^k)$ und $N := \alpha(A)$ gilt $N^k = \alpha^k(A) = 0$ nach (a). Nach Satz 14.28 besitzt N eine Jordan-Normalform J bestehend aus Blöcken der Form $J_l(0)$ mit $l \le k$. Wie in (a) gilt $e_{i+1} = Ae_i$ für $i = 1, \ldots, d-1$. Daher sind die Vektoren

$$Ne_i = A^ne_i + a_{n-1}A^{n-1}e_i + \ldots + e_i \in e_{n+i} + \langle e_1, \ldots, e_{n+i-1}\rangle$$

für $i = 1, \ldots, d-n$ linear unabhängig. Dies zeigt $\operatorname{rk}(N) \ge d-n$ und $\dim\operatorname{Ker}(N) \le n$ nach dem Homomorphiesatz. Somit besitzt J höchstens n Jordanblöcke. Diese müssen folglich alle die Größe $k \times k$ haben. □

Satz 15.21. *Sei $f \in \operatorname{End}(V)$ mit $\chi_f = \mu_f$. Dann existiert eine Basis B von V mit ${}_B[f]_B = B(\mu_f)$.*

Beweis. Sei $n := \dim V$. Nach Lemma 15.16 existiert ein $v \in V$ mit $\mu_f = \mu_v$. Für $U := \langle f^i(v) : i \in \mathbb{N}_0\rangle \le V$ gilt

$$n = \deg(\chi_f) = \deg(\mu_f) = \deg(\mu_v) \le \dim U.$$

Also ist $U = V$. Nach Bemerkung 15.15 ist $B := \{v, f(v), \ldots, f^{n-1}(v)\}$ eine Basis von V und ${}_B[f]_B = B(\mu_f)$. □

Satz 15.22. *Für jede Matrix $A \in K^{n\times n}$ gilt:*

(a) Es existieren eindeutig bestimmte normierte Polynome $\alpha_1, \ldots, \alpha_k \in K[X] \setminus K$ mit $\alpha_k \mid \alpha_{k-1} \mid \ldots \mid \alpha_1$ und

$$A \approx \operatorname{diag}\big(B(\alpha_1), \ldots, B(\alpha_k)\big) \qquad \text{(FROBENIUS-Normalform}^1\text{)}$$

Dabei ist $\mu_A = \alpha_1$ und $\chi_A = \alpha_1 \ldots \alpha_k$.

(b) Es existieren irreduzible Polynome $\gamma_1, \ldots, \gamma_s \in K[X]$ und $a_1, \ldots, a_s \in \mathbb{N}$ mit

$$A \approx \operatorname{diag}\big(B(\gamma_1^{a_1}), \ldots, B({\gamma_s}^{a_s})\big) \qquad \text{(WEIERSTRASS-Normalform)}$$

Dabei sind die Potenzen $\gamma_1^{a_1}, \ldots, \gamma_s^{a_s}$ bis auf die Reihenfolge eindeutig bestimmt und $\chi_A = \gamma_1^{a_1} \ldots \gamma_s^{a_s}$.

[1] Im Englischen: *rational canonical form*

Beweis (JACOB). Sei $V := K^n$ und $f \in \mathrm{End}(V)$ mit $[f] = A$. Nach Lemma 15.16 existiert $v \in V$ mit $\alpha_1 := \mu_v = \mu_f$. Sei $d := \deg(\alpha_1)$ und

$$U := \langle f^i(v) : i \in \mathbb{N}_0 \rangle \leq V.$$

Nach Bemerkung 15.15 ist $B(\alpha_1)$ die Darstellungsmatrix von $f_{|U}$ bzgl. $\{b_i := f^{i-1}(v) : i = 1, \ldots, d\}$. Im Fall $U = V$ sind wir fertig. Sei also $U < V$. Wir ergänzen die b_i zu einer Basis $b_1, \ldots, b_n$ von V. Sei $b_1^*, \ldots, b_n^*$ die duale Basis von V^* und $f^* \in \mathrm{End}(V^*)$ die zu f duale Abbildung. Nach Lemma 14.9 ist $U^0 \leq V^*$ f^*-invariant. Der zyklische Unterraum

$$L := \langle (f^*)^i(b_d^*) : i \in \mathbb{N}_0 \rangle \leq V^*$$

ist ebenfalls f^*-invariant. Wegen $[f^*] = A^{\mathrm{t}}$ (Satz 7.49) ist $\mu_{b_d^*} \mid \mu_{f^*} = \mu_f$. Insbesondere ist $\dim L \leq d$. Angenommen es existieren $\lambda_1, \ldots, \lambda_t \in K$ mit $t \leq d$, $\lambda_t \neq 0$ und

$$v^* := \sum_{i=1}^{t} \lambda_i (f^*)^{i-1}(b_d^*) \in L \cap U^0.$$

Nach Bemerkung 7.50 gilt $(f^*)^i = (f^i)^*$ für $i \in \mathbb{N}_0$. Es folgt der Widerspruch

$$0 = v^*(b_{d-t+1}) = b_d^*\Big(\sum_{i=1}^{t} \lambda_i f^{i-1}(b_{d-t+1})\Big) = b_d^*(\lambda_t b_d) = \lambda_t.$$

Daher ist $\{(f^*)^i(b_d^*) : i = 0, \ldots, d-1\}$ eine Basis von L und $L \cap U^0 = \{0\}$. Aus $\dim U^0 = n - \dim U = n - d$ folgt $V^* = L \oplus U^0$. Nach Lemma 7.43 ist

$$V = L_0 \oplus U,$$

wobei L_0 nach Lemma 14.9 f-invariant ist. Das Minimalpolynom von $f_1 := f_{|L_0}$ teilt α_1. Durch Induktion nach n besitzt f_1 eine Frobenius-Normalform $\mathrm{diag}(B(\alpha_2), \ldots, B(\alpha_k))$ mit $\alpha_k \mid \alpha_{k-1} \mid \ldots \mid \alpha_2 = \mu_{f_1} \mid \alpha_1$. Insgesamt existiert eine Frobenius-Normalform für f. Aus der Blockdiagonalform und Lemma 15.20 ergibt sich $\chi_A = \alpha_1 \ldots \alpha_k$.

Für die Eindeutigkeit konstruieren wir zunächst die Weierstraß-Normalform. Dafür sei $\mu_f = \gamma_1^{c_1} \ldots \gamma_l^{c_l}$ die Primfaktorzerlegung, $V_i := \mathrm{Ker}(\gamma_i^{c_i}(f))$ und $f_i := f_{|V_i}$ für $i = 1, \ldots, l$. Nach der Primärzerlegung ist

$$V = V_1 \oplus \ldots \oplus V_l$$

und $\mu_{f_i} = \gamma_i^{c_i}$. Eine Frobenius-Normalform von f_i hat somit die Form $\mathrm{diag}(B(\gamma_i^{a_1}), \ldots, B(\gamma_i^{a_s}))$ mit $1 \leq a_1, \ldots, a_s \leq c_i$. Daraus erhält man eine Weierstraß-Normalform für f. Nach Lemma 15.20 ist

$$\mathrm{diag}\big(\underbrace{J_{a_1}(0), \ldots, J_{a_1}(0)}_{\deg(\gamma_i)}, \ldots, \underbrace{J_{a_s}(0), \ldots, J_{a_s}(0)}_{\deg(\gamma_i)}\big)$$

die Jordan-Normalform von $\gamma_i(f_i)$. Daher sind die Zahlen $a_1, \ldots, a_s$ eindeutig bestimmt. Sei umgekehrt $B(\rho)$ mit $\rho \in K[X]$ ein Block einer beliebigen Weierstraß-Normalform W von f. Dann existiert ein f-invarianter Unterraum $U \leq V$, sodass ρ das Minimalpolynom von $f_{|U}$ ist. Es folgt $\rho \mid \mu_f$. Nach Folgerung 15.11 ist $\rho \mid \gamma_i^{c_i}$ für ein i und $U \leq V_i$. Die entsprechenden Blöcke von W liefern also auch eine Weierstraß-Normalform von f_i. Da die Zahlen $a_1, \ldots, a_s$ eindeutig bestimmt sind, ist Weierstraß-Normalform von f ebenfalls eindeutig bestimmt (bis auf Reihenfolge der Blöcke).

Wir betrachten schließlich die f-invariante Zerlegung $V = U_1 \oplus \ldots \oplus U_k$ einer Frobenius-Normalform von f wie oben. Sei $\alpha_i = \gamma_{i1}^{a_{i1}} \ldots \gamma_{is}^{a_{is}}$ die Primfaktorzerlegung des Minimalpolynoms von $f_{|U_i}$. In der Weierstraß-Normalform von $f_{|U_i}$ müssen dann die Blöcke $B(\gamma_{ij}^{a_{ij}})$ mit $j = 1, \ldots, s$ auftreten. Da α_i auch das charakteristische Polynom von $f_{|U_i}$ ist (Lemma 15.20), können keine weiteren Blöcke auftreten. Zusammen ergeben alle $B(\gamma_{ij}^{a_{ij}})$ die eindeutige Weierstraß-Normalform von f. Auf diese Weise sind auch $\alpha_1, \ldots, \alpha_k$ eindeutig bestimmt (siehe Beispiel 15.24). $\square$

Bemerkung 15.23.

(a) Im Gegensatz zur Jordan-Normalform sind die Blöcke $B(\alpha_i)$ der Frobenius-Normalform in einer festen Reihenfolge. Außerdem hängt die Frobenius-Normalform nicht von der Faktorisierung des Minimalpolynoms ab. Insbesondere verändert sich die Frobenius-Normalform nicht, wenn man K durch einen größeren Körper ersetzt (zum Beispiel $\mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$).

(b) Die Polynome $\alpha_1, \ldots, \alpha_k$ in der Frobenius-Normalform lassen sich alternativ durch die Folge $\beta_1 := \alpha_1/\alpha_2$, $\beta_2 := \alpha_2/\alpha_3, \ldots, \beta_k := \alpha_k$ beschreiben. Sind umgekehrt beliebige normierte Polynome $\beta_1, \ldots, \beta_k$ gegeben, so erhält man die Frobenius-Normalform

$$\operatorname{diag}(\beta_1 \ldots \beta_k, \beta_2 \ldots \beta_k, \ldots, \beta_k) \in K^{n\times n}$$

mit

$$n = \deg(\beta_1) + 2\deg(\beta_2) + \ldots + k\deg(\beta_k).$$

Auf diese Weise lassen sich die Ähnlichkeitsklassen von Matrizen systematisch aufzählen. Möchte man nur invertierbare Matrizen zählen, so müssen alle β_i ein Absolutglied $\neq 0$ haben (anderenfalls wäre X ein Teiler des charakteristischen Polynoms)

(c) Ist $q := |K| < \infty$, so gibt es genau q^d normierte Polynome vom Grad $d \geq 0$. Unter diesen haben $q^d - q^{d-1}$ ein Absolutglied $\neq 0$. Sei $n = 4$. Mit den Bezeichnungen aus (b) gibt es folgende Möglichkeiten:

$\deg(\beta_1)$	$\deg(\beta_2)$	$\deg(\beta_3)$	$\deg(\beta_4)$	Anzahl	invertierbar
0	0	0	1	q	$q-1$
1	0	1		q^2	$(q-1)^2$
2	1			q^3	$(q^2-q)(q-1)$
0	2			q^2	q^2-q
4				q^4	q^4-q^3

Insgesamt gibt es $q^4 + q^3 + 2q^2 + q$ Ähnlichkeitsklassen von Matrizen in $K^{4\times 4}$. Davon bestehen $q^4 - q$ aus invertierbaren Matrizen.

(d) Ist A nilpotent, so stimmen Jordan-, Frobenius- und Weierstraß-Normalform überein.

(e) Ein Nachteil der Frobenius-Normalform ist, dass man Begleitmatrizen weniger leicht multiplizieren kann als Jordanblöcke. Ein weiterer Nachteil ist, dass die Frobenius-Normalform F nur dann eine Diagonalmatrix ist, wenn $A = F$ eine Skalarmatrix ist (beachte $\deg(\mu_A) = \deg(\alpha_1) = 1$). Man kann am ersten Block $B(\alpha_1)$ jedoch erkennen, ob A diagonalisierbar ist (Satz 10.52).

(f) Man beachte, dass die irreduziblen Polynome γ_i in der Weierstraß-Normalform nicht unbedingt verschieden sind. Da diese Polynome aus der Primfaktorzerlegung der α_i entstehen, hängt die Weierstraß-Normalform, im Gegensatz zu Frobenius-Normalform, von K ab (in einem größeren Körper zerfallen die γ_i möglicherweise). Genau dann ist A diagonalisierbar, wenn die Weierstraß-Normalform eine Diagonalmatrix ist. In diesem Fall stimmt die Weierstraß-Normalform also mit der Jordan-Normalform überein.

Beispiel 15.24.

(a) Seien $\alpha, \beta, \gamma \in K[X]$ irreduzibel. Die Umrechnung zwischen Frobenius-Normalform und Weierstraß-Normalform geht wie folgt:

$$\operatorname{diag}\big(B(\alpha^2\beta^3\gamma), B(\alpha^2\beta^2), B(\alpha)\big) \approx \operatorname{diag}\big(B(\alpha), B(\alpha^2), B(\alpha^2), B(\beta^2), B(\beta^3), B(\gamma)\big).$$

(b) Die Berechnung der Frobenius/Weierstraß-Normalform (von Hand) ist naturgemäß sehr aufwendig. Unser Beweis von Lemma 15.16 ist zum Beispiel nicht konstruktiv, aber es gibt entsprechende Algorithmen.[2] Oft kann man Ad-hoc-Argumente benutzen. Die Matrix

$$A := \begin{pmatrix} 7 & -1 & -3 \\ 30 & -4 & -15 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{Q}^{3\times 3}$$

hat das charakteristisches Polynom

$$\chi_A = \big((X-7)(X+4)+30\big)(X-1) = (X^2-3X+2)(X-1) = (X-1)^2(X-2).$$

Wegen

$$(A-1_2)(A-2\cdot 1_2) = \begin{pmatrix} 6 & -1 & -3 \\ 30 & -5 & -15 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 5 & -1 & -3 \\ 30 & -6 & -15 \\ 0 & 0 & -1 \end{pmatrix} = 0$$

ist $\mu_A = (X-1)(X-2)$. Also ist

$$\begin{pmatrix} B(X-1) & 0 \\ 0 & B(\mu) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 1 & 3 \end{pmatrix}$$

die Frobenius-Normalform von A. Die Weierstraß-Normalform ist hingegen $\operatorname{diag}(1,1,2)$.

(c) Wir beschreiben die Ähnlichkeitsklassen von Matrizen in $\mathbb{F}_2^{3\times 3}$ mit der Folge $(\beta_1,\ldots,\beta_k)$ aus Bemerkung 15.23:

$$\begin{array}{ccc}
(1,1,X) = 0_3, & (1,1,X+1) = 1_3, & (X,X), \\
(X,X+1) \approx \operatorname{diag}(1,1,0), & (X+1,X) \approx \operatorname{diag}(1,0,0), & (X+1,X+1) = P_{(1,2)}, \\
(X^3) = J_3(0), & (X^3+1) = P_{(1,2,3)}, & (X^3+X), \\
(X^3+X+1), & (X^3+X^2), & (X^3+X^2+1), \\
(X^3+X^2+X), & (X^3+X^2+X+1). &
\end{array}$$

Also gibt es genau 14 Ähnlichkeitsklassen, von denen sechs zu invertierbaren Matrizen gehören (welche?).

Satz 15.25. *Für alle $A \in K^{n\times n}$ gilt $A \approx A^{\mathrm{t}}$.*

Beweis. Nach der Frobenius-Normalform kann man $A = B(\alpha)$ für ein normiertes Polynom $\alpha \in K[X]\setminus K$ annehmen. Nach Lemma 15.20 und Aufgabe 35 gilt $\chi_{A^{\mathrm{t}}} = \chi_A = \mu_A = \mu_{A^{\mathrm{t}}}$. Nach Satz 15.21 ist $A^{\mathrm{t}} \approx A$. □

Bemerkung 15.26. Einen direkten Beweis von Satz 15.25 findet man in Aufgabe 58.

[2] siehe [M. Geck, *On Jacob's construction of the rational canonical form of a matrix*, Electron. J. Linear Algebra 36 (2020), 177–182]

15.3 Zentralisatoren

Definition 15.27. Für $f \in \operatorname{End}(V)$ und $A \in K^{n\times n}$ sei

$$\begin{aligned}\mathrm{C}(f) &:= \{g \in \operatorname{End}(V) : f\circ g = g\circ f\} \subseteq \operatorname{End}(V),\\ \mathrm{C}(A) &:= \{B \in K^{n\times n} : AB = BA\} \subseteq K^{n\times n}\end{aligned}$$

der *Zentralisator* von f bzw. A.

Bemerkung 15.28. Offenbar sind Zentralisatoren Unterräume. Für $C, D \in \mathrm{C}(A)$ ist außerdem $CD \in \mathrm{C}(A)$. Für $S \in \mathrm{GL}(n,K)$ gilt $\mathrm{C}(SAS^{-1}) = S\,\mathrm{C}(A)S^{-1}$. Daher ist $\dim \mathrm{C}(A)$ eine Invariante der Ähnlichkeitsklasse von A. Es gilt

$$\{\alpha(A) : \alpha \in K[X]\} = \langle A^i : i \in \mathbb{N}_0\rangle \subseteq \mathrm{C}(A).$$

Wir untersuchen, wann Gleichheit gilt.

Beispiel 15.29.

(a) Sei $A \in K^{n\times n}$ eine Diagonalmatrix. Nach Permutation der Basisvektoren können wir $A = \operatorname{diag}(\lambda_1 1_{d_1},\ldots,\lambda_k 1_{d_k})$ mit paarweise verschiedenen $\lambda_1,\ldots,\lambda_k \in K$ annehmen. Nach Aufgabe 17 gilt $\mathrm{C}(A) = \{\operatorname{diag}(A_1,\ldots,A_k) : \forall i : A_i \in K^{d_i\times d_i}\}$. Insbesondere ist $\dim \mathrm{C}(A) = \sum_{i=1}^k d_i^2$. Dies wird in Aufgabe 61 und Bemerkung 15.33 verallgemeinert.

(b) Sei $A \in K^{d\times d}$ und $D := \operatorname{diag}(A,\ldots,A) \in K^{dk\times dk}$. Sei $B = (B_{ij}) \in K^{dk\times dk}$ mit Blöcken $B_{ij} \in K^{d\times d}$. Dann gilt

$$\begin{pmatrix} AB_{11} & \cdots & AB_{1k}\\ \vdots & & \vdots\\ AB_{k1} & \cdots & AB_{kk}\end{pmatrix} = DB = BD = \begin{pmatrix} B_{11}A & \cdots & B_{1k}A\\ \vdots & & \vdots\\ B_{k1}A & \cdots & B_{kk}A\end{pmatrix} \iff \forall i,j : B_{ij} \in \mathrm{C}(A).$$

Insbesondere ist $\dim \mathrm{C}(D) = k^2 \dim \mathrm{C}(A)$.

(c) Nach Aufgabe 53 gilt $\dim \mathrm{C}(J_n(\lambda)) = n$ für alle $\lambda \in K$. Dies wird in Folgerung 15.34 verallgemeinert.

Lemma 15.30. *Seien V, W Vektorräume und $f \in \operatorname{End}(V)$, $g \in \operatorname{End}(W)$ mit $\chi_f = \mu_f \mid \chi_g = \mu_g$. Dann ist*

$$H := \{h \in \operatorname{Hom}(V,W) : h\circ f = g\circ h\}$$

ein Vektorraum mit $\dim H = \dim V$.

Beweis. Sei $d := \dim V$ und $e := \dim W$. Wegen $\chi_f = \mu_f$ und $\chi_g = \mu_g$ existieren $v \in V$ und $w \in W$ mit

$$V = \langle f^i(v) : i = 0,\ldots,d-1\rangle, \qquad W = \langle g^i(w) : i = 0,\ldots,e-1\rangle.$$

Sei $h \in H$. Dann existiert genau ein Polynom $\alpha \in K[X]$ mit $h(v) = \alpha(g)(w)$ und $\deg(\alpha) < e$. Für $i = 1,\ldots,d-1$ folgt $h(f^i(v)) = g^i(h(v))$. Also ist h durch α eindeutig bestimmt. Nach Lemma 10.15 ist die Abbildung $H \to K[X]$, $h \mapsto \alpha$ linear und injektiv. Wegen

$$0 = h\big(\mu_f(f)(v)\big) = \mu_f(g)(h(v)) = \mu_f(g)\big(\alpha(g)(w)\big) = (\alpha\mu_f)(g)(w)$$

gilt außerdem $\mu_g \mid \alpha\mu_f$, d. h. $\tau := \frac{\mu_g}{\mu_f} \mid \alpha$. Sei $P_d \subseteq K[X]$ der Vektorraum aller Polynome vom Grad $< d$. Dann ist

$$\Gamma\colon H \to P_d, \qquad h \mapsto \alpha/\tau$$

eine injektive lineare Abbildung.

Sei umgekehrt $\beta \in P_d$ gegeben und $\alpha := \beta\tau$. Dann gilt $\mu_g \mid \alpha\mu_f$. Wir definieren $h \in \operatorname{Hom}(V, W)$ durch $h(v) = \alpha(g)(w)$ und $h(f^i(v)) = g^i(h(v))$ für $i = 1, \ldots, d-1$. Für $i = 0, \ldots, d-2$ gilt dann

$$(h \circ f)(f^i(v)) = g^{i+1}(h(v)) = g(g^i(h(v))) = (g \circ h)(f^i(v)).$$

Sei $\mu_f = X^d + a_{d-1}X^{d-1} + \ldots + a_0$. Wegen $\mu_f(f) = 0$ gilt

$$\begin{aligned}(h \circ f)(f^{d-1}(v)) &= h(f^d(v)) = h\big(-a_{d-1}f^{d-1}(v) - \ldots - a_0 v\big) = -a_{d-1}g^{d-1}(h(v)) - \ldots - a_0 h(v)\\ &= g^d(h(v)) - \mu_f(g)(h(v)) = g^d(h(v)) - (\mu_f\alpha)(g)(w) = g^d(h(v)) = (g \circ h)(f^{d-1}(v)).\end{aligned}$$

Dies zeigt $h \in H$ mit $\Gamma(h) = \beta$. Also ist Γ ein Isomorphismus und $\dim H = \dim P_d = d = \dim V$. □

Beispiel 15.31. Sei $V = K$ und $f = \lambda\,\mathrm{id}_V$ für ein $\lambda \in K$. Für $h \in H$ gilt $\lambda h(1) = h(f(1)) = g(h(1))$, d. h. $h(1)$ ist ein Eigenvektor von g zum Eigenwert λ. Außerdem ist $H \to E_\lambda(g)$, $h \mapsto h(1)$ ein Isomorphismus. Wegen $\dim H = 1$ hat λ geometrische Vielfachheit 1. Das kann man natürlich auch direkt aus der Bedingung $\chi_f = X - \lambda \mid \chi_g = \mu_g$ ableiten.

Satz 15.32 (Frobenius). *Sei $A \in K^{n\times n}$ mit Frobenius-Normalform* $\operatorname{diag}(B(\alpha_1), \ldots, B(\alpha_k))$. *Dann gilt*

$$\boxed{\dim \mathrm{C}(A) = \sum_{i=1}^{k}(2i-1)\deg(\alpha_i).}$$

Insbesondere ist $\dim \mathrm{C}(A) \geq n$ *mit Gleichheit genau dann, wenn* $\chi_A = \mu_A$.

Beweis. Sei $d_i := \deg(\alpha_i)$ und $A_i := B(\alpha_i) \in K^{d_i\times d_i}$ für $i = 1, \ldots, k$. Nach Bemerkung 15.28 können wir $A = \operatorname{diag}(A_1, \ldots, A_k)$ annehmen. Sei $C = (C_{ij}) \in K^{n\times n}$ eine Blockmatrix mit $C_{ij} \in K^{d_i\times d_j}$ für $1 \leq i, j \leq k$. Dann gilt

$$C \in \mathrm{C}(A) \iff CA = AC \iff \forall i, j : C_{ij}A_j = A_iC_{ij}.$$

Für $i \leq j$ gilt $\chi_{A_j} = \mu_{A_j} = \alpha_j \mid \alpha_i = \chi_{A_i} = \mu_{A_i}$ nach Lemma 15.20. Nach Lemma 15.30 gibt es d_j linear unabhängige Möglichkeiten für die Wahl von C_{ij}. Im Fall $i > j$ können wir

$$C_{ij}^{\mathrm{t}}A_i^{\mathrm{t}} = (A_iC_{ij})^{\mathrm{t}} = (C_{ij}A_j)^{\mathrm{t}} = A_j^{\mathrm{t}}C_{ij}^{\mathrm{t}}$$

betrachten. Wegen $\chi_{A_i^{\mathrm{t}}} = \alpha_i = \mu_{A_i^{\mathrm{t}}}$ (Aufgabe 35) gibt es d_i linear unabhängige Möglichkeiten für C_{ij}^{t} (und damit auch für C_{ij}). Da die Blöcke C_{ij} in C unabhängig voneinander gewählt werden können, erhält man

$$\dim \mathrm{C}(A) = \sum_{i,j=1}^{k}\min\{d_i, d_j\} = d_1 + 3d_2 + 5d_3 + \ldots + (2k-1)d_k \geq d_1 + \ldots + d_k = n$$

mit Gleichheit genau dann, wenn $k = 1$ und $\mu_A = \alpha_1 = \chi_A$ (Satz 15.22). □

Bemerkung 15.33. In Bemerkung 15.23 haben wir die Frobenius-Normalform durch die Polynome $\beta_1 := \alpha_1/\alpha_2,\ \beta_2 := \alpha_2/\alpha_3, \ldots, \beta_k := \alpha_k$ beschrieben. Es gilt $\deg(\alpha_i) = \deg(\beta_i\beta_{i+1}\ldots\beta_k) = \sum_{j=i}^k \deg(\beta_j)$ und $\sum_{i=1}^m (2i-1) = m^2$ nach Aufgabe 4. Dies zeigt

$$\boxed{\dim \mathrm{C}(A) = \sum_{l=1}^k l^2 \deg(\beta_l).}$$

Folgerung 15.34. *Sei $A \in K^{n\times n}$ mit $\chi_A = \mu_A$. Dann gilt $\mathrm{C}(A) = \langle A^i : i = 0, \ldots, n-1\rangle$.*

Beweis. Aus $\deg(\mu_A) = \deg(\chi_A) = n$ folgt $\dim \mathrm{C}(A) \ge \dim\langle A^i : i \in \mathbb{N}_0\rangle = n$. Nach Frobenius ist andererseits $\dim \mathrm{C}(A) = n$. □

15.4 Zerfällungskörper

Satz 15.35. *Für $n, k \in \mathbb{N}$ und $\lambda \in \mathbb{C}$ existieren genau k Matrizen $W \in \mathbb{C}^{n\times n}$ mit $W^k = J_n(\lambda)$.*

Beweis. Sei $J := J_n(\lambda)$. Nach Satz 14.42 existiert zumindest eine k-te Wurzel W mit $W^k = J$. Für jede k-te Einheitswurzel $\zeta \in \mathbb{C}$ gilt auch $(\zeta W)^k = J$. Nach Lemma 11.27 existieren also mindestens k Wurzeln von J. Nach Folgerung 15.34 gilt $W \in \mathrm{C}(J) = \langle J^i : i \in \mathbb{N}_0\rangle$. Insbesondere ist W eine untere Dreiecksmatrix mit Diagonale $(\mu, \ldots, \mu)$ für eine k-te Einheitswurzel μ (vgl. Aufgabe 53). Indem man W durch $\mu^{-1}W$ ersetzt, kann man $\mu = 1$ annehmen.

Sei umgekehrt $A \in \mathbb{C}^{n\times n}$ mit $A^k = J$. Dann ist auch A eine untere Dreiecksmatrix mit Diagonale $(\zeta, \ldots, \zeta)$ für eine k-te Einheitswurzel ζ. Da $B := \zeta W \in \mathrm{C}(J)$ ein Polynom in J ist, sind A und B vertauschbar. Es folgt

$$0 = A^k - B^k = (A-B)(\underbrace{A^{k-1} + A^{k-2}B + \ldots + AB^{k-2} + B^{k-1}}_{=:C}).$$

Offenbar ist C eine untere Dreiecksmatrix mit Hauptdiagonale $k\zeta^{k-1}(1, \ldots, 1) \ne 0$. Insbesondere ist C invertierbar und $A - B = 0 \cdot C^{-1} = 0$, d. h. $A = B$. □

Satz 15.36 (Schurs Lemma)**.** *Sei $f \in \mathrm{End}(V)$. Genau dann ist χ_f irreduzibel, wenn $\{0\}$ und V die einzigen f-invarianten Unterräume sind.*

Beweis. Seien $\{0\}$ und V die einzigen f-invarianten Unterräume von V. Nach der Frobenius-Normalform ist $\chi_f = \mu_f$. Sei γ ein irreduzibler Teiler von μ_f. Dann ist $U := \mathrm{Ker}(\gamma(f)) \le V$ ein f-invarianter Unterraum. Für $v \in V$ mit $\mu_v = \mu_f$ (Lemma 15.16) gilt $0 \ne \frac{\mu_f}{\gamma}(f)(v) \in U$. Dies zeigt $U = V$ und $\chi_f = \mu_f = \gamma$ ist irreduzibel.

Sei umgekehrt χ_f irreduzibel und $U < V$ f-invariant. Wir ergänzen eine Basis B_1 von U zu einer Basis B von V. Dann gilt

$$_B[f]_B = \begin{pmatrix} {}_{B_1}[f_{|U}]_{B_1} & * \\ 0 & * \end{pmatrix}.$$

Dies zeigt $\chi_{f_{|U}} \mid \chi_f$ und $\chi_{f_{|U}} \in K$. Es folgt $U = \{0\}$. □

Satz 15.37. *Für jedes irreduzible Polynom $\gamma \in K[X]$ ist $L := \mathrm{C}(B(\gamma))$ ein Körper mit $\dim_K L = \deg(\gamma)$.*

Beweis. Sei $n := \deg(\gamma)$, $V = K^n$ und $f \in \operatorname{End}(V)$ mit $A := B(\gamma) = [f]$. Nach Lemma 15.20 ist $\chi_f = \mu_f = \gamma$. Nach Folgerung 15.34 ist $L = \langle A^i : i \in \mathbb{N}_0 \rangle \subseteq K^{n\times n}$. Insbesondere ist die (Matrizen-)Multiplikation in L kommutativ. Für $g \in \mathrm{C}(f) \setminus \{0\}$ ist $\operatorname{Ker}(g) \le V$ wie üblich ein f-invarianter Unterraum. Nach Schurs Lemma gilt $\operatorname{Ker}(g) = \{0\}$ und $g \in \mathrm{GL}(V)$. Mit g ist auch $g^{-1} \in \mathrm{C}(f)$. Also besitzt jedes nicht-triviale Element in L ein Inverses bzgl. Multiplikation. Die verbleibenden Körperaxiome für $\mathrm{C}(A)$ folgen aus den Rechenregeln in $K^{n\times n}$. Nach Folgerung 15.34 ist $\dim L = n$. □

Beispiel 15.38.

(a) Bekanntlich ist $\gamma := X^2 + 1 \in \mathbb{R}[X]$ irreduzibel. Sei $A := B(\gamma) = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$. Nach Satz 15.37 ist $\mathrm{C}(A) = \left\{\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) : a, b \in \mathbb{R}\right\}$ ein Körper. Tatsächlich stimmt $\mathrm{C}(A)$ mit der Konstruktion von $\mathbb{C}$ im Beweis von Lemma 11.24 überein.

(b) Für $\gamma = X^2 - 2 \in \mathbb{Q}[X]$ ist $\mathrm{C}(B(\gamma)) = \mathbb{Q}(\sqrt{2})$ der Körper aus Aufgabe 14.

(c) Für ist $\gamma := X^2 + X + 1 \in \mathbb{F}_2[X]$ ist

$$\mathrm{C}(B(\gamma)) = \left\{0_2, 1_2, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\right\}$$

ein Körper mit vier Elementen.

Satz 15.39. *Für jedes normierte Polynom $\alpha \in K[X]$ existiert ein Körper $L \supseteq K$, sodass α in $L[X]$ in Linearfaktoren zerfällt.*

Beweis. Im Fall $\deg(\alpha) = 1$ ist α selbst ein Linearfaktor. Sei also $\deg(\alpha) \ge 2$. Sei γ ein irreduzibler Teiler von α mit $d := \deg(\gamma)$. Nach Satz 15.37 ist $L_1 := \mathrm{C}(B(\gamma_1))$ ein Körper. Wir können K mit den Skalarmatrizen $\{\lambda 1_d \in L_1 : \lambda \in K\}$ identifizieren (die Rechenregeln für Skalarmatrizen entsprechen denen in K). Für $x := B(\gamma) \in L_1$ gilt $\gamma(x) = 0$, d. h. x ist eine Nullstelle von γ. Nach Lemma 10.22 gilt $\alpha = (X - x)\beta$ für ein $\beta \in L_1[X]$ mit $\deg(\beta) = \deg(\alpha) - 1$. Durch Induktion nach $\deg(\alpha)$ existiert ein Körper $L \supseteq L_1$, in dem β in Linearfaktoren zerfällt. Offenbar zerfällt auch α in $L[X]$ in Linearfaktoren. □

Definition 15.40. In der Situation von Satz 15.39 nennt man L einen *Zerfällungskörper* von α (nicht eindeutig bestimmt).

16 Die Jordan-Chevalley-Zerlegung

16.1 Der chinesische Restsatz

Bemerkung 16.1. Wir erweitern unsere Kenntnisse über Polynome. Mit diesem Werkzeug zerlegen wir eine Matrix auf eindeutige Weise in einen diagonalisierbaren Teil (über einem Zerfällungskörper) und einen nilpotenten Teil.

Definition 16.2. Seien $\alpha, \beta, \delta \in K[X]$ mit $\delta \neq 0$. Wir sagen α ist *kongruent* zu β *modulo* δ, falls $\delta \mid \alpha - \beta$. Ggf. schreiben wir $\alpha \equiv \beta \pmod{\delta}$.

Lemma 16.3. *Die Kongruenz modulo $\delta \in K[X] \setminus \{0\}$ ist eine Äquivalenzrelation auf $K[X]$. Für $\alpha_i, \beta_i \in K[X]$ mit $\alpha_i \equiv \beta_i \pmod{\delta}$ $(i = 1, 2)$ gilt $\alpha_1 + \alpha_2 \equiv \beta_1 + \beta_2 \pmod{\delta}$ und $\alpha_1\alpha_2 \equiv \beta_1\beta_2 \pmod{\delta}$.*

Beweis. Wegen $\delta \mid 0 = \alpha - \alpha$ ist $\alpha \equiv \alpha \pmod{\delta}$. Aus $\alpha \equiv \beta \pmod{\delta}$ folgt $\beta \equiv \alpha \pmod{\delta}$. Sei nun $\alpha \equiv \beta \pmod{\delta}$ und $\beta \equiv \gamma \pmod{\delta}$. Dann gilt $\delta \mid (\alpha - \beta) + (\beta - \gamma) = \alpha - \gamma$ und $\alpha \equiv \gamma \pmod{\delta}$. Daher ist $\equiv$ eine Äquivalenzrelation. Aus $\alpha_i \equiv \beta_i \pmod{\delta}$ folgt

$$\begin{aligned}
&\delta \mid (\alpha_1 - \beta_1) + (\alpha_2 - \beta_2) = (\alpha_1 + \alpha_2) - (\beta_1 + \beta_2), \\
&\delta \mid (\alpha_1 - \beta_1)\alpha_2 + (\alpha_2 - \beta_2)\beta_1 = \alpha_1\alpha_2 - \beta_1\beta_2.
\end{aligned}$$

Dies zeigt die zweite Behauptung. $\square$

Beispiel 16.4.

(a) Für $\delta \in K^\times$ ist $\alpha \equiv \beta \pmod{\delta}$ für alle $\alpha, \beta \in K[X]$, d. h. $\equiv$ ist die triviale Relation.

(b) Es gilt $\alpha \equiv \beta \pmod{X}$ genau dann, wenn α und β das gleiche Absolutglied haben.

(c) Für alle $\alpha \in K[X]$ existiert ein $\beta \in K[X]$ mit $\alpha \equiv \beta \pmod{\delta}$ und $\deg(\beta) < \deg(\delta)$. Dies ergibt sich aus der Division mit Rest.

(d) Lemma 16.3 vereinfacht viele Teilbarkeitsbetrachtungen: Um zu prüfen, ob

$$\alpha = (X^2 + X + 2)^4 + (X^5 - 1)(X^3 + X) \in \mathbb{Q}[X]$$

durch $\delta = X + 1 \in \mathbb{Q}[X]$ teilbar ist, kann man die Kongruenz mit jedem einzelnen Term durchführen. Wegen

$$\begin{aligned}
X^2 + X + 2 &= X(X + 1) + 2 \equiv 2 \pmod{\delta}, \\
X^5 - 1 &= (X + 1)(X^4 - X^3 + X^2 - X + 1) - 2 \equiv -2 \pmod{\delta}, \\
X^3 + X + 10 &= (X + 1)(X^2 - X + 2) + 8 \equiv 8 \pmod{\delta}
\end{aligned}$$

gilt $\alpha \equiv 2^4 + (-2)8 \equiv 0 \pmod{\delta}$, d. h. $\delta \mid \alpha$.

Satz 16.5 (Chinesischer Restsatz). *Seien* $\alpha_1, \ldots, \alpha_n \in K[X] \backslash \{0\}$ *paarweise teilerfremd und* $\beta_1, \ldots, \beta_n \in K[X]$ *beliebig. Dann existiert ein* $\gamma \in K[X]$ *mit* $\gamma \equiv \beta_i \pmod{\alpha_i}$ *für* $i = 1, \ldots, n$.

Beweis. Für $i = 1, \ldots, n$ sei $\alpha_i' := \prod_{j \neq i} \alpha_j$. Nach Folgerung 15.11 muss jeder irreduzible Teiler von α_i' ein α_j mit $j \neq i$ teilen. Da α_i und α_j teilerfremd sind, müssen auch α_i und α_i' teilerfremd sein. Nach Bézout existieren $\gamma_1, \ldots, \gamma_n \in K[X]$ mit $\alpha_i'\gamma_i \equiv 1 \pmod{\alpha_i}$. Sei

$$\gamma := \sum_{i=1}^{n} \alpha_i'\beta_i\gamma_i.$$

Dann gilt $\gamma \equiv \alpha_i'\beta_i\gamma_i \equiv \beta_i \pmod{\alpha_i}$ für $i = 1, \ldots, n$. □

Bemerkung 16.6. Mit γ erfüllt auch $\gamma + \rho \prod_{i=1}^{n} \alpha_i$ für alle $\rho \in K[X]$ die Kongruenzen aus dem chinesischen Restsatz. Daher gibt es unendlich viele Lösungen.

Beispiel 16.7. Da die Polynome $X^2 - 2, X^2 + 1 \in \mathbb{Q}[X]$ keine gemeinsamen Nullstellen haben, sind sie teilerfremd. Gesucht sei $\gamma \in \mathbb{Q}[X]$ mit

$$\gamma \equiv X \pmod{X^2 - 2}, \qquad \gamma \equiv 3 \pmod{X^2 + 1}.$$

Bézout liefert $\frac{1}{3}(X^2 + 1) \equiv 1 \pmod{X^2 - 2}$ und $-\frac{1}{3}(X^2 - 2) \equiv 1 \pmod{X^2 + 1}$ (im Zweifel muss man den euklidischen Algorithmus bemühen). Wie im Beweis von Satz 16.5 ist

$$\gamma = \frac{1}{3}(X^2 + 1)X - \frac{1}{3}(X^2 - 2)3 = \frac{1}{3}X^3 - X^2 + \frac{1}{3}X + 2$$

eine Lösung der Kongruenzen.

Definition 16.8.

- Für $\alpha = \sum_{k=0}^{\infty} a_k X^k \in K[X]$ definieren wir die (formale) *Ableitung*

$$\alpha' = \sum_{k=1}^{\infty} k a_k X^{k-1}$$

  von α wie in der Analysis.

- Ein irreduzibles Polynom α heißt *separabel*, wenn $\alpha' \neq 0$. Ein beliebiges Polynom α heißt *separabel*, wenn alle irreduziblen Teiler von α separabel sind (das Gegenteil ist *inseparabel*).

Beispiel 16.9.

(a) Die konstanten Polynome sind separabel, da sie keine irreduziblen Teiler haben.

(b) Ist die Abbildung $\mathbb{N} \to K$, $n \mapsto n \cdot 1_K$ injektiv, so gilt $\alpha' \neq 0$ für alle $\alpha \in K[X] \setminus K$. Insbesondere ist jedes Polynom in $\mathbb{C}[X]$ separabel.

(c) Über endlichen Körpern ist die Bedingung $\alpha' \neq 0$ weniger offensichtlich. Zum Beispiel gilt $(X^2)' = 0$ in $\mathbb{F}_2[X]$. Sei allgemein $\alpha = \sum a_k X^k \in \mathbb{F}_2[X]$ mit $\alpha' = \sum k a_k X^{k-1} = 0$. Für ungerade k folgt $a_k X^{k-1} = k a_k X^{k-1} = 0$, also $a_k = 0$. Dies zeigt

$$\alpha = a_0 + a_2 X^2 + \ldots + a_{2n} X^{2n}$$

für ein $n \in \mathbb{N}$. Wegen $(x+y)^2 = x^2 + 2xy + y^2 = x + y$ in $\mathbb{F}_2$ erhält man induktiv

$$\alpha = (a_0 + a_2 X + \ldots + a_{2n} X^n)^2.$$

Insbesondere ist α reduzibel (d. h. nicht irreduzibel). Damit ist jedes Polynom in $\mathbb{F}_2[X]$ separabel.

(d) Man konstruiert $\mathbb{Q}$ aus $\mathbb{Z}$, indem man Brüche einführt. Auf die gleiche Weise kann man den Körper der *rationalen Funktionen*

$$K(X) := \left\{ \frac{\alpha}{\beta} : \alpha, \beta \in K[X], \beta \neq 0 \right\}$$

aus $K[X]$ gewinnen, indem man Brüche von Polynomen einführt. Sei speziell $K := \mathbb{F}_2(X)$. Man kann zeigen, dass $\alpha := X^2 + Y \in K[Y]$ irreduzibel und inseparabel ist.

Lemma 16.10. *Für $\alpha, \beta \in K[X]$ gilt*

$$(\alpha + \beta)' = \alpha' + \beta', \qquad \text{(Summenregel)}$$
$$(\alpha\beta)' = \alpha'\beta + \alpha\beta'. \qquad \text{(Produktregel)}$$

Beweis. Für $\alpha = \sum a_k X^k$ und $\beta = \sum b_k X^k$ ist

$$\begin{aligned}(\alpha + \beta)' &= \Big(\sum (a_k + b_k) X^k\Big)' = \sum k(a_k + b_k) X^{k-1} \\ &= \sum k a_k X^{k-1} + \sum k b_k X^{k-1} = \alpha' + \beta'.\end{aligned}$$

Daraus folgt

$$\begin{aligned}(\alpha\beta)' &= \Big(\sum_{k,l=0}^{\infty} a_k b_l X^{k+l}\Big)' = \sum_{k,l} a_k b_l \big(X^{k+l}\big)' = \sum_{k,l} (k+l) a_k b_l X^{k+l-1} \\ &= \sum_{k,l} k a_k b_l X^{k-1+l} + \sum_{k,l} l a_k b_l X^{l-1+k} = \alpha'\beta + \alpha\beta'.\end{aligned}$$ □

Lemma 16.11. *Sei $\gamma \in K[X]$ irreduzibel. Genau dann ist γ separabel, wenn γ keine mehrfachen Nullstellen in einem Zerfällungskörper besitzt.*

Beweis. Sei $\lambda \in L$ eine Nullstelle von γ in einem Zerfällungskörper L. Dann existiert ein $\rho \in L[X]$ mit $\gamma = (X - \lambda)\rho$. Aus der Produktregel folgt

$$\gamma' = \rho + (X - \lambda)\rho'$$

und $\gamma'(\lambda) = \rho(\lambda)$. Ist γ inseparabel, so ist $\rho(\lambda) = \gamma'(\lambda) = 0$, d. h. λ ist eine mehrfache Nullstelle. Sei umgekehrt γ separabel, d. h. $\gamma' \neq 0$. Dann gilt $0 \leq \deg(\gamma') < \deg(\gamma)$. Da γ irreduzibel ist, sind γ' und γ teilerfremd. Nach Bézout existieren $\alpha, \beta \in K[X]$ mit $\alpha\gamma + \beta\gamma' = 1$. Es folgt $\beta(\lambda)\gamma'(\lambda) = 1$ und $\rho(\lambda) = \gamma'(\lambda) \neq 0$. Also ist λ eine einfache Nullstelle. Da λ beliebig war, sind alle Nullstellen einfach. □

16.2 Separable und halbeinfache Abbildungen

Definition 16.12. Wir nennen $f \in \mathrm{End}(V)$ (bzw. $A \in K^{n\times n}$)

- *separabel*, falls μ_f (bzw. μ_A) separabel ist.
- *halbeinfach*, falls μ_f (bzw. μ_A) in paarweise verschiedene irreduzible Polynome zerfällt.

Beispiel 16.13.

(a) Über $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ und $\mathbb{F}_2$ sind nach Beispiel 16.9 alle Endomorphismen separabel.

(b) Sei $K = \mathbb{C}$. Nach dem Fundamentalsatz der Algebra und Satz 10.52 ist $f \in \mathrm{End}(V)$ genau dann halbeinfach, wenn f diagonalisierbar. Dies wird in Lemma 16.15 verallgemeinert. Andererseits sind $B(X^2-2) \in \mathbb{Q}^{2\times 2}$ und $B(X^2+X+1) \in \mathbb{F}_2^{2\times 2}$ halbeinfach, aber nicht diagonalisierbar.

Satz 16.14. *Genau dann ist $f \in \mathrm{End}(V)$ halbeinfach, wenn jeder f-invariante Unterraum $U \le V$ ein f-invariantes Komplement besitzt.*

Beweis. Sei f halbeinfach und $\mu_f = \gamma_1 \ldots \gamma_k$ mit paarweise verschiedenen irreduziblen Polynomen $\gamma_1, \ldots, \gamma_k$. Die Weierstraß-Normalform von f liefert eine Zerlegung $V = V_1 \oplus \ldots \oplus V_s$ in f-invariante Unterräume, sodass die Darstellungsmatrix von f auf V_i durch $B(\gamma_j)$ für ein $1 \le j \le k$ gegeben ist (beachte $k \le s$). Nach Lemma 15.20 und Schurs Lemma besitzt V_i keine echten nicht-trivialen f-invarianten Unterräume für $i = 1, \ldots, s$. Sei $U \le V$ f-invariant. Sei $W \le V$ f-invariant mit $U \cap W = \{0\}$, sodass $\dim W$ möglichst groß ist (notfalls $W = \{0\}$). Angenommen es gilt $U \oplus W < V$. Dann existiert ein i mit $V_i \nsubseteq U + W$. Offenbar ist

$$L := V_i \cap (U + W) < V_i$$

f-invariant. Aus Schurs Lemma folgt $L = \{0\}$. Sei $u = w + v \in U \cap (W + V_i)$ mit $w \in W$ und $v \in V_i$. Dann gilt $v = u - w \in V_i \cap (U + W) = \{0\}$ und $u = w \in U \cap W = \{0\}$. Also ist $U \cap (W + V_i) = \{0\}$ im Widerspruch zur Wahl von W. Dies zeigt $V = U \oplus W$.

Nehmen wir umgekehrt an, dass jeder f-invariante Unterraum von V ein f-invariantes Komplement besitzt. Wir nehmen indirekt $\mu_f = \gamma^2\delta$ für ein irreduzibles Polynom γ an. Nach Voraussetzung besitzt $U := \mathrm{Ker}(\gamma(f)) \le V$ ein f-invariantes Komplement W. Für $w \in W$ gilt

$$(\gamma\delta)(f)(w) \in U \cap W = \{0\}.$$

Wegen $(\gamma\delta)(f)(U) = \{0\}$ gilt sogar $(\gamma\delta)(f)(v) = 0$ für alle $v \in V$. Dann wäre aber $\mu_f \mid \gamma\delta$. Widerspruch. □

Lemma 16.15. *Sei $A \in K^{n\times n}$ separabel. Genau dann ist A halbeinfach, wenn ein Körper $L \supseteq K$ existiert, sodass A in $L^{n\times n}$ diagonalisierbar ist.*

Beweis. Sei A halbeinfach und $\mu_A = \gamma_1 \ldots \gamma_k$ mit paarweise verschiedenen (separablen) irreduziblen Polynomen $\gamma_1, \ldots, \gamma_k$. Sei L ein Zerfällungskörper von μ_A. Nach Lemma 16.11 hat jedes γ_i keine mehrfachen Nullstellen in L. Für $i \ne j$ existieren $\alpha, \beta \in K[X]$ mit $\alpha\gamma_i + \beta\gamma_j = 1$ nach Bézout. Daher haben γ_i und γ_j keine gemeinsamen Nullstellen in L. Insgesamt zerfällt μ_A in paarweise verschiedene Linearfaktoren in $L[X]$. Nach Satz 10.52 ist A in $L^{n\times n}$ diagonalisierbar. Ist A nicht halbeinfach, so hat μ_A offensichtlich mehrfache Nullstellen in $L[X]$. Damit kann A nicht diagonalisierbar sein. □

Lemma 16.16. *Sind $A, B \in K^{n\times n}$ vertauschbar, separabel und halbeinfach, so ist auch $A+B$ halbeinfach.*

Beweis. Sei L ein Zerfällungskörper von $\mu_A\mu_B$. Nach Lemma 16.15 und Lemma 14.11 sind A und B in $L^{n\times n}$ simultan diagonalisierbar. Also ist auch $A+B$ in $L^{n\times n}$ diagonalisierbar. Nach Satz 10.52 hat μ_{A+B} keine mehrfache Nullstellen in L. Nach Lemma 16.11 ist $A+B$ separabel und nach Lemma 16.15 halbeinfach. $\square$

16.3 Verallgemeinerte Jordanblöcke

Bemerkung 16.17. Ist $\gamma \in K[X]$ irreduzibel, so kann man $\lambda := B(\gamma)$ als Element des Körpers $L := \mathrm{C}(B(\gamma))$ auffassen. Im folgenden Lemma studieren wir den Jordanblock $J_k(\lambda) \in L^{k\times k}$ als Matrix in $K^{n\times n}$, wobei $n := k\deg(\gamma)$. Für $\gamma = X-\mu$ erhält man $J_k(\gamma) = J_k(\mu)$.

Satz 16.18 (Verallgemeinerter Jordanblock)**.** *Sei $\gamma \in K[X]$ irreduzibel und separabel vom Grad d und $k \in \mathbb{N}$. Dann gilt*

$$J_k(\gamma) := \begin{pmatrix} B(\gamma) & & & 0 \\ 1_d & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1_d & B(\gamma) \end{pmatrix} \approx B(\gamma^k).$$

Beweis. Sei $J := J_k(\gamma)$. Eine einfache Induktion zeigt

$$J^m = \begin{pmatrix} B(\gamma)^m & & & 0 \\ mB(\gamma)^{m-1} & \ddots & & \\ & \ddots & \ddots & \\ * & & mB(\gamma)^{m-1} & B(\gamma)^m \end{pmatrix}$$

für $m \in \mathbb{N}_0$ (vgl. Satz 14.42). Es folgt

$$\gamma(J) = \begin{pmatrix} 0 & & & 0 \\ \gamma'(B(\gamma)) & \ddots & & \\ & \ddots & \ddots & \\ * & & \gamma'(B(\gamma)) & 0 \end{pmatrix}.$$

Insbesondere ist $\gamma(J)$ nilpotent und $\mu_J \mid \gamma^k$. Eine weitere Induktion zeigt

$$\gamma(J)^2 = \begin{pmatrix} 0 & & & & 0 \\ 0 & \ddots & & & \\ \gamma'(B(\gamma))^2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ * & & \gamma'(B(\gamma))^2 & 0 & 0 \end{pmatrix}, \ldots, \gamma(J)^{k-1} = \begin{pmatrix} 0 & & & 0 \\ \vdots & \ddots & & \\ 0 & & \ddots & \\ \gamma'(B(\gamma))^{k-1} & 0 & \cdots & 0 \end{pmatrix}.$$

Da γ separabel ist, gilt $0 \le \deg(\gamma') < \deg(\gamma)$. Da γ irreduzibel ist, müssen γ und γ' teilerfremd sind. Nach Bézout existieren $\alpha, \beta \in K[X]$ mit $\alpha\gamma + \beta\gamma' = 1$. Es folgt

$$1 = \alpha(B(\gamma))\gamma(B(\gamma)) + \beta(B(\gamma))\gamma'(B(\gamma)) = \beta(B(\gamma))\gamma'(B(\gamma)).$$

Also ist $\gamma'(B(\gamma))$ invertierbar und $\gamma'(B(\gamma))^{k-1} \neq 0$. Dies zeigt $\mu_J = \gamma^k$. Die Behauptung folgt aus Satz 15.21. □

Satz 16.19 (JORDAN-CHEVALLEY-Zerlegung). *Für jede separable Matrix $A \in K^{n\times n}$ existieren eindeutig bestimmte Matrizen $D, N \in K^{n\times n}$ mit folgenden Eigenschaften:*

(a) $A = D + N$ und $DN = ND$.

(b) D ist halbeinfach und N nilpotent.

Ggf. existiert ein $\alpha \in K[X]$ mit $\alpha(A) = D$.

Beweis. Wir transformieren A zunächst in die Weierstraß-Normalform. Da μ_A separabel ist, können wir anschließend jeden Block $B(\gamma^a)$ in der Weierstraß-Normalform zu $J_a(\gamma)$ umformen (Satz 16.18). Sei also $S \in \mathrm{GL}(n, K)$ mit

$$W := SAS^{-1} = \operatorname{diag}(J_{a_1}(\gamma_1), \ldots, J_{a_s}(\gamma_s))$$

Für $i = 1, \ldots, s$ sei $d_i := \deg(\gamma_i)$ und

$$D_i := \operatorname{diag}(B(\gamma_i), \ldots, B(\gamma_i)) \in K^{a_id_i\times a_id_i}, \qquad N_i := J_{a_i}(\gamma_i) - D_i,$$
$$\check{D} := \operatorname{diag}(D_1, \ldots, D_s), \qquad \check{N} := \operatorname{diag}(N_1, \ldots, N_s).$$

Dann gilt $W = \check{D} + \check{N}$. O. B. d. A. seien $\gamma_1, \ldots, \gamma_k$ die *verschiedenen* Primteiler von μ_A. Für $1 \le i \le k$ und $1 \le j \le s$ mit $\gamma_i = \gamma_j$ sei außerdem $a_i \ge a_j$. Nach Lemma 15.20 ist $\mu_{D_i} = \gamma_i$ und $\mu_{\check{D}} = \gamma_1 \ldots \gamma_k$, d. h. $\check{D}$ ist halbeinfach. Andererseits ist $\check{N}$ eine strikte untere Dreiecksmatrix und daher nilpotent (Beispiel 14.26). Eine Rechnung wie in Satz 16.18 zeigt $D_iN_i = N_iD_i$ und $\check{D}\check{N} = \check{D}\check{N}$. Da das Minimalpolynom nicht von der Basiswahl abhängt, ist $D := S^{-1}\check{D}S$ halbeinfach und $N := S^{-1}\check{N}S$ nilpotent. Außerdem gilt $DN = ND$ und $A = S^{-1}WS = D + N$.

Wegen $J_i := J_{a_i}(\gamma_i) = D_i + N_i$ ist $D_i \in \mathrm{C}(J_i)$. Nach Satz 16.18 und Folgerung 15.34 existiert ein $\alpha_i \in K[X]$ mit $\alpha_i(J_i) = D_i$ für $i = 1, \ldots, k$. Sei $1 \le j \le s$ mit $\gamma_j = \gamma_i$. Dann ist $a_j \le a_i$ und $J_i = \left(\begin{smallmatrix} J_j & 0 \\ * & * \end{smallmatrix}\right)$. Es folgt

$$\begin{pmatrix} D_j & 0 \\ * & * \end{pmatrix} = D_i = \alpha_i(J_i) = \begin{pmatrix} \alpha_i(J_j) & 0 \\ * & * \end{pmatrix}$$

und $\alpha_i(J_j) = D_j$. Nach den chinesischen Restsatz existiert ein $\alpha \in K[X]$ mit $\alpha \equiv \alpha_i \pmod{\gamma_i^{a_i}}$ für $i = 1, \ldots, k$. Aus $\gamma_i^{a_i}(J_i) = 0$ folgt

$$\alpha(W) = \operatorname{diag}\bigl(\alpha_1(J_1), \ldots, \alpha_s(J_s)\bigr) = \operatorname{diag}(D_1, \ldots, D_s) = \check{D}$$

und $\alpha(A) = S^{-1}\alpha(W)S = D$. Damit ist die Existenzaussage bewiesen. Da μ_D die gleichen Primteiler wie μ_A hat, ist D separabel.

Sei nun $A = \grave{D} + \grave{N}$ mit den gleichen Eigenschaften. Wir zeigen zunächst, dass $\grave{D}$ separabel ist. Über einem Zerfällungskörper sind $\grave{D}$ und $\grave{N}$ nach Satz 14.8 und Lemma 14.12 simultan trigonalisierbar. Bzgl. einer geeigneten Basis ist dann A eine Dreiecksmatrix. Als nilpotente Matrix muss $\grave{N}$ bzgl. dieser Basis eine strikte Dreiecksmatrix sein. Daher haben A und $\grave{D}$ die gleiche Hauptdiagonale und damit das gleiche charakteristische Polynom (beachte: χ_A hängt nicht vom Zerfällungskörper ab). Da A separabel ist, muss auch $\grave{D}$ separabel sein. Nach Voraussetzung ist $\grave{D}$ mit A und mit $\alpha(A) = D$ vertauschbar.

Nach Lemma 16.16 ist $D - \grave{D}$ halbeinfach. Analog sind auch $\grave{N}$ und N vertauschbar. Multipliziert man $(\grave{N} - N)^{2n}$ aus, so erhält man Summanden der Form $\grave{N}^i N^{2n-i}$ mit $0 \le i \le 2n$. Im Fall $i \ge n$ ist $\grave{N}^i = 0$. Anderenfalls ist $2n - i \ge n$ und $N^{2n-i} = 0$. In jedem Fall ist $(\grave{N} - N)^{2n} = 0$. Insgesamt ist $D - \grave{D} = \grave{N} - N$ halbeinfach und nilpotent. Das geht nur mit dem Minimalpolynom X, d. h. $\grave{D} = D$ und $\grave{N} = N$. □

Folgerung 16.20. *Für jede Matrix $A \in \mathbb{C}^{n\times n}$ existieren eindeutig bestimmte Matrizen $D, N \in \mathbb{C}^{n\times n}$ mit folgenden Eigenschaften:*

(a) $A = D + N$ und $DN = ND$.

(b) D ist diagonalisierbar und N nilpotent.

Beweis. Nach Beispiel 16.13 ist A separabel und D ist genau dann halbeinfach, wenn D diagonalisierbar ist. □

Bemerkung 16.21. Man kann die Jordan-Chevalley-Zerlegung von $A \in K^{n\times n}$ berechnen, indem man A über einem Zerfällungskörper von μ_A betrachtet und dort die Jordan-Normalform benutzt. Da die Jordan-Chevalley-Zerlegung eindeutig bestimmt ist, liegen D und N (dennoch) in $K^{n\times n}$.

Beispiel 16.22. Sei $\gamma := X^2 + 1 \in \mathbb{Q}[X]$ und

$$A := B(\gamma^2) = \begin{pmatrix} . & . & . & -1 \\ 1 & . & . & . \\ . & 1 & . & -2 \\ . & . & 1 & . \end{pmatrix} \in \mathbb{Q}[X].$$

Wir bestimmen die Jordan-Chevalley-Zerlegung von A ohne den in Bemerkung 16.21 beschriebenen Umweg über $\mathbb{C}$ (oder $\mathbb{Q}(\mathrm{i})$, vgl. Aufgabe 14). Nach Satz 16.18 gilt $A \approx J_2(\gamma)$. Um diesen Basiswechsel zu realisieren, benötigen wir ein Basisvektor $b_3 \in \mathbb{Q}^4$ mit $\mu_{b_3} = \gamma$. Dafür eignet sich

$$b_3 := \gamma(A)e_1 = e_1 + e_3 = (1, 0, 1, 0)^{\mathrm{t}}.$$

Dann ist $b_4 := Ab_3 = (0, 1, 0, 1)^{\mathrm{t}}$. Für b_1 und b_2 soll gelten: $Ab_1 = b_2 + b_3$ und $Ab_2 = -b_1 + b_4$. Damit erhält man das Gleichungssystem

$$\gamma(A)b_1 = (A^2 + 1_4)b_1 = 2b_4 = (0, 2, 0, 2)^{\mathrm{t}}$$

mit der Lösung $b_1 = 2e_2$. Schließlich ist $b_2 = Ab_1 - b_3 = (-1, 0, 1, 0)^{\mathrm{t}}$. Für

$$S := \begin{pmatrix} . & -1 & 1 & . \\ 2 & . & . & 1 \\ . & 1 & 1 & . \\ . & . & . & 1 \end{pmatrix} \in \mathrm{GL}(4, \mathbb{Q})$$

gilt $S^{-1}AS = J_2(\gamma)$. Dies ergibt die Jordan-Chevalley-Zerlegung $A = D + N$ mit

$$D := S\,\mathrm{diag}(B(\gamma), B(\gamma))S^{-1} = \frac{1}{2}\begin{pmatrix} . & -1 & . & -1 \\ 3 & . & -1 & . \\ . & 1 & . & -3 \\ 1 & . & 1 & . \end{pmatrix},$$

$$N := S \begin{pmatrix} 0_2 & 0_2 \\ 1_2 & 0_2 \end{pmatrix} S^{-1} = \frac{1}{2} \begin{pmatrix} . & 1 & . & -1 \\ -1 & . & 1 & . \\ . & 1 & . & -1 \\ -1 & . & 1 & . \end{pmatrix}.$$

Satz 16.23. *Sei $A \in K^{n\times n}$ separabel mit Jordan-Chevalley-Zerlegung $A = D + N$. Dann gilt* $\mathrm{C}(A) = \mathrm{C}(D) \cap \mathrm{C}(N)$.

Beweis. Für $B \in \mathrm{C}(D) \cap \mathrm{C}(N)$ gilt $BA = BD + BN = DB + NB = AB$ und $B \in \mathrm{C}(A)$. Sei $\alpha \in K[X]$ mit $\alpha(A) = D$. Für $B \in \mathrm{C}(A)$ gilt dann $BD = B\alpha(A) = \alpha(A)B = DB$ und $BN = B(A - \alpha(A)) = (A - \alpha(A))B = NB$. Dies zeigt $B \in \mathrm{C}(D) \cap \mathrm{C}(N)$. $\square$

Aufgaben

Aufgabe 32. Seien $\alpha, \beta \in K[X]$ mit $\alpha \mid \beta \mid \alpha$. Zeigen Sie, dass ein $c \in K^\times$ mit $\alpha = c\beta$ existiert.

Aufgabe 33. Sei $n \in \mathbb{N}$. Zeigen Sie:

(a) $A \in \mathbb{F}_2^{n\times n}$ ist genau dann diagonalisierbar, wenn $A^2 = A$.

(b) $A \in \mathrm{GL}(n, \mathbb{F}_2)$ ist genau dann diagonalisierbar, wenn $A = 1_n$.
Hinweis: Satz 10.52.

(c) $|\mathrm{GL}(n, \mathbb{F}_2)| = (2^n - 1)(2^n - 2)\dots(2^n - 2^{n-1})$.
Hinweis: Satz 7.45.

Aufgabe 34. Sei $A \in K^{n\times m}$ und $B \in K^{m\times n}$. Beweisen Sie die *Determinanten-Formel* von Sylvester

$$\det(1_n + AB) = \det(1_m + BA).$$

Hinweis: Lemma 10.38.

Aufgabe 35. Seien $A, B \in K^{n\times n}$.

(a) Beweisen oder widerlegen Sie:

$$\chi_A = \chi_{A^{\mathrm{t}}}, \qquad \mu_A = \mu_{A^{\mathrm{t}}}, \qquad \mu_{AB} = \mu_{BA}.$$

(b) Wie lassen sich $\chi_{A^{-1}}$ und $\mu_{A^{-1}}$ aus χ_A und μ_A berechnen, falls A invertierbar ist?

Aufgabe 36. Zeigen Sie $\chi_A = X^3 - \mathrm{tr}(A)X^2 + \frac{1}{2}\big(\mathrm{tr}(A)^2 - \mathrm{tr}(A^2)\big)X - \det(A)$ für alle $A \in K^{3\times 3}$.

Aufgabe 37. Sei V ein euklidischer Raum und $U, W \le V$. Zeigen Sie:

(a) $(U + W)^\perp = U^\perp \cap W^\perp$.

(b) $(U \cap W)^\perp = U^\perp + W^\perp$.

Hinweis: Man kann Lemma 16.3 benutzen.

Aufgabe 38. Sei K ein Körper, $n \in \mathbb{N}$ und $S \in K^{n\times n}$. Zeigen Sie, dass

$$\{A \in \mathrm{GL}(n, K) : ASA^{\mathrm{t}} = S\}$$

eine Untergruppe von $\mathrm{GL}(n, K)$ ist. Für $S = 1_n$ erhält man $\mathrm{O}(n, K)$.

Aufgabe 39. Seien $v, w \in \mathbb{R}^2$ linear unabhängig. Dann bilden $0, v, w$ die Eckpunkte eines Dreiecks mit Seitenlängen $A := |v|$, $B := |w|$ und $C := |v - w|$. Seien α, β, γ die Winkel gegenüber von A, B, C. Zeigen Sie:

(a) (Sinussatz) $\frac{\sin\alpha}{A} = \frac{\sin\beta}{B} = \frac{\sin\gamma}{C}$.

(b) (Kosinussatz) $C^2 = A^2 + B^2 - 2AB\cos\gamma$.

(c) (Trigonometrischer Pythagoras) $\sin(\alpha)^2 + \cos(\alpha)^2 = 1$.

Aufgabe 40. Für $\zeta := \cos(\pi/5) + \mathrm{i}\sin(\pi/5) \in \mathbb{C}$ gilt $\zeta^5 = -1$ (siehe Beweis von Lemma 11.27). Sei $\omega := \zeta + \zeta^{-1} = 2\mathrm{Re}(\zeta) \in \mathbb{R}$. Zeigen Sie:

(a) $\omega^2 - \omega - 1 = 0$.
Hinweis: $X^5 + 1 = (X + 1)(X^4 - X^3 + X^2 - X + 1)$.

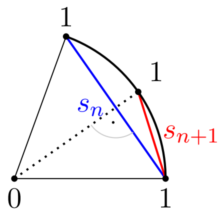
(b) $\omega = \frac{1}{2}(\sqrt{5} + 1)$.

(c)
$$D(\pi/5) = \frac{1}{4}\begin{pmatrix} \sqrt{5}+1 & -\sqrt{10-2\sqrt{5}} \\ \sqrt{10-2\sqrt{5}} & \sqrt{5}+1 \end{pmatrix}.$$
Hinweis: $\cos(\varphi)^2 + \sin(\varphi)^2 = 1$.

Aufgabe 41. Wir approximieren die Halbkreisbogenlänge π durch den halben Umfang eines regelmäßigen 2^n-Ecks mit „Radius“ 1. Dafür sei s_n die Seitenlänge des regelmäßigen 2^n-Ecks.

(a) Zeigen Sie $s_2 = \sqrt{2}$.

(b) Zeigen Sie $s_{n+1} = \sqrt{2 - \sqrt{4 - s_n^2}}$ durch zweimalige Anwendung von Pythagoras:



(c) Zeigen Sie
$$2^n \underbrace{\sqrt{2 - \sqrt{2 + \sqrt{2 + \ldots \sqrt{2}}}}}_{n \text{ Wurzeln}} = \lim_{n\to\infty} 2^n s_{n+1} = \pi.$$

Aufgabe 42. Sei $V := \mathbb{R}^n$ der euklidische Raum bzgl. des Standardskalarprodukts. Sei $v \in V$ normiert und $S_v \in \mathrm{O}(V)$ die Spiegelung an $v^\perp$. Zeigen Sie $[S_v] = 1_n - 2v^{\mathrm{t}}v$.
Bemerkung: Solche Matrizen nennt man *Householder-Transformationen.*

Aufgabe 43. Sei $\alpha \in \mathbb{R}[X]$. Zeigen Sie, dass Polynome $\gamma_1, \ldots, \gamma_k \in \mathbb{R}[X]$ mit $\alpha = \gamma_1 \ldots \gamma_k$ und $\deg(\gamma_i) \leq 2$ für $i = 1, \ldots, k$ existieren.
Hinweis: Bemerkung 11.37.

Aufgabe 44. Zeigen Sie, dass man im Hauptachsensatz und im Spektralsatz die Transformationsmatrix S mit $\det(S) = 1$ wählen kann.

Aufgabe 45. Sei V ein $\mathbb{R}$-Vektorraum und $\beta \in \mathrm{Bil}(V)$ symmetrisch mit $\mathrm{ind}(\beta) = (r, s, t)$. Zeigen Sie, dass r (bzw. s) die maximale Dimension eines Unterraums $U \le V$ ist, sodass die Einschränkung von β auf $U \times U$ positiv (bzw. negativ) definit ist.

Aufgabe 46. Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch mit $\mathrm{ind}(A) = (r, s, t)$ und $\lambda \in \mathbb{R}$. Zeigen Sie

$$\mathrm{ind}(\lambda A) = \begin{cases} (r,s,t) & \text{falls } \lambda > 0, \\ (s,r,t) & \text{falls } \lambda < 0, \\ (0,0,n) & \text{falls } \lambda = 0. \end{cases}$$

Aufgabe 47. Seien $A, B \in \mathbb{R}^{n\times n}$ vertauschbare positiv (semi)definite Matrizen. Zeigen Sie, dass $A + B$ positiv (semi)definit ist.
Hinweis: Simultan diagonalisieren.

Aufgabe 48. Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Wie kann man an χ_A ablesen, ob A positiv semidefinit ist? Beweisen Sie ein Kriterium in Analogie zu Satz 12.44.

Aufgabe 49. Eine hermitesche Matrix $A \in \mathbb{C}^{n\times n}$ heißt *positiv (semi)definit*, falls $\overline{v}Av^{\mathrm{t}} > 0$ bzw. $\overline{v}Av^{\mathrm{t}} \ge 0$ für alle $v \in \mathbb{C}^n \setminus \{0\}$ gilt. Zeigen Sie:

(a) Genau dann ist A positiv (semi)definit, wenn alle Eigenwerte von A positiv (bzw. nicht-negativ) sind.

(b) Für $k \in \mathbb{N}$ besitzt jede positiv (semi)definite Matrix genau eine positiv (semi)definite k-te Wurzel.

Aufgabe 50. Seien $A_1, \ldots, A_k \in K^{n\times n}$ diagonalisierbar und paarweise vertauschbar. Zeigen Sie, dass $A_1, \ldots, A_k$ simultan diagonalisierbar sind.
Hinweis: Die Induktion nach k ist nicht-trivial! Man kann Aufgabe 17 benutzen.

Aufgabe 51. Seien $A_1, \ldots, A_k \in K^{n\times n}$ paarweise vertauschbare trigonalisierbare Matrizen. Zeigen Sie, dass $A_1, \ldots, A_k$ simultan trigonalisierbar sind.

Aufgabe 52. Seien $n, k \in \mathbb{N}$ und $\lambda \in K$. Zeigen Sie

$$J_n(\lambda)^k = \begin{pmatrix} \lambda^k & & & & 0 \\ k\lambda^{k-1} & \ddots & & & \\ \binom{k}{2}\lambda^{k-2} & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \binom{k}{n-1}\lambda^{k-n+1} & \cdots & \binom{k}{2}\lambda^{k-2} & k\lambda^{k-1} & \lambda^k \end{pmatrix},$$

wobei $\binom{k}{l} = \frac{k(k-1)\ldots(k-l+1)}{l!}$ für $l = 1, \ldots, n-1$.

Aufgabe 53. Sei $\lambda \in K$ und $A := J_n(\lambda)$ ein Jordanblock. Zeigen Sie, dass $C \in K^{n\times n}$ genau dann mit A vertauschbar ist, wenn $c_1, \ldots, c_n \in K$ mit

$$C = \begin{pmatrix} c_1 & & & 0 \\ c_2 & \ddots & & \\ \vdots & \ddots & \ddots & \\ c_n & \cdots & c_2 & c_1 \end{pmatrix}$$

existieren.

Aufgabe 54. Sei V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Sei $b_1, \ldots, b_n$ eine Basis von V. Zeigen Sie, dass μ_f das kleinste gemeinsame Vielfache von $\mu_{b_1}, \ldots, \mu_{b_n}$ ist, d. h. es gibt kein normiertes Polynom kleineren Grades, dass durch $\mu_{b_1}, \ldots, \mu_{b_n}$ teilbar ist.

Aufgabe 55. Sei V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Sei $\lambda \in K$ ein Eigenwert von f, der mit Vielfachheit k als Nullstelle von μ_f auftritt. Zeigen Sie

$$E_\lambda(f) \subsetneq \mathrm{Ker}((f - \lambda\,\mathrm{id})^2) \subsetneq \ldots \subsetneq \mathrm{Ker}((f - \lambda\,\mathrm{id})^k) = H_\lambda(f).$$

Hinweis: Man kann die Weierstraß-Normalform benutzen.

Aufgabe 56. Sei $\alpha \in K[X] \setminus K$ normiert. Zeigen Sie $\chi_{B(\alpha)} = \alpha$ mit der Definition des charakteristischen Polynoms (und nicht über das Minimalpolynom wie in Lemma 15.20).

Aufgabe 57. Sei $\alpha = (X - \lambda_1) \ldots (X - \lambda_n) \in K[X]$ mit paarweise verschiedenen $\lambda_1, \ldots, \lambda_n$. Sei $V := (\lambda_i^{j-1}) \in K^{n\times n}$ die Vandermonde-Matrix. Zeigen Sie $VB(\alpha)V^{-1} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Aufgabe 58. Sei $\alpha = X^n + a_{n-1}X^{n-1} + \ldots + a_0$ und

$$S = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & 1 \\ a_2 & a_3 & \iddots & 1 & \\ \vdots & \iddots & \iddots & & \\ a_{n-1} & 1 & & & \\ 1 & & & & 0 \end{pmatrix} \in \mathrm{GL}(n, K)$$

(a) Zeigen Sie $S^{-1}B(\alpha)S = B(\alpha)^{\mathrm{t}}$.

(b) (Taussky) Folgern Sie, dass jede quadratische Matrix das Produkt von zwei symmetrischen Matrizen ist.

Aufgabe 59. Sei $A \in \mathbb{C}^{n\times n}$ mit lauter reellen Eigenwerten. Zeigen Sie, dass A zu einer reellen Matrix ähnlich ist.
Hinweis: Frobenius-Normalform.

Aufgabe 60. Beweisen Sie Folgerung 14.37 über einen beliebigen Körper K anstelle von $\mathbb{C}$.

Aufgabe 61.

(a) Seien $A \in K^{n\times n}$ und $B \in K^{m\times m}$, sodass μ_A und μ_B teilerfremd sind. Zeigen Sie

$$\mathrm{C}(\mathrm{diag}(A,B)) = \big\{\mathrm{diag}(C,D) : C \in \mathrm{C}(A),\ D \in \mathrm{C}(B)\big\} \cong \mathrm{C}(A) \times \mathrm{C}(B).$$

Hinweis: Lemma von Bézout.

(b) Sei A halbeinfach mit $\chi_A = \gamma_1^{a_1} \dots \gamma_k^{a_k}$ (Primfaktorzerlegung). Zeigen Sie

$$\dim \mathrm{C}(A) = \sum_{i=1}^{k} a_i^2 \deg(\gamma_i)$$

mit (a). Vergleichen Sie mit Bemerkung 15.33.

Aufgabe 62. Sei V ein K-Vektorraum und $F \subseteq \mathrm{End}(V)$. Ein Unterraum $U \le V$ heißt F-*invariant*, falls $f(U) \subseteq U$ für alle $f \in F$ gilt. Sei

$$\mathrm{C}(F) := \bigcap_{f\in F} \mathrm{C}(f)$$

der *Zentralisator* von F (analog für Matrizen).

(a) Angenommen $\{0\}$ und V sind die einzigen F-invarianten Unterräume. Zeigen Sie, dass alle $g \in \mathrm{C}(F) \setminus \{0\}$ invertierbar sind.

(b) Sei $F = \{f_1, f_2\}$ mit

$$f_1 := \mathrm{diag}\big(B(X^2+1), B(X^2+1)\big) \in \mathbb{R}^{4\times 4}, \qquad f_2 := \begin{pmatrix} 0_2 & B(X^2-1) \\ -B(X^2-1) & 0_2 \end{pmatrix} \in \mathbb{R}^{4\times 4}.$$

Zeigen Sie, dass $\mathbb{H} := \mathrm{C}(F) \subseteq \mathbb{R}^{4\times 4}$ ein 4-dimensionaler $\mathbb{R}$-Vektorraum ist, in dem jedes 0 verschiedene Element invertierbar ist.
Bemerkung: Im Gegensatz zu Satz 15.39 ist die Multiplikation in $\mathbb{H}$ nicht kommutativ. Man nennt $\mathbb{H}$ den Hamilton*schen Schiefkörper*.

Stichwortverzeichnis

L

M

N

O

P