Lineare Algebra und Analytische Geometrie 2

Vorlesung Sommersemester 2023

Prof. Meinolf Geck, Lehrstuhl für Algebra, Universität Stuttgart
https://pnp.mathematik.uni-stuttgart.de/idsr/idsr1/geckmf

Dies ist das Skript zur Vorlesung Lineare Algebra und Analytische Geometrie 2 im Sommersemester (V4Ü2, 14 Wochen). Es ist eine Fortsetzung des Skripts zur Vorlesung LAAG 1 im vorherigen Wintersemester; insbesondere führen wir die Nummerierung der Kapitel, Abschnitte etc. aus dem letzten Semester fort.

Im Durchschnitt werden wiederum pro Woche etwa 6-7 Seiten dieses Skriptes behandelt.

Kommentare sehr willkommen! (Insbesondere Druckfehler, sonstige Unklarheiten, Verbesserungsvorschläge etc.)

Stuttgart, März 2023

Literatur

Besonders geeignet für diese Vorlesung:

- S. Axler, Linear Algebra done right. Undergraduate texts in mathematics, Springer-Verlag, 2015.
- T. S. Blyth and E. F Robertson, Further Linear Algebra, Springer Undergraduate Mathematics Series, Springer–Verlag 2002.
- G. Fischer, Lineare Algebra: Eine Einführung für Studienanfänger, Vieweg + Teubner Verlag; 17. Auflage 2010.
- B. Huppert und W. Willems, Lineare Algebra: Mit zahlreichen Anwendungen in Kryptographie, Codierungstheorie, Mathematischer Physik und Stochastischen Prozessen, Vieweg + Teubner Verlag, 2. Auflage 2010.
- R. Kaye and R. Wilson, Linear Algebra. Oxford University Press, New York, 1998.
- M. Koecher, Lineare Algebra und analytische Geometrie, Grundwissen Mathematik, Springer-Verlag, 4. Auflage, 2002.
- P. Petersen, Linear Algebra. Undergraduate texts in mathematics, Springer-Verlag, 2012.

Zum Auffrischen von Schulwisssen und Grundlagen:

- T. Glosauser, (Hoch)Schulmathematik, Ein Sprungbrett vom Gymnasium zur Uni. Springer-Spektrum, 2015.
- M. Liebeck, A Concise Introduction to Pure Mathematics. Chapman Hall/CRC Mathematics Series, CRC Press, 3rd edition 2010.
- MINT Kolleg Baden-Württemberg, Mathematik-Vorkurs (Online), siehe `http://www.mint-kolleg.de/stuttgart/angebote/online_kurse` (folge dort auch den Links für den Mathematik-Vorkurs und dann zu `Mathematik-Online`).

Frei verfügbare mathematische Software zum Ausprobieren/Experimentieren:

- GAP - Groups, Algorithms, and Programming, siehe `http://www.gap-system.org/` (Exaktes Rechnen mit Zahlen und diskreten algebraischen Strukturen.)
- SageMath, siehe `https://www.sagemath.org/` (Basiert auf der Programmiersprache Python; siehe `https://www.python.org/`)

Einige weiterführende Texte (Auswahl, wird laufend ergänzt):

- M. Artin, Algebra. Aus dem Englischen übersetzt von Annette A'Campo. Birkhäuser Verlag, 1993.
- N. L. Biggs, Discrete Mathematics, 2nd Edition. Oxford University Press, 2002.
- N. Bourbaki, Éléments de Mathématiques. Algèbre. Chap. 1 à 3, Masson, Paris, 1970; Chap. 4 à 7, Masson, Paris, 1981.
- J. G. Broida and S. Gill Williamson, Comprehensive Introduction to Linear Algebra, 2012; Web Version, Creative Commons CC0 1.0; see https://cseweb.ucsd.edu/~gill/CILASite/.
- H. D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel und R. Remmert, Zahlen. Grundwissen Mathematik, vol. 1, Springer-Verlag, Berlin, 1983.
- S. H. Friedberg, A. J. Insel und L. E. Spence, Linear Algebra, 4th ed., Pearson, 2002.
- P. R. Halmos, Naive Mengenlehre, Vandenhoeck & Ruprecht, 5. Auflage, 1994.
- F. Lorenz, Lineare Algebra, 2 Bände. Spektrum Akademischer Verlag; 1. Band, 4. Auflage 2008; 2. Band, 3. Auflage, 1992.
- D. Poole, Linear Algebra: A Modern Introduction. Brooks Cole Pub Co., 3. Auflage, 2010.
- D. Serre, Matrices: Theory and Applications. Graduate Texts in Mathematics 216, Springer-Verlag, 2. Auflage, 2010.

Inhaltsverzeichnis

Kapitel V: Determinanten und Determinantenfunktionen

Determinanten von Matrizen sind in einer einführenden Vorlesung zur Matrix-Theorie (oder zur Linearen Algebra) meist ein etwas kniffliges Thema. Dies liegt einerseits daran, dass die Definition von $\det(A)$ "vom Himmel zu fallen scheint". (Es wird erst im Laufe der Zeit klarer, warum die etwas künstlich aussehende Definition die einzig mögliche ist, damit $\det(A)$ bestimmte Eigenschaften hat.) Andererseits ist es auch so, dass Beweise zu Aussagen über Determinanten meist technisch anspruchsvoller sind als das, was man bis dahin gesehen hat.

21. *Definition der Determinante*

Vermutlich ist die Formel für 2×2-Matrizen bekannt: $\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}$.

Vielleicht haben Sie auch die ***Regel von Sarrus*** für 3×3-Matrizen schon einmal gesehen:

$$\det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

Versuchen wir, ein allgemeines Muster in der obigen Formel zu finden. Die Terme, die aufsummiert werden, haben alle die Form $\pm a_{1i_1}a_{2i_2}a_{3i_3}$ wobei i_1, i_2, i_3 eine Umordnung der Ziffern $1, 2, 3$ ist. Eine solche Umordung ist nichts Anderes als eine bijektive Abbildung $\sigma\colon \{1, 2, 3\} \to \{1, 2, 3\}$, wobei $i_1 = \sigma(1)$, $i_2 = \sigma(2)$, $i_3 = \sigma(3)$. Es gibt genau $3! = 6$ solche bijektiven Abbildungen; das passt also genau zu den obigen 6 Summanden.

Für $n \in \mathbb{N}$ beliebig definieren wir nun S_n als die Menge aller bijektiven Abbildungen $\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$. Die Elemente von S_n heißen auch ***Permutationen***. Nach Kapitel I, Lemma 5.16, ist S_n eine endliche Menge mit $|S_n| = n!$. Der erste Ansatz für eine allgemeine Definition von $\det(A)$ wäre dann

$$\det(A) = \sum_{\sigma \in S_n} \pm\, a_{1\sigma(1)}\, a_{2\sigma(2)} \cdots a_{n\sigma(n)} \qquad \text{wobei} \qquad A = [a_{ij}]_{1 \leqslant i,j \leqslant n}.$$

Es ist nun nicht ganz leicht, die "richtige" allgemeine Formel für die Vorzeichen zu raten. Dazu: Sei $n \in \mathbb{N}$ und $\mathcal{T}_2$ die Menge der 2-elementigen Teilmengen von $\{1, 2, \dots, n\}$. Ist $X = \{i, j\} \in \mathcal{T}_2$, so setze $\sigma(X) := \{\sigma(i), \sigma(j)\} \in \mathcal{T}_2$ für $\sigma \in S_n$. Dann sei $N(\sigma)$ die Menge aller $X \in \mathcal{T}_2$, so dass σ die Reihenfolge der beiden Ziffern in X vertauscht. D.h., ist $X = \{i, j\}$ mit $1 \leqslant i < j \leqslant n$, so gilt $X \in N(\sigma) \Leftrightarrow \sigma(i) > \sigma(j)$. Wir bezeichnen $N(\sigma)$ als die Menge der ***Fehlstände*** von σ. Dann wird das ***Signum*** von σ definiert als

$$\operatorname{sgn}(\sigma) := \begin{cases} +1 & \text{falls } |N(\sigma)| \text{ gerade ist,} \\ -1 & \text{falls } |N(\sigma)| \text{ ungerade ist.} \end{cases}$$

Überzeugen Sie sich davon, dass das so definierte Signum genau die obigen Vorzeichen in den Formeln für 2×2- und 3×3-Matrizen ergibt.

Beispiel 21.1. (a) Ist id $\in S_n$ die identische Abbildung (also $\mathrm{id}(i) = i$ für $i = 1, \ldots, n$), so gibt es offenbar keine Fehlstände, also gilt $N(\mathrm{id}) = \varnothing$ und $\mathrm{sgn}(\mathrm{id}) = 1$.
(b) Seien $k, l \in \{1, \ldots, n\}$ fest mit $k < l$. Dann definieren wir $\tau_{kl} \in S_n$ durch $\tau_{kl}(k) := l$, $\tau_{kl}(l) := k$ und $\tau_{kl}(i) := i$ für alle $i \neq k, l$. Also vertauscht τ_{kl} die beiden Ziffern k und l, und lässt alle anderen Ziffern fest. Eine solche Permutation heißt auch ***Transposition***. Es gilt $\tau_{kl} \circ \tau_{kl} = \mathrm{id}$, also $\tau_{kl}^{-1} = \tau_{kl}$. Behauptung: $N(\tau_{12}) = \{\{1, 2\}\}$ und $\mathrm{sgn}(\tau_{12}) = -1$.
[Denn: Sei $\{i, j\} \in \mathcal{T}_2$ mit $i < j$. Sei zuerst $j \geqslant 3$. Dann gilt $\tau_{12}(j) = j$; außerdem ist $\tau_{12}(i) = i$ für $i \geqslant 3$, und $\tau_{12}(i) \in \{1, 2\}$ für $i = 1, 2$. In jedem Fall also $\tau_{12}(i) < j = \tau_{12}(j)$, d.h., $\{i, j\} \notin N(\tau_{12})$. Sei nun $j < 3$. Dann folgt $i = 1$, $j = 2$ und damit $\{i, j\} = \{1, 2\} \in N(\tau_{12})$. Also ist $N(\tau_{12}) = \{\{1, 2\}\}$ und $\mathrm{sgn}(\tau_{12}) = -1$.]

Wir werden weiter unten sehen, dass $\mathrm{sgn}(\tau_{kl}) = -1$ für beliebige $1 \leqslant k < l \leqslant n$ gilt.
(Oder Übung: Bestimmen Sie direkt $N(\tau_{kl})$ und zeigen Sie, dass $|N(\tau_{kl})| = 2(l - k - 1) + 1$ gilt.)

Definition 21.2 (***Leibniz–Formel***). Sei R ein kommutativer Ring mit 1. Dann ist die ***Determinante*** einer Matrix $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(R)$ definiert als

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma)\, a_{1\sigma(1)}\, a_{2\sigma(2)} \cdots a_{n\sigma(n)} \in R.$$

Beachte, dass diese Formel für große Werte von n praktisch völlig unbrauchbar ist (weil über $|S_n| = n!$ Terme summiert wird). Immerhin kann man die Formel aber für Matrizen mit bestimmten Eigenschaften auswerten. Beispiele:

Lemma 21.3. *Sei* $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(R)$ *so, dass alle Einträge in einer Zeile gleich* 0 *sind. Dann ist* $\det(A) = 0$.

Beweis. Sei $k \in \{1, \ldots, n\}$ so, dass $a_{kj} = 0$ für alle j gilt. Ist $\sigma \in S_n$, so ist also $a_{k\sigma(k)} = 0$ und damit der entsprechende Term $\mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ in der Formel für $\det(A)$ ebenfalls gleich 0. Also ist insgesamt $\det(A) = 0$. $\square$

Lemma 21.4. *Sei* $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(R)$ *eine obere Dreiecksmatrix, d.h.,* $a_{ij} = 0$ *für* $1 \leqslant j < i \leqslant n$. *Dann gilt* $\det(A) = a_{11} a_{22} \cdots a_{nn}$. *Eine analoge Aussage gilt auch für untere Dreiecksmatrizen. Inbesondere ist* $\det(I_n) = 1$.

Beweis. Sei $\sigma \in S_n$ so, dass der entsprechende Term $\mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ in der Formel für $\det(A)$ ungleich 0 ist. Dann ist $a_{k\sigma(k)} \neq 0$ für alle k, also $\sigma(k) \geqslant k$, d.h.,

$$\sigma(n) \geqslant n, \qquad \sigma(n-1) \geqslant n-1, \qquad \ldots, \qquad \sigma(1) \geqslant 1.$$

Aus der ersten Bedingung folgt $\sigma(n) = n$, dann aus der zweiten $\sigma(n-1) = n-1$ und so weiter bis $\sigma(1) = 1$. Damit liefert nur der Term für $\sigma = \mathrm{id}$ einen Beitrag ungleich 0 zu $\det(A)$, also gilt die genannte Formel. Angewandt auf $A = I_n$ erhalten wir $\det(I_n) = 1$. Der Beweis für untere Dreiecksmatrizen ist völlig analog. $\square$

Um weitere Formeln zu zeigen, benötigen wir noch Aussagen über die Signum-Funktion. Zunächst bemerken wir, dass die Menge S_n eine Gruppe ist mit der Hintereinanderausführung "$\circ$" als Verknüpfung, genannt die ***symmetrische Gruppe*** vom Grad n. (Die Assoziativität von $\circ$ rechnen Sie sofort nach.) Das neutrale Element ist die identische Abbildung id; das Inverse von $\pi \in S_n$ ist die Umkehrabbildung $\pi^{-1}\colon \{1,\dots,n\} \to \{1,\dots,n\}$. Wir schreiben eine Permutation $\pi \in S_n$ üblicherweise einfach als Liste $[\pi(1)\ \ \pi(2)\ \ \dots\ \ \pi(n)]$.

Satz 21.5. *Für alle* $\pi, \sigma \in S_n$ *gilt* $\operatorname{sgn}(\pi \circ \sigma) = \operatorname{sgn}(\pi)\operatorname{sgn}(\sigma)$ *und* $\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi^{-1})$.

Beweis. Sei $X = \{i, j\} \in \mathcal{T}_2$ wobei $i < j$. Es ergeben sich die folgenden vier möglichen Fälle:

Fall	$X \in N(\sigma)$?	$\sigma(X) \in N(\pi)$?	$X \in N(\pi \circ \sigma)$?
$\sigma(i) < \sigma(j),\ \pi(\sigma(i)) < \pi(\sigma(j))$	nein	nein	nein
$\sigma(i) < \sigma(j),\ \pi(\sigma(i)) > \pi(\sigma(j))$	nein	ja	ja
$\sigma(i) > \sigma(j),\ \pi(\sigma(i)) < \pi(\sigma(j))$	ja	ja	nein
$\sigma(i) > \sigma(j),\ \pi(\sigma(i)) > \pi(\sigma(j))$	ja	nein	ja

Sei $X \in \mathcal{T}_2$ beliebig. Dann sieht man sofort mit obiger Tabelle:

$$X \in N(\pi \circ \sigma) \quad \Leftrightarrow \quad X \in N(\sigma) \text{ oder } \sigma(X) \in N(\pi) \quad \text{(aber nicht beides).}$$

Also ist $N(\pi \circ \sigma) = N(\sigma) \,\triangle\, \sigma^{-1}(N(\pi))$, wobei "$\triangle$" die symmetrische Differenz von Mengen bezeichnet. Sind X, Y zwei Teilmengen einer Menge, so gilt $X \triangle Y = (X \cup Y) \setminus (X \cap Y)$ und $|X \triangle Y| = |X \cup Y| - |X \cap Y| = |X| + |Y| - 2|X \cap Y| \equiv |X| + |Y| \bmod 2$. Also folgt $|N(\pi \circ \sigma)| \equiv |N(\sigma)| + |\sigma^{-1}(N(\pi))| \equiv |N(\sigma)| + |N(\pi)| \bmod 2$ und damit $\operatorname{sgn}(\pi \circ \sigma) = \operatorname{sgn}(\pi)\operatorname{sgn}(\sigma)$.

Zur Formel für $\operatorname{sgn}(\pi^{-1})$: Mit $\sigma := \pi^{-1}$ folgt $\operatorname{sgn}(\mathrm{id}) = \operatorname{sgn}(\pi \circ \pi^{-1}) = \operatorname{sgn}(\pi)\operatorname{sgn}(\pi^{-1})$. Wegen $\operatorname{sgn}(\mathrm{id}) = 1$ und $\operatorname{sgn}(\pi) = \pm 1$ folgt daraus $\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\mathrm{id})\operatorname{sgn}(\pi)^{-1} = \operatorname{sgn}(\pi)$. □

Folgerung 21.6 (Zeilen/Spalten-Symmetrie)**.** *Es gilt* $\det(A) = \det(A^{\mathrm{tr}})$ *für* $A \in M_n(R)$.

Beweis. Sei $A = [a_{ij}]_{1 \leqslant i,j \leqslant n}$; dann ist $A^{\mathrm{tr}} = [a_{ji}]_{1 \leqslant i,j \leqslant n}$, also gilt

$$\det(A^{\mathrm{tr}}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{k=1}^{n} a_{\sigma(k)k}.$$

Sei $\sigma \in S_n$ fest und setze $l := \sigma(k)$ für $1 \leqslant k \leqslant n$. Mit k durchläuft auch l alle Ziffern von 1 bis n, nur in einer anderen Reihenfolge. Mit $k := \sigma^{-1}(l)$ erhalten wir also

$$\prod_{k=1}^{n} a_{\sigma(k)k} = \prod_{l=1}^{n} a_{l\sigma^{-1}(l)}$$

und damit $\det(A^{\mathrm{tr}}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)}$.

Schließlich sieht man sofort, dass die Abbildung $S_n \to S_n$, $\sigma \mapsto \sigma^{-1}$, bijektiv ist. Also können wir in der letzteren Formel für $\det(A^{\mathrm{tr}})$ auch überall σ^{-1} durch σ ersetzen. Wegen $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$ (siehe Satz 21.5) folgt dann $\det(A^{\mathrm{tr}}) = \det(A)$. □

Folgerung 21.7. *Sei* $1 \leqslant k < l \leqslant n$. *Dann gilt* $\operatorname{sgn}(\tau_{kl}) = -1$.

Beweis. Schreibe $\{1, \dots, n\} \setminus \{k, l\} = \{j_1, \dots, j_{n-2}\}$ und definiere $\pi \in S_n$ durch $\pi(1) := k$, $\pi(2) := l$, $\pi(i) := j_{i-2}$ für $i = 3, 4, \dots, n$. Man prüft sofort nach, dass dann $\pi \circ \tau_{12} \circ \pi^{-1} = \tau_{kl}$ gilt. Mit Satz 21.5 folgt daraus $\operatorname{sgn}(\tau_{kl}) = \operatorname{sgn}(\pi)\operatorname{sgn}(\tau_{12})\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\pi)^2\operatorname{sgn}(\tau_{12})$. Wegen $\operatorname{sgn}(\pi) = \pm 1$ und $\operatorname{sgn}(\tau_{12}) = -1$ (siehe Beispiel 21.1) gilt also auch $\operatorname{sgn}(\tau_{kl}) = -1$. □

Satz 21.8. *Jedes* $\pi \in S_n$ *ist ein Produkt von höchstens* $n - 1$ *Transpositionen.*
Ist π *ein Produkt von* r *Transpositionen, so gilt* $\operatorname{sgn}(\pi) = (-1)^r$.

Anstelle eines formalen Beweises illustrieren wir an einem Beispiel ein Verfahren, wie man die gewünschte Produktdarstellung findet. (Wir überlassen es als Übung, aus dieser beispielhaften Beschreibung einen formalen Beweis zu destillieren.)
Sei etwa $\pi = [3\ 5\ 1\ 2\ 4] \in S_5$. Wir gehen wie folgt vor: Suche die kleinste Ziffer i_1 mit $\pi(i_1) > i_1$; in diesem Fall ist dies $i_1 = 1$, mit $\pi(i_1) = 3$. Bilde die Transposition, die i_1 und $\pi(i_1)$ vertauscht, in diesem Fall also τ_{13}. Setze dann $\pi_1 := \tau_{13} \circ \pi = [1\ 5\ 3\ 2\ 4] \in S_5$.
Wiederhole das Verfahren mit π_1: Die kleinste Ziffer i_2 mit $\pi_1(i_2) > i_2$ ist $i_2 = 2$, mit $\pi_1(i_2) = 5$. Bilde dann die Transposition τ_{25} und setze $\pi_2 := \tau_{25} \circ \pi_1 = [1\ 2\ 3\ 5\ 4] \in S_5$.
Wiederhole das Verfahren mit π_2: Die kleinste Ziffer i_3 mit $\pi_2(i_3) > i_3$ ist $i_3 = 4$, mit $\pi_2(i_3) = 5$. Bilde dann die Transposition τ_{45} und setze $\pi_3 := \tau_{45} \circ \pi_2 = [1\ 2\ 3\ 4\ 5] \in S_5$.
Hier bricht also das Verfahren ab, mit $\tau_{45} \circ \tau_{25} \circ \tau_{13} \circ \pi = \mathrm{id}$, oder anders ausgedrückt: $\pi = \tau_{13} \circ \tau_{25} \circ \tau_{45}$. Mit Satz 21.5 und Folgerung 21.7 erhalten wir auch $\operatorname{sgn}(\pi) = (-1)^3 = -1$.

Beachte, dass es auf die Reihenfolge der Faktoren ankommt: Zum Beispiel ist $\tau_{kl} \circ \tau_{lm} \neq \tau_{lm} \circ \tau_{kl}$ für $k < l < m$. (Wende beide Seiten auf k an; auf der linken Seite ist das Ergebnis l, auf der rechten m.) Mit anderen Worten: S_n ist eine nicht-abelsche Gruppe für $n \geqslant 3$.

Ab hier Woche 2

Sei nun K ein Körper. Wir betrachten den Polynomring $R = K[X]$ über K in der Unbestimmten X. Da $R = K[X]$ ein kommutativer Ring mit 1 ist, können wir auch Matrizen mit Einträgen in $K[X]$ bilden und davon die Determinante; es gilt also

$$\det(F) \in K[X] \qquad \text{für alle} \qquad F \in M_n\big(K[X]\big).$$

Damit können wir die folgende zentrale Definition formulieren. (Dies ist motiviert durch Kapitel III, Bemerkung 14.1: Um zu sehen, ob $\lambda \in K$ ein Eigenwert von $A \in M_n(K)$ ist, geht man zur Matrix $A - \lambda I_n$ über und testet, ob diese invertierbar ist; ersetze nun λ durch X.)

Definition 21.9. Sei $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(K)$; dazu bilden wir

$$F := A - XI_n \in M_n\big(K[X]\big) \qquad \text{und} \qquad \chi_A := \det(F) = \det(A - XI_n) \in K[X].$$

Dann heißt χ_A das ***charakteristische Polynom*** von A.

Beispiel 21.10. (a) Für $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \in M_2(K)$ ist $A - XI_2 = \begin{bmatrix} -X & 1 \\ 1 & 1-X \end{bmatrix} \in M_2\big(K[X]\big)$ und damit $\chi_A = \det\Big(\begin{bmatrix} -X & 1 \\ 1 & 1-X \end{bmatrix}\Big) = (-X) * (1-X) - 1 = X^2 - X - 1$.

Wir bemerken, dass hier χ_A gleich dem Minimalpolynom μ_A ist (siehe Beispiel 27.2(b)).

(b) Sei $K = \mathbb{Q}$ und $A = \begin{bmatrix} -1 & 0 & 1 \\ 3 & 0 & -3 \\ 1 & 0 & -1 \end{bmatrix} \in M_3(\mathbb{Q})$. Mit der Regel von Sarrus erhalten wir

$$\begin{aligned} \chi_A = \det\Big(\begin{bmatrix} -1-X & 0 & 1 \\ 3 & -X & -3 \\ 1 & 0 & -1-X \end{bmatrix}\Big) &= (-1-X) * (-X) * (-1-X) - 1 * (-X) * 1 \\ &= -X^3 - 2X^2 - X + X = -X^2 * (X+2). \end{aligned}$$

Andererseits rechnet man einfach nach, dass $A^2 = -2A$ gilt; also ist das Minimalpolynom gegeben durch $\mu_A = X * (X+2)$. Hier ist also χ_A nur fast gleich μ_A.

(c) Ist $A \in M_n(K)$ eine obere Dreiecksmatrix mit $c_1, \ldots, c_n \in K$ auf der Diagonalen, so ist $F = A - XI_n$ eine obere Dreiecksmtarix mit $c_1 - X, \ldots, c_n - X$ auf der Diagonalen, also folgt $\chi_A = \det(A - XI_n) = (c_1 - X) * (c_2 - X) * \ldots * (c_n - X)$ (siehe Lemma 21.4). Insbesondere ist $\chi_{0_{n\times n}} = (-X)^n = (-1)^n X^n$ und $\chi_{I_n} = (1-X)^n = (-1)^n (X-1)^n$.

Den genauen Zusammenhang zwischen χ_A und μ_A werden wir später aufklären.

Lemma 21.11. *Sei* $\lambda \in K$. *Dann gilt* $\chi_A(\lambda) = \det(A - \lambda I_n) \in K$. *Insbesondere ist* $\det(A) = \chi_A(0)$ *der konstante Term von* χ_A.

Beweis. Sei $F = [f_{ij}]_{1\leqslant i,j\leqslant n} = A - XI_n \in M_n(K[X])$ und $\chi_A = \det(F) = \sum_{\pi\in S_n} \operatorname{sgn}(\pi) f_\pi$ wobei $f_\pi := f_{1\pi(1)} * \ldots * f_{n\pi(n)}$. Mit den Regeln zum Einsetzen von Elementen aus K in Polynome in Kapitel II, §9, folgt $\chi_A(\lambda) = \sum_{\pi\in S_n} \operatorname{sgn}(\pi) f_\pi(\lambda)$ und $f_\pi(\lambda) = f_{1\pi(1)}(\lambda) \cdots f_{n\pi(n)}(\lambda)$. Ist $i = \pi(i)$, so ist $f_{i\pi(i)} = a_{ii} - X$ also $f_{i\pi(i)}(\lambda) = a_{ii} - \lambda$; ist $i \neq \pi(i)$, so ist $f_{i\pi(i)} = a_{i\pi(i)}$ also $f_{i\pi(i)}(\lambda) = a_{i\pi(i)}$. Damit ist $\sum_{\pi\in S_n} \operatorname{sgn}(\pi) f_\pi(\lambda) = \det(B)$, wobei $B = [b_{ij}]_{1\leqslant i,j\in n} \in M_n(K)$ gegeben ist durch $b_{ij} = a_{ij}$ falls $i \neq j$, und $b_{ii} = a_{ii} - \lambda$, d.h., $B = A - \lambda I_n$. □

Satz 21.12. *Es gilt* $\chi_A \neq 0$ *und* $\operatorname{Grad}(\chi_A) = n$. *Schreiben wir* $\chi_A = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0$, *so gilt* $a_n = (-1)^n$ *und* $a_{n-1} = (-1)^{n-1} \operatorname{Spur}(A)$, *wobei* $\operatorname{Spur}(A) := \sum_{i=1}^n a_{ii}$ *(Summe der Diagonaleinträge) als die* ***Spur*** *von* A *bezeichnet wird.*

Beweis. Schreibe $\chi_A = \det(F) = \sum_{\pi\in S_n} \operatorname{sgn}(\pi) f_\pi$ wie im obigen Beweis. Sei $\pi \in S_n$ mit $f_\pi \neq 0$, also $f_{i\pi(i)} \neq 0$ für alle i. Für $i = \pi(i)$ ist $f_{i\pi(i)} = a_{ii} - X$, hat also Grad 1; für $i \neq \pi(i)$ ist $f_{i\pi(i)} = a_{ij}$, hat also Grad 0. Damit ist $\operatorname{Grad}(f_\pi) \leqslant n$. Falls $\pi \neq \mathrm{id}$, so gibt es mindestens zwei Ziffern $i \neq j$ mit $\pi(i) \neq i$ und $\pi(j) \neq j$, also ist dann $\operatorname{Grad}(f_\pi) \leqslant n-2$. Für $\pi = \mathrm{id}$ ist $f_{\mathrm{id}} = (a_{11} - X) * \ldots * (a_{nn} - X)$. Damit folgt

$$\chi_A = (a_{11} - X) * \ldots * (a_{nn} - X) + \text{Terme mit Grad} \leqslant n - 2.$$

Sind irgendwelche $c_1, \ldots, c_n \in K$ gegeben und multipliziert man $(c_1 - X) * \ldots * (c_n - X)$ aus, so erhält man $(-X)^n + (-X)^{n-1}(c_1 + \ldots + c_n) +$ Terme vom Grad $\leqslant n - 2$ (einfacher Beweis mit vollständiger Induktion nach n). □

22. ***Determinantenfunktionen***

Wir kommen nun zu einer weiteren fundamentalen Eigenschaft der Determinante. Sei wieder $n \in \mathbb{N}$ und R ein beliebiger kommutativer Ring mit 1. Ist $\pi \in S_n$ und $\operatorname{sgn}(\pi) = 1$, so heißt π eine ***gerade Permutation***; andernfalls eine ***ungerade Permutation***. Seien

$$A_n := \{\pi \in S_n \mid \operatorname{sgn}(\pi) = 1\} \qquad \text{und} \qquad A_n' := \{\pi \in S_n \mid \operatorname{sgn}(\pi) = -1\}.$$

Dann ist $S_n = A_n \cup A_n'$ (disjunkte Vereinigung). Wegen $\operatorname{sgn}(\mathrm{id}) = 1$ ist $\mathrm{id} \in A_n$.

Lemma 22.1. *Sei $n \geqslant 2$. Dann gilt $|S_n| = 2|A_n|$, d.h., je eine Hälfte der Permutationen in S_n sind gerade und die andere Hälfte sind ungerade. Ist $\tau \in S_n$ eine beliebige Permutation mit $\operatorname{sgn}(\tau) = -1$ (zum Beispiel eine Transposition), so ist $A_n \to A_n'$, $\pi \mapsto \pi \circ \tau$, eine bijektive Abbildung, also $S_n = A_n \cup \{\pi \circ \tau \mid \pi \in A_n\}$.*

Beweis. Sei $\tau \in S_n$ fest mit $\operatorname{sgn}(\tau) = -1$. Für $\pi \in A_n$ gilt dann $\operatorname{sgn}(\pi \circ \tau) = \operatorname{sgn}(\pi)\operatorname{sgn}(\tau) = -1$, also erhalten wir eine Abbildung $f \colon A_n \to A_n'$, $\pi \mapsto \pi \circ \tau$. Umgekehrt: Für $\sigma \in A_n'$ ist $\operatorname{sgn}(\sigma \circ \tau^{-1}) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau^{-1}) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) = (-1)(-1) = 1$, also erhalten wir auch eine Abbildung $g \colon A_n' \to A_n$, $\sigma \mapsto \sigma \circ \tau^{-1}$. Offenbar ist $f \circ g = \mathrm{id}_{A_n'}$ und $g \circ f = \mathrm{id}_{A_n}$, also sind f, g bijektiv und $g = f^{-1}$. Damit folgt $|A_n| = |A_n'|$ und $|S_n| = |A_n| + |A_n'| = 2|A_n|$. □

Satz 22.2. *Sei $A \in M_n(R)$. Sind zwei Zeilen von A gleich, so gilt $\det(A) = 0$.*

Beweis. Sei $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(R)$. Seien $1 \leqslant k < l \leqslant n$ so, dass die k-te und l-te Zeile von A gleich sind, also $a_{kj} = a_{lj}$ für $1 \leqslant j \leqslant n$. Betrachte nun die Transposition $\tau_{kl} \in S_n$. Nach Lemma 22.1 ist $S_n = A_n \cup A_n'$ mit $A_n' = \{\pi \circ \tau_{kl} \mid \pi \in A_n\}$. Damit erhalten wir

$$\begin{aligned}
\det(A) &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i\pi(i)} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{k\pi(k)} a_{l\pi(l)} \prod_{1 \leqslant i \leqslant n, i \neq k,l} a_{i\pi(i)} \\
&= \sum_{\pi \in A_n} \underbrace{\operatorname{sgn}(\pi)}_{=1} a_{k\pi(k)} a_{l\pi(l)} \prod_{1 \leqslant i \leqslant n, i \neq k,l} a_{i\pi(i)} \\
&\qquad + \sum_{\pi \in A_n} \underbrace{\operatorname{sgn}(\pi \circ \tau_{kl})}_{=-1} \underbrace{a_{k,(\pi \circ \tau_{kl})(k)}}_{=a_{k\pi(l)}} \underbrace{a_{l,(\pi \circ \tau_{kl})(l)}}_{=a_{l\pi(k)}} \prod_{1 \leqslant i \leqslant n, i \neq k,l} \underbrace{a_{i,(\pi \circ \tau_{kl})(i)}}_{=a_{i\pi(i)}}.
\end{aligned}$$

Nach Voraussetzung ist $a_{k\pi(l)} = a_{l\pi(l)}$ und $a_{l\pi(k)} = a_{k\pi(k)}$. Also ist die zweite Summe gleich

$$\sum_{\pi \in A_n} (-1) a_{l\pi(l)} a_{k\pi(k)} \prod_{1 \leqslant i \leqslant n, i \neq k,l} a_{i\pi(i)},$$

d.h., genau gleich dem Negativen der ersten Summe. Also ist $\det(A) = 0$. □

Definition 22.3. Eine Abbildung $\Delta\colon M_n(R) \to R$ heißt ***Determinantenfunktion***, wenn folgende beiden Eigenschaften gelten:

(D1) $\Delta(A)$ ist linear in jeder Zeile von $A \in M_n(R)$.

(D2) Es gilt $\Delta(A) = 0$ wenn zwei Zeilen von A gleich sind.

Mit (D1) ist gemeint: Sei $A = [a_{ij}] \in M_n(R)$. Sind $a_1, \ldots, a_n \in R^{1\times n}$ die Zeilen von A (also $a_k = [a_{k1} \ \ldots \ a_{kn}] \in R^{1\times n}$ für $1 \leqslant k \leqslant n$), so schreiben wir auch $\Delta(A) = \Delta(a_1, \ldots, a_n)$. Ist nun $k \in \{1, \ldots, n\}$ und $a_k = sv + tw$ mit $s, t \in K$ und $v, w \in R^{1\times n}$, so gelte

$$\Delta(A) = s\Delta(a_1, \ldots, a_{k-1}, v, a_{k+1}, \ldots, a_n) + t\Delta(a_1, \ldots, a_{k-1}, w, a_{k+1}, \ldots, a_n).$$

Aus (D1) und (D2) folgt auch:

(D2') Entsteht $A' \in M_n(R)$ durch Vertauschen von zwei Zeilen in $A \in M_n(R)$, so gilt $\Delta(A') = -\Delta(A)$.

Denn: Seien $a_1, \ldots, a_n \in R^{1\times n}$ die Zeilen von A. Seien $1 \leqslant k < l \leqslant n$ so, dass A' durch Vertauschen von a_k und a_l in A entsteht. Dann betrachte

$$\Delta(a_1, \ldots, a_{k-1}, a_k + a_l, a_{k+1}, \ldots, a_{l-1}, a_k + a_l, a_{l+1}, \ldots, a_n).$$

Wir interessieren uns nur für die k-te und l-te Zeile, schreiben obigen Ausdruck also kurz als $\Delta(\ldots, a_k + a_l, \ldots, a_k + a_l, \ldots)$. Wegen (D2) ist dies gleich 0. Mit (D1) erhalten wir

$$\begin{aligned} 0 &= \Delta(\ldots, a_k + a_l, \ldots, a_k + a_l, \ldots) \\ &= \Delta(\ldots, a_k, \ldots, a_k + a_l, \ldots) + \Delta(\ldots, a_l, \ldots, a_k + a_l, \ldots) \\ &= \underbrace{\Delta(\ldots, a_k, \ldots, a_k, \ldots)}_{=0 \text{ wegen (D2)}} + \underbrace{\Delta(\ldots, a_k, \ldots, a_l, \ldots)}_{=\Delta(A)} \\ &\qquad + \underbrace{\Delta(\ldots, a_l, \ldots, a_k, \ldots)}_{=\Delta(A')} + \underbrace{\Delta(\ldots, a_l, \ldots, a_l, \ldots)}_{=0 \text{ wegen (D2)}}. \end{aligned}$$

Also ist $\Delta(A) = -\Delta(A')$, d.h., es gilt (D2').

Satz 22.4. *Die Abbildung* $\det\colon M_n(R) \to R$ *ist eine Determinantenfunktion.*

Beweis. Sei $A = [a_{ij}] \in M_n(R)$. Sei $1 \leqslant k \leqslant n$ und $a_k = sv + tw$ mit $s, t \in R$ und $v, w \in R^{1\times n}$, mit Komponenten $v_1, \ldots, v_n \in R$ bzw. $w_1, \ldots, w_n \in R$. Dann gilt

$$\begin{aligned} \det(a_1, \ldots, a_n) &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i\pi(i)} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{k\pi(k)} \prod_{1\leqslant i\leqslant n, i\neq k} a_{i\pi(i)} \\ &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi)(s v_{\pi(k)} + t w_{\pi(k)}) \prod_{1\leqslant i\leqslant n, i\neq k} a_{i\pi(i)} \\ &= s\underbrace{\Big(\sum_{\pi \in S_n} \operatorname{sgn}(\pi) v_{\pi(k)} \prod_{1\leqslant i\leqslant n, i\neq k} a_{i\pi(i)}\Big)}_{=\det(a_1,\ldots,a_{k-1},v,a_{k+1},\ldots,a_n)} + t\underbrace{\Big(\sum_{\pi \in S_n} \operatorname{sgn}(\pi) w_{\pi(k)} \prod_{1\leqslant i\leqslant n, i\neq k} a_{i\pi(i)}\Big)}_{=\det(a_1,\ldots,a_{k-1},w,a_{k+1},\ldots,a_n)}; \end{aligned}$$

also gilt (D1) für det. Nach Satz 22.2 gilt auch (D2). □

Satz 22.5 (Cramersche Regel, 1750). *Sei* $R = K$ *ein Körper,* $b \in K^n$ *und* $A \in M_n(K)$ *mit* $\det(A) \neq 0$. *Dann ist* A *invertierbar und das* LGS *mit erweiterter Matrix* $[A|b] \in K^{n\times(n+1)}$ *hat genau eine Lösung* $x \in K^n$ *(nämlich* $x := A^{-1}\cdot b$*). Seien* $b_1, \dots, b_n \in K$ *die Komponenten von* b*. Für* $1 \leqslant k \leqslant n$ *sei* $x_k \in K$ *die* k*-te Komponente von* x*. Dann gilt*

$$x_k = \det(A)^{-1}\det(A_k) \quad \textit{mit} \quad A_k := \begin{bmatrix} a_{11} & \dots & a_{1,k-1} & b_1 & a_{1,k+1} & \dots & a_{1n} \\ a_{21} & \dots & a_{2,k-1} & b_2 & a_{2,k+1} & \dots & a_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,k-1} & b_n & a_{n,k+1} & \dots & a_{nn} \end{bmatrix} \in M_n(K).$$

Beweis. Seien $w_1, \dots, w_n \in K^n$ die Spalten von A; wir schreiben dann auch $\det(A) = \det(w_1, \dots, w_n)$. Nehmen wir zunächst an, es gibt eine Lösung $x \in K^n$ des LGS; wegen $A \cdot x = b$ ist dann $b = x_1 w_1 + \dots + x_n w_n$. Mit der Zeilen/Spalten-Symmetrie von det gelten (D1) und (D2) analog auch für Spalten. Damit folgt:

$$\begin{aligned} \det(A_k) &= \det(w_1, \dots, w_{k-1}, \underbrace{\textstyle\sum_{i=1}^n x_i w_i}_{=b}, w_{k+1}, \dots, w_n) \\ &= \sum_{i=1}^n x_i \underbrace{\det(w_1, \dots, w_{k-1}, w_i, w_{k+1}, \dots, w_n)}_{=0 \text{ wenn } i \neq k, \text{ wegen (D2)}} = x_k \underbrace{\det(w_1, \dots, w_{k-1}, w_k, w_{k+1}, \dots, w_n)}_{=\det(A)}. \end{aligned}$$

Wegen $\det(A) \neq 0$ ergibt dies $x_k = \det(A)^{-1}\det(A_k)$; insbesondere ist x eindeutig bestimmt. Wenden wir dies auf den Fall $b = 0_n$ an, so hat das homogene LGS mit Matrix A nur die triviale Lösung. Nach Kapitel III, Satz 13.3, können wir nun schließen, dass A invertierbar ist. Aber dann hat das LGS auch für $b \neq 0_n$ eine eindeutige Lösung, nämlich $x = A^{-1} \cdot b$, und die obige Rechnung liefert die gewünschten Formeln für die Komponenten x_k von x. □

Die Cramersche Regel hat keine große praktische Bedeutung für das Lösen von LGSen. Denn um damit die Komponenten der Lösung $x \in K^n$ zu bestimmen, müsste man $n + 1$ Determinanten berechnen; dies geht mit dem Gauß–Verfahren viel effizienter. Aber die Regel ist für manche theoretische Zwecke sehr nützlich, zum Beispiel:

Beispiel 22.6. Sei $A \in M_n(\mathbb{Z})$ und $b \in \mathbb{Z}^n$. Behauptung:

Ist $\det(A) = \pm 1$*, so gibt es genau ein* $x \in \mathbb{Z}^n$ *mit* $A \cdot x = b$.

Denn: Wir fassen A als Matrix in $M_n(\mathbb{Q})$ auf. Nach der Cramerschen Regel ist dann A invertierbar und $x = A^{-1} \cdot b$; die Komponenten von x berechnen sich als $x_k = \det(A)^{-1}\det(A_k)$ für $k = 1, \dots, n$. Nach Voraussetzung ist $\det(A) = \pm 1$; da außerdem alle Einträge von A und von b in $\mathbb{Z}$ sind, folgt mit der Leibniz-Formel $\det(A_k) \in \mathbb{Z}$. Also ist auch $x_k \in \mathbb{Z}$. □

Bemerkung 22.7. Gegeben seien zwei Punkte im $\mathbb{R}^2$:

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \in \mathbb{R}^2 \qquad \text{und} \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2 \qquad \text{mit } a_1, a_2, b_1, b_2 \geqslant 0.$$

Dann kann man sich überlegen, dass der Flächeninhalt des von a und b aufgespannten Parallelogramms durch den Absolutbetrag von $\det\left(\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}\right)$ gegeben ist ("geometrische Interpretation" der Determinante für $n = 2$). Allgemeiner sei $n \geqslant 2$ beliebig und seien Vektoren $v_1, \dots, v_d \in \mathbb{R}^n$ gegeben, so dass das Tupel $(v_1, \dots, v_d)$ linear unabhängig ist. Dann heißt

$$P = P(v_1, \dots, v_d) := \{s_1 v_1 + \dots + s_d v_d \mid s_i \in \mathbb{R} \text{ mit } 0 \leqslant s_i \leqslant 1 \text{ für } 1 \leqslant i \leqslant d\} \subseteq \mathbb{R}^n$$

ein d-dimensionales ***Parallelotop*** (oder ***Parallelepiped***). Wir bilden dann die Matrix $G_P = [g_{ij}] \in M_d(\mathbb{R})$ mit $g_{ij} := \langle v_i, v_j \rangle$ für $1 \leqslant i, j \leqslant d$, wobei $\langle\ ,\ \rangle \colon \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ das Standard-Skalarprodukt ist. Dann ist $\det(G_P) > 0$ und $\mathrm{vol}(P) := \sqrt{\det(G_P)}$ kann als ***Volumen*** von P interpretiert werden; siehe zum Kapitel 5, §4, Abschnitt 7*, im Buch von Koecher.
Für $n = d = 2$ und $a, b \in \mathbb{R}^2$ wie oben veranschaulicht man sich sofort, dass $P(a, b) \subseteq \mathbb{R}^2$ das von a und b aufgespannte Parallelogramm ist; außerdem rechnet man in der Tat nach:

$$G_{P(a,b)} = \begin{bmatrix} a_1^2 + a_2^2 & a_1 b_1 + a_2 b_2 \\ a_1 b_1 + a_2 b_2 & b_1^2 + b_2^2 \end{bmatrix} \quad \text{und} \quad \det(G_{P(a,b)}) = (a_1 b_2 - a_2 b_1)^2 = \det\left(\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}\right)^2.$$

23. *Die Produktregel*

Sei $\Delta \colon M_n(R) \to R$ eine Determinantenfunktion. Für $\tau \in S_n$ definieren wir die Matrix

$$A^\tau = (a^\tau_{ij})_{1 \leqslant i,j \leqslant n} \in M_n(R) \qquad \text{wobei} \qquad a^\tau_{ij} := \begin{cases} 1 & \text{falls } i = \tau(j), \\ 0 & \text{sonst.} \end{cases}$$

Die Matrix A^τ heißt die zu τ gehörige ***Permutationsmatrix***. Einige Beispiele für $n = 3$:

$$A^{[2\,1\,3]} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A^{[3\,1\,2]} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad A^{[3\,2\,1]} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

(Hier wird jeweils τ einfach durch die Liste $[\tau(1)\ \tau(2)\ \tau(3)]$ bezeichnet; wir erhalten A^τ durch Permutieren der Zeilen der Einheitsmatrix gemäß τ.) Es gilt $A^\tau \cdot e_i = e_{\tau(i)}$ für $1 \leqslant i \leqslant n$.

Lemma 23.1. *Es gilt* $\Delta(A^\pi) = \mathrm{sgn}(\pi)\Delta(I_n)$ *für alle* $\pi \in S_n$.

Beweis. Die Aussage gilt natürlich für $\pi = \mathrm{id}$, da $A^{\mathrm{id}} = I_n$. Sei nun π eine Transposition; dann ist $\mathrm{sgn}(\pi) = -1$. Vertauscht π die Ziffern i und j, so ist A^π die Elementarmatrix V_{ij}. Mit (D2') folgt $\Delta(A^\pi) = \Delta(V_{ij}) = -\Delta(I_n) = \mathrm{sgn}(\pi)\Delta(I_n)$, also gilt die Formel. Für den allgemeinen Fall benötigen wir die Regel

$$(*) \qquad A^{\pi \circ \sigma} = A^\pi \cdot A^\sigma \qquad \text{für alle } \pi, \sigma \in S_n.$$

Dazu: Sei $i \in \{1, \dots, n\}$ und $j := \sigma(i)$. Dann gilt $A^\sigma \cdot e_i = e_{\sigma(i)} = e_j$ und $A^\pi \cdot e_j = e_{\pi(j)}$. Es folgt $(A^\pi \cdot A^\sigma) \cdot e_i = A^\pi \cdot \big(A^\sigma \cdot e_i\big) = e_{\pi(j)} = e_{(\pi \circ \sigma)(i)} = A^{\pi \circ \sigma}(e_i)$, also gilt $(*)$.

Sei nun $\mathrm{id} \neq \pi \in S_n$. Nach Satz 21.8 gibt es Transpositionen $\tau_1, \dots, \tau_r \in S_n$ mit $1 \leqslant r \leqslant n-1$ und $\pi = \tau_1 \circ \dots \circ \tau_r$. Setzen wir $\pi' := \tau_2 \circ \dots \circ \tau_r \in S_n$, so folgt mit $(*)$ und (D2'): $\Delta(A^\pi) = \Delta(A^{\tau_1 \circ \pi'}) = \Delta\big(A^{\tau_1} \cdot A^{\pi'}\big) = -\Delta(A^{\pi'})$.

Danach wenden wir ein analoges Argument auf $A^{\pi'}$ an; nach insgesamt r Schritten folgt $\Delta(A^\pi) = (-1)^r \Delta(I_n)$, wobei $(-1)^r = \operatorname{sgn}(\pi)$, siehe Satz 21.8. □

Satz 23.2 (Produktregel für Determinantenfunktionen)**.** *Seien* $A, B \in M_n(R)$. *Dann gilt* $\Delta(A \cdot B) = \det(A)\Delta(B)$. *Insbesondere gilt* $\det(A \cdot B) = \det(A)\det(B)$.

Beweis. Sei $B \in M_n(R)$ fest. Dann definiere $\Delta' \colon M_n(R) \to R$ durch $\Delta'(A) := \Delta(A \cdot B)$ für alle $A \in M_n(R)$. Wir zeigen, dass (D1) und (D2) für Δ' gelten. Seien dazu $a_1, \ldots, a_n \in R^{1\times n}$ die Zeilen von A. Dann sind (nach Definition des Matrixproduktes) die Zeilen von $A \cdot B$ gegeben durch $a_1 \cdot B, \ldots, a_n \cdot B$. Gibt es $1 \leqslant i < j \leqslant n$ mit $a_i = a_j$, so gilt auch $a_i \cdot B = a_j \cdot B$ und damit $\Delta'(A) = \Delta(a_1 \cdot B, \ldots, a_n \cdot B) = 0$, weil (D2) für Δ gilt.

Nun zu (D1). Sei $k \in \{1, \ldots, n\}$ und $a_k = sv + tw$ mit $s, t \in K$ und $v, w \in R^{1\times n}$. Dann gilt $a_k \cdot B = s(v \cdot B) + t(w \cdot B)$, Da wir uns nur für die k-te Zeile interessieren, schreiben wir kurz $\Delta'(A) = \Delta(\ldots, a_k \cdot B, \ldots)$. Nun folgt $\Delta'(A) = \Delta(\ldots, s(v \cdot B) + t(w \cdot B), \ldots) = s\Delta(\ldots, v \cdot B, \ldots) + t\Delta(\ldots, w \cdot B, \ldots)$, weil (D1) für Δ gilt. Damit gilt (D1) auch für Δ'.

Also ist Δ' eine Determinantenfunktion. Seien nun $e_1, \ldots, e_n \in R^{1\times n}$ die Standard-Zeilenvektoren, d.h., e_k hat Eintrag 1 an der Stelle k und 0 sonst. Damit ist

$$a_k = a_{k1}e_1 + a_{k2}e_2 + \ldots + a_{kn}e_n = \textstyle\sum_{i=1}^n a_{ki}e_i \qquad \text{für } 1 \leqslant k \leqslant n.$$

Durch wiederholte Anwendung von (D1) folgt

$$\begin{aligned}
\Delta'(A) &= \Delta'\big(a_1, a_2, \ldots, a_n\big) = \Delta'\Big(\sum_{i_1=1}^n a_{1i_1}e_{i_1}, a_2, \ldots, a_n\Big) = \sum_{i_1=1}^n a_{1i_1}\Delta'\big(e_{i_1}, a_2, \ldots, a_n\big) \\
&= \sum_{i_1=1}^n a_{1i_1}\Delta'\Big(e_{i_1}, \sum_{i_2=1}^n a_{2i_2}e_{i_2}, a_3, \ldots, a_n\Big) = \sum_{i_1=1}^n \sum_{i_2=1}^n a_{1i_1}a_{2i_2}\Delta'\big(e_{i_1}, e_{i_2}, a_3, \ldots, a_n\big) \\
&= \ldots = \sum_{i_1=1}^n \sum_{i_2=1}^n \ldots \sum_{i_n=1}^n a_{1i_1}a_{2i_2}\cdots a_{ni_n}\Delta'\big(e_{i_1}, e_{i_2}, \ldots, e_{i_n}\big).
\end{aligned}$$

Wegen (D2) ist $\Delta'\big(e_{i_1}, e_{i_2}, \ldots, e_{i_n}\big) = 0$ wenn zwei der Indizes $i_1, i_2, \ldots, i_n$ gleich sind. Also brauchen wir nur die Tupel $(i_1, i_2, \ldots, i_n)$ zu betrachten, bei denen alle Komponenten verschieden sind; dies ist aber genau dann der Fall, wenn $[i_1\ i_2\ \ldots\ i_n] \in S_n$ gilt. Also folgt

$$\Delta'(A) = \sum_{\pi \in S_n} a_{1\pi(1)}a_{2\pi(2)}\cdots a_{n\pi(n)}\Delta'\big(e_{\pi(1)}, e_{\pi(2)}, \ldots, e_{\pi(n)}\big).$$

Für $\pi \in S_n$ sei nun $Q^\pi = [q_{ij}^\pi] \in M_n(R)$ die Matrix mit Zeilen $e_{\pi(1)}, e_{\pi(2)}, \ldots, e_{\pi(n)}$. Für $i, j \in \{1, \ldots, n\}$ gilt $q_{ij}^\pi = 1 \Leftrightarrow j = \pi(i) \Leftrightarrow i = \pi^{-1}(j)$. Also ist $Q^\pi = A^{\pi^{-1}}$ und mit Lemma 23.1 folgt $\Delta'(Q^\pi) = \Delta'(A^{\pi^{-1}}) = \operatorname{sgn}(\pi^{-1})\Delta'(I_n) = \operatorname{sgn}(\pi)\Delta'(I_n)$, wobei die letzte Gleichheit nach Satz 21.5 gilt. Also ist $\Delta'(A) = \det(A)\Delta'(I_n) = \det(A)\Delta(B)$. □

Folgerung 23.3 (Charakterisierung von det)**.** *Die Abbildung* $\det \colon M_n(R) \to \mathbb{R}$ *ist die eindeutige Determinantenfunktion mit Wert* 1 *auf der Einheitsmatrix.*

Beweis. Die Abbildung $\det\colon M_n(R) \to R$ ist eine Determinantenfunktion mit $\det(I_n) = 1$. Sei umgekehrt $\Delta\colon M_n(R) \to \mathbb{R}$ eine beliebige Determinantenfunktion mit $\Delta(I_n) = 1$. Nach Satz 23.2 (mit $B := I_n$) folgt $\Delta(A) = \Delta(A \cdot I_n) = \det(A)\Delta(I_n) = \det(A)$ für $A \in M_n(R)$. □

Ab hier Woche 3

Satz 23.4. *Sei* $1 \leqslant d < n$ *und* $A \in M_n(R)$ *eine obere Blockdreiecksmatrix der Form*

$$A = \left[\begin{array}{c|c} B & D \\ \hline 0 & C \end{array}\right] \quad \textit{wobei} \quad B \in M_d(R), \quad C \in M_{n-d}(R), \quad D \in R^{d\times(n-d)}.$$

Dann gilt $\det(A) = \det(B)\det(C)$.

Beweis. Zunächst rechnet man sofort nach, dass folgendes gilt:

$$A = \left[\begin{array}{c|c} B & D \\ \hline 0 & C \end{array}\right] = \left[\begin{array}{c|c} I_d & D \\ \hline 0 & C \end{array}\right] \cdot \left[\begin{array}{c|c} B & 0 \\ \hline 0 & I_{n-d} \end{array}\right].$$

Halte nun $D \in R^{d\times(n-d)}$ fest; wir definieren $\Delta_1\colon M_{n-d}(R) \to R$ und $\Delta_2\colon M_d(R) \to R$ durch

$$\Delta_1(C) := \det\left(\left[\begin{array}{c|c} I_d & D \\ \hline 0 & C \end{array}\right]\right) \qquad \text{und} \qquad \Delta_2(B) := \det\left(\left[\begin{array}{c|c} B & 0 \\ \hline 0 & I_{n-d} \end{array}\right]\right)$$

für alle $B \in M_d(R)$ und alle $C \in M_{n-d}(R)$. Nach Satz 23.2 ist $\det(A) = \Delta_1(C)\Delta_2(B)$, also genügt es zu zeigen, dass $\Delta_1(C) = \det(C)$ und $\Delta_2(B) = \det(B)$ gilt.

Behauptung: Δ_1 ist eine Determinantenfunktion. Nun, schreiben wir eine Zeile von C als Linearkomination von zwei Zeilenvektoren, so ist auch die entsprechende Zeile von

$$\tilde{C} := \left[\begin{array}{c|c} I_d & D \\ \hline 0 & C \end{array}\right]$$

eine solche Linearkombination (weil der linke untere $(n-d) \times d$ Block nur aus 0 besteht), also gilt (D1) für Δ_1. Analog: Sind zwei Zeilen von C gleich, so sind die entsprechenden Zeilen auch in $\tilde{C}$ gleich, also gilt auch (D2) für Δ_1. Nach Satz 23.2 folgt $\Delta_1(C) = \Delta_1(C \cdot I_{n-d}) = \det(C)\Delta_1(I_{n-d})$. Nun ist $\tilde{I}_{n-d}$ eine obere Dreiecksmatrix mit 1 entlang der Diagonalen; also gilt $\Delta_1(I_{n-d}) = \det(\tilde{I}_{n-d}) = 1$ (siehe Lemma 21.4). Damit ist $\Delta_1(C) = \det(C)$. Völlig analog sieht man, dass auch $\Delta_2(B) = \det(B)$ gilt. □

Beispiel: $$\det\left(\left[\begin{array}{cc|cc} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ \hline 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{array}\right]\right) = \det\left(\begin{bmatrix} 1 & -1 \\ 0 & 4 \end{bmatrix}\right) \cdot \det\left(\begin{bmatrix} -4 & -8 \\ 4 & 0 \end{bmatrix}\right) = 4 \cdot 32 = 128.$$

Bemerkung 23.5. (a) Seien $A, A' \in M_n(R)$. Wir halten folgende Regeln fest:

- Entsteht A' aus A durch Multiplikation einer Zeile mit $c \in R$, so $\det(A') = c\det(A)$.
- Entsteht A' durch Vertauschen von zwei Zeilen in A, so gilt $\det(A') = -\det(A)$.
- Entsteht A' aus A, indem man ein Vielfaches einer Zeile von A zu einer anderen Zeile

addiert, so gilt $\det(A') = \det(A)$.

Mit der Zeilen/Spalten-Symmetrie von det gelten analoge Regeln auch für Spalten.

Die erste Regel ist ein Spezialfall von (D1); die zweite Regel ist (D2'). Zur dritten Regel: Seien $i \neq j$ und $c \in R$ so, dass A' dadurch entsteht, dass man das c-Fache der j-ten Zeile in A zur i-ten Zeile addiert. Dann ist $A' = I_{ij}(c) \cdot A$, wobei $I_{ij}(c)$ eine Elementarmatrix wie in Kapitel III, Satz 12.4, ist. Mit der Produktregel folgt $\det(A') = \det(I_{ij}(c)) \det(A)$; weil $I_{ij}(c)$ eine Dreiecksmatrix mit 1 entlang der Diagonalen ist, gilt $\det(I_{ij}(c)) = 1$.

(b) Sei $R = K$ ein Körper und $A \in M_n(K)$. Mit dem Gauß-Verfahren (also einer Folge von elementaren Umformungen wie oben) erhalten wir $A \to A' \in M_n(K)$, wobei A' in Stufenform ist. Die obigen Formeln ergeben dann eine Konstante $0 \neq c \in K$ mit $\det(A) = c \det(A')$; Beachte: Gibt es $< n$ Stufen, so ist $\det(A') = 0$; andernfalls ist $A' = I_n$ und $\det(A') = 1$.

Beispiel 23.6. Seien $x_1, \ldots, x_n \in R$. Dann betrachte die ***Vandermonde-Matrix***

$$V(x_1, \ldots, x_n) := \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{bmatrix} \in M_n(R).$$

In Kapitel III, Beispiel 13.6, haben wir gesehen, dass $V(x_1, \ldots, x_n)$ invertierbar ist, wenn $R = K$ ein Körper ist und die x_i paarweise verschieden sind. Wir behaupten nun:

$$\det\big(V(x_1, \ldots, x_n)\big) = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i).$$

Dazu wenden wir geschickt elementare Zeilenumformungen auf die transponierte Matrix $V(x_1, \ldots, x_n)^{\mathrm{tr}}$ an: Ziehe zuerst das x_1-Fache der $(n-1)$-ten Zeile von der n-ten Zeile ab, dann das x_1-Fache der $(n-2)$-ten Zeile von der $(n-1)$-ten Zeile und so fort. Dann ist

$$\det\big(V(x_1, \ldots, x_n)\big) = \det\big(V(x_1, \ldots, x_n)^{\mathrm{tr}}\big) = \det\left(\begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1^2 & x_2^2 & \ldots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-2} & x_2^{n-2} & \ldots & x_n^{n-2} \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{bmatrix}\right)$$

$$= \det\left(\left[\begin{array}{c|ccc} 1 & 1 & \ldots & 1 \\ \hline 0 & x_2 - x_1 & \ldots & x_n - x_1 \\ 0 & x_2(x_2 - x_1) & \ldots & x_n(x_n - x_1) \\ \vdots & \vdots & & \vdots \\ 0 & x_2^{n-2}(x_2 - x_1) & \ldots & x_n^{n-2}(x_n - x_1) \end{array}\right]\right) = \det\left(\begin{bmatrix} x_2 - x_1 & \ldots & x_n - x_1 \\ x_2(x_2 - x_1) & \ldots & x_n(x_n - x_1) \\ \vdots & & \vdots \\ x_2^{n-2}(x_2 - x_1) & \ldots & x_n^{n-2}(x_n - x_1) \end{bmatrix}\right),$$

wobei im letzten Schritt Satz 23.4 angewandt wurde. In der letzten Matrix (die nur noch $n-1$ Zeilen und Spalten hat) können wir den Faktor $x_2 - x_1$ aus der ersten Spalte herausziehen, dann den Faktor $x_3 - x_1$ aus der zweiten Spalte und so fort bis zum Faktor $x_n - x_1$ in der letzten Spalte. Damit erhalten wir

$$\det\bigl(V(x_1,\dots,x_n)\bigr) = (x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)\det\left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_2 & x_3 & \dots & x_n \\ \vdots & \vdots & & \vdots \\ x_2^{n-2} & x_3^{n-2} & \dots & x_n^{n-2} \end{bmatrix}\right).$$

Die Matrix auf der rechten Seite ist nun eine (transponierte) Vandermonde-Matrix der Größe $(n-1)\times(n-1)$. Also folgt die Behauptung mit vollständiger Induktion nach n (wobei der Induktionsanfang $n=1$ einfach durch $\det(V(x_1)) = 1$ gegeben ist).

24. *Laplace-Entwicklung und der Satz von Cayley–Hamilton*

Schließlich kommen wir noch zu einer rekursiven Formel zur Berechnung von Determinanten. Sei dazu $n \geqslant 2$ und $A = [a_{ij}]_{1\leqslant i,j\leqslant n} \in M_n(R)$. Seien $1 \leqslant k,l \leqslant n$ fest. Dann sei $A^{kl} \in M_{n-1}(R)$ die Matrix, die aus A entsteht, wenn man die k-Zeile und l-te Spalte streicht, also

$$A^{kl} = \begin{bmatrix} a_{1,1} & \dots & a_{1,l-1} & a_{1,l+1} & \dots & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k-1,1} & \dots & a_{k-1,l-1} & a_{k-1,l+1} & \dots & a_{k-1,n} \\ a_{k+1,1} & \dots & a_{k+1,l-1} & a_{k+1,l+1} & \dots & a_{k+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,l-1} & a_{n,l+1} & \dots & a_{n,n} \end{bmatrix} \in M_{n-1}(R).$$

Wir definieren dann eine neue Matrix $\tilde{A} = [\tilde{a}_{kl}]_{1\leqslant k,l\leqslant n} \in M_n(R)$ durch

$$\tilde{a}_{kl} := (-1)^{k+l}\det(A^{lk}) \qquad \text{für } 1 \leqslant k,l \leqslant n.$$

Diese Matrix $\tilde{A}$ heißt auch die ***Adjunkte*** zu A (englisch: "adjugate" oder "adjunct") und wird manchmal mit $\operatorname{adj}(A)$ bezeichnet. Der folgende Satz zeigt, dass man $\det(A)$ rekursiv mit Hilfe der Einträge der Adjunkten berechnen kann.

Satz 24.1 (Laplace-Entwicklung). *Mit obigen Bezeichnungen gilt* $A\cdot\tilde{A} = \tilde{A}\cdot A = \det(A)I_n$, *wobei* $\tilde{A} = \operatorname{adj}(A)$ *die Adjunkte zu* A *ist. Insbesondere erhält man damit:*

$$\det(A) = \sum_{l=1}^{n}(-1)^{k+l}a_{kl}\det(A^{kl}) \qquad \text{und} \qquad \det(A) = \sum_{k=1}^{n}(-1)^{k+l}a_{kl}\det(A^{kl})$$

(Entwicklung nach der k*-ten Zeile)* *(Entwicklung nach der* l*-ten Spalte).*

Beweis. Seien $a_1,\dots,a_n \in R^{1\times n}$ die Zeilen von A. Für $1 \leqslant l \leqslant n$ sei $e_l \in R^{1\times n}$ der l-te Zeilen-Standardvektor; es gilt dann also $a_j = \sum_{l=1}^n a_{jl}e_l$ für $1 \leqslant j \leqslant n$. Für $1 \leqslant k,l \leqslant n$ sei

$B^{kl} \in M_n(R)$ die Matrix mit Zeilen $a_1, \dots, a_{k-1}, e_l, a_{k+1}, \dots, a_n$, also

$$B^{kl} = \begin{bmatrix} a_{1,1} & \dots & a_{1,l-1} & a_{1,l} & a_{1,l+1} & \dots & a_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{k-1,1} & \dots & a_{k-1,l-1} & a_{k-1,l} & a_{k-1,l+1} & \dots & a_{k-1,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{k+1,1} & \dots & a_{k+1,l-1} & a_{k+1,l} & a_{k+1,l+1} & \dots & a_{k+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,l-1} & a_{n,l} & a_{n,l+1} & \dots & a_{n,n} \end{bmatrix} \in M_n(R).$$

Wenden wir nacheinander die $k-1$ Transpositionen $\tau_{k-1,k}, \tau_{k-2,k-1}, \dots, \tau_{1,2}$ auf die Zeilen von B^{kl} an, so wird einfach die k-te Zeile (die aus e_l besteht) in die erste Zeile nach oben versetzt. Wendet man anschließend die $l-1$ Transpositionen $\tau_{l-1,l}, \tau_{l-2,l-1}, \dots, \tau_{12}$ auf die Spalten an, so erhält man am Ende die Matrix

$$\begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ a_{1,l} & a_{1,1} & \dots & a_{1,l-1} & a_{1,l+1} & \dots & a_{1,n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{k-1,l} & a_{k-1,1} & \dots & a_{k-1,l-1} & a_{k-1,l+1} & \dots & a_{k-1,n} \\ a_{k+1,l} & a_{k+1,1} & \dots & a_{k+1,l-1} & a_{k+1,l+1} & \dots & a_{k+1,n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{n,l} & a_{n,1} & \dots & a_{n,l-1} & a_{n,l+1} & \dots & a_{nn} \end{bmatrix} = \left[\begin{array}{c|c} 1 & 0 \ \dots \ 0 \\ \hline * & \\ \vdots & A^{kl} \\ * & \end{array}\right] \in M_n(R).$$

Mit den obigen Regeln zur Berechnung von Determinanten folgt

$$\det(B^{kl}) = (-1)^{k-1}(-1)^{l-1}\det(A_{kl}) = (-1)^{k+l}\det(A^{kl}) = \tilde{a}_{lk}.$$

Damit können wir nun wie folgt argumentieren. Sei $1 \leqslant j \leqslant n$ und $C \in M_n(R)$ die Matrix mit Zeilen $a_1, \dots, a_{k-1}, a_j, a_{k+1}, \dots, a_n$. Weil det linear in der k-ten Zeile ist, gilt

$$\begin{aligned} \det(C) &= \det\big(a_1, \dots, a_{k-1}, a_j, a_{k+1}, \dots, a_n\big) = \det\Big(a_1, \dots, a_{k-1}, \sum_{l=1}^{n} a_{jl}e_l, a_{k+1}, \dots, a_n\Big) \\ &= \sum_{l=1}^{n} a_{jl}\det(a_1, \dots, a_{k-1}, e_l, a_{k+1}, \dots, a_n) = \sum_{l=1}^{n} a_{jl}\det(B^{kl}) \\ &= \sum_{l=1}^{n} (-1)^{k+l} a_{jl}\det(A^{kl}) = \sum_{l=1}^{n} a_{jl}\tilde{a}_{lk} = (A \cdot \tilde{A})_{jk}. \end{aligned}$$

Ist $j \neq k$, so sind zwei Zeilen von C gleich, also $\det(C) = 0$. Ist $j = k$, so ist $C = A$. Also:

$$\sum_{l=1}^{n} a_{jl}\tilde{a}_{lk} = \begin{cases} \det(A) & \text{falls } j = k, \\ 0 & \text{falls } j \neq k, \end{cases}$$

d.h., es gilt $A \cdot \tilde{A} = \det(A)I_n$. Insbesondere gilt damit die Formel zur Entwicklung nach der k-ten Zeile. — Nun beachte: Die Gleichheit $A \cdot \operatorname{adj}(A) = \det(A)I_n$ gilt für alle Matrizen $A \in M_n(K)$, also auch für die Matrix A^{tr} und damit $A^{\mathrm{tr}} \cdot \operatorname{adj}(A^{\mathrm{tr}}) = \det(A^{\mathrm{tr}})I_n$. Jetzt beachte noch $\det(A^{\mathrm{tr}}) = \det(A)$ und $\operatorname{adj}(A^{\mathrm{tr}}) = \big(\operatorname{adj}(A)\big)^{\mathrm{tr}}$. Transponieren wir die Gleichung

$A^{tr}\cdot\mathrm{adj}(A^{tr}) = \det(A^{tr})I_n$, so folgt also auch $\mathrm{adj}(A)\cdot A = \det(A)I_n$, und dies ergibt schließlich die Formel zur Entwicklung nach der l-ten Spalte von A. □

Beispiel Entwicklung nach der 1. Spalte bzw. 3. Zeile:

$$\begin{aligned}
\det\left(\begin{bmatrix}2&1&3\\4&2&5\\1&6&9\end{bmatrix}\right) &= 2\cdot\det\left(\begin{bmatrix}2&5\\6&9\end{bmatrix}\right) - 4\cdot\det\left(\begin{bmatrix}1&3\\6&9\end{bmatrix}\right) + 1\cdot\det\left(\begin{bmatrix}1&3\\2&5\end{bmatrix}\right)\\
&= 2\cdot(18-30) - 4\cdot(9-18) + 1\cdot(5-6) = 11,\\
\text{oder (3. Zeile)} \quad &= 1\cdot\det\left(\begin{bmatrix}1&3\\2&5\end{bmatrix}\right) - 6\cdot\det\left(\begin{bmatrix}2&3\\4&5\end{bmatrix}\right) + 9\cdot\det\left(\begin{bmatrix}2&1\\4&2\end{bmatrix}\right)\\
&= 1\cdot(5-6) - 6\cdot(10-12) + 9\cdot(4-4) = 11.
\end{aligned}$$

In manchen Büchern wird die Formel aus der Laplace-Entwicklung als *Definition* der Determinante benutzt. (Dann muss man am Ende zeigen, dass auch die Leibniz-Formel gilt.)

Folgerung 24.2. *Sei $R = K$ ein Körper und $A \in M_n(K)$. Genau dann ist A invertrierbar, wenn $\det(A) \neq 0$ gilt. Ist $\det(A) \neq 0$, so ist $A^{-1} = \det(A)^{-1}\tilde{A}$ und $\det(A^{-1}) = \det(A)^{-1}$.*

Beweis. Sei A invertierbar. Mit der Produkregel folgt $\det(A)\det(A^{-1}) = \det(A\cdot A^{-1}) = \det(I_n) = 1$ und damit $\det(A) \neq 0$ und $\det(A^{-1}) = \det(A)^{-1}$. Sei nun $\det(A) \neq 0$ und setze $B := \det(A)^{-1}\tilde{A}$. Nach Satz 24.1 gilt $A\cdot B = I_n$, also $B = A^{-1}$. □

Sei weiterhin K ein Körper. Wie in §21 definieren wir für eine Matrix $A \in M_n(K)$ das ***charakteristische Polynom*** $\chi_A := \det(A - XI_n) \in K[X]$. Wir haben bereits gesehen:

$$\chi_A \neq \underline{0}, \quad \mathrm{Grad}(\chi_A) = n \quad \text{und der Leitkoeffizient von } \chi_A \text{ ist gleich } (-1)^n.$$

Mit den obigen Ergebnissen können wir nun weitere Eigenschaften zeigen.

Lemma 24.3. *Sind $A, B \in M_n(K)$ ähnlich (d.h., es gibt $T \in M_n(K)$ invertierbar mit $B = T^{-1}\cdot A\cdot T$), so gilt $\chi_A = \chi_B$, also auch $\det(A) = \det(B)$ und $\mathrm{Spur}(A) = \mathrm{Spur}(B)$.*

Beweis. Es gilt $T^{-1}\cdot(A - XI_n)\cdot T = T^{-1}\cdot A\cdot T - XT^{-1}\cdot T = B - XI_n$. Mit der Produktregel folgt:

$$\chi_B = \det\big(T^{-1}\cdot(A - XI_n)\cdot T\big) = \det(T^{-1}) * \chi_A * \det(T) = \det(T^{-1}\cdot T) * \chi_A = \det(I_n) * \chi_A = \chi_A.$$

Nach Lemma 21.11 und Lemma 21.12 kommen $\mathrm{Spur}(A)$ und $\det(A)$ als Koeffizienten von χ_A vor; eine analoge Aussage gilt auch für $\mathrm{Spur}(B)$ und $\det(B)$. Aus $\chi_A = \chi_B$ folgt also insbesondere $\det(A) = \det(B)$ und $\mathrm{Spur}(A) = \mathrm{Spur}(B)$. □

Die folgende Aussage liefert eine neue Charakterisierung von Eigenwerten: Die Eigenwerte von $A \in M_n(K)$ sind genau die Nullstellen von $\chi_A \in K[X]$.

Lemma 24.4. *Sei $\lambda \in K$. Dann gilt:*

$$\lambda \textit{ Eigenwert von } A \iff A - \lambda I_n \textit{ nicht invertierbar} \iff \chi_A(\lambda) = \det(A - \lambda I_n) = 0.$$

Beweis. Für die erste Äquivalenz siehe Kapitel III, Bemerkung 14.1. Nach Lemma 21.11 gilt $\chi_A(\lambda) = \det(A - \lambda I_n)$. Die zweite Äquivalenz folgt also sofort aus Folgerung 24.2. □

Der folgende Satz ist eine berühmte Aussage der Matrix-Theorie.

Satz 24.5 (Cayley–Hamilton 1853/58)**.** *Für eine Matrix* $A \in M_n(K)$ *gilt* $\chi_A(A) = 0_{n\times n}$*; insbesondere ist* $\mu_A \mid \chi_A$ *(siehe Kapitel* III*, Folgerung* 15.4*) und* $\mathrm{Grad}(\mu_A) \leqslant n$.

Für 2×2 Matrizen kann man dies direkt nachrechnen; für $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(K)$ gilt

$$\chi_A = \det\left(\begin{bmatrix} a - X & b \\ c & d - X \end{bmatrix}\right) = X^2 - (a+d)X + ad - bc = X^2 - \mathrm{Spur}(A)X + \det(A).$$

Damit erhalten wir $\chi_A(A) = A^2 - (a+b)A + (ad - bc)I_2 = \ldots = 0_{2\times 2}$.
Für den allgemeinen Fall braucht man einen ziemlich cleveren "Trick".

Beweis. (Frobenius 1878) Sei $F = [f_{ij}]_{1\leqslant i,j\leqslant n} = A - XI_n \in M_n\big(K[X]\big)$ und $\chi_A = \det(F)$. Mit Laplace–Entwicklung gilt dann $F \cdot \mathrm{adj}(F) = \det(F)I_n$; mit $\mathrm{adj}(F) = [\tilde{f}_{ij}] \in M_n\big(K[X]\big)$ also:

$$\sum_{l=1}^{n} \tilde{f}_{lj} * f_{il} = \sum_{l=1}^{n} f_{il} * \tilde{f}_{lj} = \big(F \cdot \mathrm{adj}(F)\big)_{ij} = \begin{cases} \chi_A & \text{falls } i = j, \\ 0 & \text{falls } i \neq j. \end{cases}$$

(Für die erste Gleichheit haben wir nur die Reihenfolge der Faktoren $\tilde{f}_{lj}$ und f_{il} getauscht.)
Wir setzen A ein und erhalten: $\displaystyle\sum_{l=1}^{n} \tilde{f}_{lj}(A) \cdot f_{il}(A) = \begin{cases} \chi_A(A) & \text{falls } i = j, \\ 0_{n\times n} & \text{falls } i \neq j. \end{cases}$

Für $1 \leqslant i \leqslant n$ sei $e_i \in K^n$ der i-te Standard-Basisvektor. Für festes $1 \leqslant j \leqslant n$ gilt dann

$$\begin{aligned} \chi_A(A) \cdot e_j &= \sum_{i=1}^{n} \underbrace{\Big(\sum_{l=1}^{n} \tilde{f}_{lj}(A) \cdot f_{il}(A)\Big)}_{=0_{n\times n} \text{ falls } i \neq j} \cdot e_i = \sum_{i=1}^{n} \sum_{l=1}^{n} \tilde{f}_{lj}(A) \cdot \big(f_{il}(A) \cdot e_i\big) \\ &= \sum_{l=1}^{n} \sum_{i=1}^{n} \tilde{f}_{lj}(A) \cdot \big(f_{il}(A) \cdot e_i\big) = \sum_{l=1}^{n} \tilde{f}_{lj}(A) \cdot \Big(\sum_{i=1}^{n} f_{il}(A) \cdot e_i\Big). \end{aligned}$$

Betrachten wir die letzte Summe. Für $i = l$ ist $f_{ll} = a_{ll} - X$ also $f_{ll}(A) \cdot e_l = a_{ll}e_l - A \cdot e_l$. Für $i \neq l$ ist $f_{il} = a_{il}$ also $f_{il}(A) \cdot e_i = a_{il}e_i$. Damit erhalten wir

$$\sum_{i=1}^{n} f_{il}(A) \cdot e_i = (a_{ll}e_l - A \cdot e_l) + \sum_{1\leqslant i\leqslant n, i\neq l} a_{il}e_i = \Big(\sum_{i=1}^{n} a_{il}e_i\Big) - A \cdot e_l = 0_n$$

(weil $A \cdot e_l$ genau die l-te Spalte von A ist). Also ist $\chi_A(A) \cdot e_j = 0_n$ für $1 \leqslant j \leqslant n$. Damit sind alle Spalten von $\chi_A(A)$ gleich 0_n, also $\chi_A(A) = 0_{n\times n}$. □

Beispiel 24.6. Sei K ein Körper, $d \in \mathbb{N}$ und $f \in K[X]$ ein beliebiges normiertes Polynom mit $\mathrm{Grad}(f) = d$; also $f = a_0 + a_1X + \ldots + a_dX^d \in K[X]$ mit $a_d = 1$. In Kapitel III, §15, haben wir die (Frobenius-)***Begleitmatrix*** zu f definiert als

$$A_f := \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \dots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{d-1} \end{bmatrix} \in M_d(K).$$

Nach Kapitel III, Lemma 15.10, gilt $\mu_{A_f} = f$. Nach dem Satz von Cayley–Hamilton ist $\mu_{A_f} \mid \chi_{A_f}$. Wegen $\mathrm{Grad}(\mu_{A_f}) = d = \mathrm{Grad}(\chi_{A_f})$ folgt $\chi_{A_f} = c\mu_{A_f}$ mit einem $0 \neq c \in K$. Weil $(-1)^d$ der Leitkoeffizient von χ_A ist, muss $c = (-1)^d$ gelten, also $\chi_{A_f} = (-1)^d f$.

Es ist eine gute Übungsaufgabe, direkt die Determinante von $A_f - XI_d \in M_d(K[X])$ zu berechnen, ohne den Satz von Cayley–Hamilton zu benutzen. (Hinweis: Siehe noch einmal Beispiel 23.6.)

25. ***Beweis des Fundamentalsatzes der Algebra*** (Selbststudium)

In Kapitel II, Abschnitt 10, haben wir den ***Fundamentalsatz der Algebra*** formuliert:

Sei $n \in \mathbb{N}$ *und* $\underline{0} \neq f \in \mathbb{C}[X]$ *ein Polynom mit* $\mathrm{Grad}(f) = n$. *Dann zerfällt* f *in Linearfaktoren über* $\mathbb{C}$, *d.h., es gibt* $a, z_1, \dots, z_n \in \mathbb{C}$ *mit* $a \neq 0$ *und* $f = a(X - z_1) * (X - z_2) * \dots * (X - z_n)$. *Insbesondere hat jedes nicht-konstante Polynom* $f \in \mathbb{C}[X]$ *eine Nullstelle in* $\mathbb{C}$.

Um diesen Satz zu beweisen, genügt es zu zeigen, dass jedes nicht-konstante $f \in \mathbb{C}[X]$ eine Nullstelle in $\mathbb{C}$ hat. Denn ist $f(z) = 0$ für ein $z \in \mathbb{C}$, so ist $f = (X - z) * f_1$ mit $\mathrm{Grad}(f_1) = \mathrm{Grad}(f_1) - 1$ und wir können mit f_1 fortfahren. Weiterhin können wir annehmen, dass f normiert ist. (Denn die Nullstellen eines Polynoms ändern sich nicht, wenn wir das Polynom normieren, also mit einer Konstante ungleich 0 multiplizieren).
Startpunkt für viele Beweise ist die folgende Aussage, die den Zwischenwertsatz der reellen Analysis benutzt (der sicherlich in der Vorlesung Analysis 1 behandelt wurde).

Lemma 25.1. *Sei* $f \in \mathbb{R}[X]$ *nicht-konstant mit* $2 \nmid \mathrm{Grad}(f)$. *Dann hat* f *eine Nullstelle in* $\mathbb{R}$.

Beweis. Wir können annehmen, dass f normiert ist. Sei $\dot{f} \colon \mathbb{R} \to \mathbb{R}$ die durch f definierte Polynomfunktion; diese ist stetig. Da $\mathrm{Grad}(f)$ ungerade (und f normiert) ist, gilt $\lim_{x \to +\infty} \dot{f}(x) = +\infty$ und $\lim_{x \to -\infty} \dot{f}(x) = -\infty$. Also gibt es auch $a < b$ in $\mathbb{R}$ mit $\dot{f}(a) < 0$ und $\dot{f}(b) > 0$. Nach dem Zwischenwertsatz hat $\dot{f}$ auf dem abgeschlossenen Intervall $[a, b]$ eine Nullstelle. □

Die Kunst besteht anschließend darin, allein aus obiger Aussage die Existenz von komplexen Nullstellen für beliebige nicht-konstante Polynome $f \in \mathbb{C}[X]$ herzuleiten. Wir wollen hier einen Beweis geben, der nur mit Methoden der Linearen Algebra auskommt (also linearen Abbildungen, Determinanten, Eigenwerten, etc.)[1]. Zunächst einige allgemeine Aussagen,

[1] Dies beruht auf dem Artikel: H. Derksen, The Fundamental Theorem of Algebra and Linear Algebra, Amer. Math. Monthly **110**, No. 7 (2003), 620–623.

die über beliebigen Körpern K gelten. Wir haben bisher Eigenwerte und Eigenvektoren für Matrizen definiert. Das geht natürlich auch etwas allgemeiner für Endomorphismen.

Bemerkung 25.2. Sei V ein beliebiger Vektorraum über einem Körper K. Sei $\varphi \in \mathrm{End}(V)$, also $\varphi\colon V \to V$ eine lineare Abbildung. Ein Vektor $v \in V$ heißt ***Eigenvektor*** von φ, wenn $v \neq 0_V$ gilt und es ein $\lambda \in K$ gibt mit $\varphi(v) = \lambda v$; in diesem Fall heißt λ der zu v gehörige ***Eigenwert***. Sei nun $1 \leqslant \dim V < \infty$. Ist B eine Basis von V und $A = M_B(\varphi) \in M_n(K)$ die darstellende Matrix von φ bezüglich B, so sieht man sofort mit Kapitel IV, Lemma 20.2, dass $x := M_B(v) \in K^n$ ein Eigenvektor von A ist mit Eigenwert λ (und umgekehrt). Um die Eigenwerte von φ zu bestimmen, wählt man also eine Basis B von V, stellt $A = M_B(\varphi) \in M_n(K)$ auf, und bestimmt schließlich die Eigenwerte von A.

Lemma 25.3. *Sei K ein Körper und $n \in \mathbb{N}$. Die folgenden beiden Aussagen sind äquivalent.*

(a) *Jedes nicht-konstante normierte $f \in K[X]$ mit $\mathrm{Grad}(f) = n$ hat eine Nullstelle in K.*
(b) *Jeder Endomorphismus $\varphi \in \mathrm{End}(V)$, wobei V ein K-Vektorraum mit $n = \dim V$ ist, besitzt einen Eigenwert in K (d.h., es gibt ein $0_V \neq v \in V$ und $\lambda \in K$ mit $\varphi(v) = \lambda v$).*

Beweis. “(a) $\Rightarrow$ (b)” Sei B eine Basis von V und $A := M_B(\varphi) \in M_n(K)$ die darstellende Matrix von φ. Sei $\chi_A \in K[X]$ das charakteristische Polynom von A. Dann ist $\chi_A = (-1)^n f$ wobei $f \in K[X]$ normiert ist mit $\mathrm{Grad}(f) = n$ (siehe §24). Nach (a) hat f eine Nullstelle $\lambda \in K$. Dann ist λ ein Eigenwert von A, also auch von φ (siehe Bemerkung 25.2).
“(b) $\Rightarrow$ (a)” Sei $f \in K[X]$ nicht-konstant, normiert mit $n := \mathrm{Grad}(f)$. Sei $A = A_f \in M_n(K)$ die zugehörige Begleitmatrix (siehe Beispiel 24.6). Zur Erinnerung: Es gilt $\mu_A = f$ und $\chi_A = (-1)^n f$. Sei $\varphi_A\colon K^n \to K^n$ die durch A definierte lineare Abbildung. Nach (b) hat φ_A einen Eigenwert, d.h., es gibt eine Nullstelle von $f = \mu_A$ in K. □

Wir beschreiben nun eine allgemeine Konstruktion, mit der man aus einer gegebenen Matrix zwei neue Endomorphismen erhält. (Dies ist der entscheidende Trick von Derksens Beweis.)

Beispiel 25.4. Sei K ein Körper und $\mathrm{Sym}_n(K) := \{S \in M_n(K) \mid S^{\mathrm{tr}} = S\}$ die Menge aller symmetrischen Matrizen. Dies ist ein Teilraum von $M_n(K)$ mit $\dim \mathrm{Sym}_n(K) = n(n+1)/2$ (Beweis Übung oder selbst). Sei nun eine beliebige Matrix $A \in M_n(K)$ gegeben. Setze:

$$\varphi_1(S) := A \cdot S + S \cdot A^{\mathrm{tr}} \in M_n(K) \qquad \text{und} \qquad \varphi_2(S) := A \cdot S \cdot A^{\mathrm{tr}} \in M_n(K)$$

für $S \in M_n(K)$. Man sieht sofort, dass φ_1, φ_2 linear sind. Mit den üblichen Regeln für das Transponieren von Matrizen sieht man außerdem sehr leicht, dass für $S \in \mathrm{Sym}_n(K)$ auch die Matrizen $\varphi_1(S)$ und $\varphi_2(S)$ wieder symmetrisch sind. Also erhalten wir $\varphi_1, \varphi_2 \in \mathrm{End}\big(\mathrm{Sym}_n(K)\big)$. Weiterhin gilt $\varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1$, denn

$$(\varphi_1 \circ \varphi_2)(S) = \varphi_1\big(A \cdot S \cdot A^{\mathrm{tr}}\big) = A \cdot (A \cdot S \cdot A^{\mathrm{tr}}) + (A \cdot S \cdot A^{\mathrm{tr}}) \cdot A^{\mathrm{tr}},$$

$$(\varphi_2 \circ \varphi_1)(S) = \varphi_2\big(A \cdot S + S \cdot A^{\mathrm{tr}}\big) = A \cdot (A \cdot S + S \cdot A^{\mathrm{tr}}) \cdot A^{\mathrm{tr}},$$

und die beiden rechten Seiten sind gleich.

Bemerkung 25.5. Sei V ein K-Vektorraum. Seien $\varphi_1, \varphi_2 \in \mathrm{End}(V)$ gegeben mit $\varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1$. Dann gilt $\varphi_2(\mathrm{Kern}(\varphi_1)) \subseteq \mathrm{Kern}(\varphi_1)$ und $\varphi_2(\mathrm{Bild}(\varphi_1)) \subseteq \mathrm{Bild}(\varphi_1)$.
Dazu: Ist $v \in \mathrm{Kern}(\varphi_1)$, so ist $\varphi_1(v) = 0_V$ und damit $\varphi_1(\varphi_2(v)) = \varphi_2(\varphi_1(v)) = 0_V$, also auch $\varphi_2(v) \in \mathrm{Kern}(\varphi_1)$. Ist $w \in \mathrm{Bild}(\varphi_1)$, so ist $w = \varphi_1(v)$ für ein $v \in V$ und damit $\varphi_2(w) = \varphi_2(\varphi_1(v)) = \varphi_1(\varphi_2(v)) \in \mathrm{Bild}(\varphi_1)$.

Lemma 25.6. *Seien* $A \in M_n(K)$ *und* $\varphi_1, \varphi_2 \in \mathrm{End}\big(\mathrm{Sym}_n(K)\big)$ *wie in Beispiel 25.4. Nehmen wir an, es gibt einen gemeinsamen Eigenvektor für* φ_1 *und* φ_2 *in* $\mathrm{Sym}_n(K)$*. Dann gibt es eine Matrix* $0_{n\times n} \neq S \in \mathrm{Sym}_n(K)$ *und* $\lambda, \mu \in K$ *mit* $(A^2 - \lambda A + \mu I_n) \cdot S = 0_{n\times n}$*.*

Beweis. Nach Voraussetzung gibt es ein $0_{n\times n} \neq S \in \mathrm{Sym}_n(K)$ und $\lambda, \mu \in K$ mit $\varphi_1(S) = \lambda S$ und $\varphi_2(S) = \mu S$. Dann gilt $A \cdot S + S \cdot A^{\mathrm{tr}} = \lambda S$ und $A \cdot S \cdot A^{\mathrm{tr}} = \mu S$. Multiplikation der ersten Gleichung mit A von links ergibt $A^2 \cdot S + A \cdot S \cdot A^{\mathrm{tr}} = \lambda\, A \cdot S$. Durch Einsetzen von $A \cdot S \cdot A^{\mathrm{tr}} = \mu S$ erhalten wir $A^2 \cdot S + \mu S = \lambda\, A \cdot S$ und damit $(A^2 - \lambda A + \mu I_n) \cdot S = 0_{n\times n}$. □

Betrachte nun folgende Aussage für $m \in \mathbb{N}$, $m \geqslant 2$:

$\mathcal{P}(K, m)$: *Jedes nicht-konstante, normierte* $f \in K[X]$ *mit* $m \nmid \mathrm{Grad}(f)$ *hat eine Nullstelle in* K.

Zum Beispiel ist Lemma 25.1 gerade die Aussage $\mathcal{P}(\mathbb{R}, 2)$.

Lemma 25.7. *Es gelte* $\mathcal{P}(K, m)$ *für ein* $m \in \mathbb{N}$*,* $m \geqslant 2$*. Seien* V *ein* K*-Vektorraum und* $\varphi_1, \varphi_2 \in \mathrm{End}(V)$ *mit* $\varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1$*. Ist* $m \nmid \dim V < \infty$*, so gibt es einen gemeinsamen Eigenvektor in* V *für* φ_1 *und* φ_2*.*

Beweis. Wir gehen mit vollständiger Induktion nach $\dim V$ vor. Ist $\dim V = 1$, so ist nichts zu zeigen. (Jedes $0_V \neq v \in V$ ist gemeinsamer Eigenvektor von φ_1 und φ_2.) Sei nun $\dim V > 1$. Ist $m \mid \dim V$, so ist nichts zu zeigen. Sei also $m \nmid \dim V$. Da $\mathcal{P}(K, m)$ gilt, gibt es nach Lemma 25.3 ("(a) ⇒ (b)") einen Eigenwert $\lambda \in K$ von φ_1. Setze $\varphi := \varphi_1 - \lambda\, \mathrm{id}_V \in \mathrm{End}(V)$ sowie $U := \mathrm{Kern}(\varphi)$ und $W := \mathrm{Bild}(\varphi)$. Beachte: Für $i = 1, 2$ gilt dann auch $\varphi \circ \varphi_i = \varphi_i \circ \varphi$, also jeweils $\varphi_i(U) \subseteq U$ und $\varphi_i(W) \subseteq W$ nach Bemerkung 25.5. Durch Einschränkung erhalten wir also Endomorphismen $\varphi_i|_U \in \mathrm{End}(U)$ und $\varphi_i|_W \in \mathrm{End}(W)$, die für $i = 1, 2$ jeweils wieder vertauschbar sind. Nun gilt nach dem Kern-Bild-Dimensionssatz $\dim V = \dim U + \dim W$. Wegen $m \nmid \dim V$ ist also $m \nmid \dim U$ oder $m \nmid \dim W$. Sei $0_V \neq v \in V$ ein zu λ gehöriger Eigenvektor von φ_1. Dann ist $\varphi(v) = 0_V$, also $v \in U$ und damit $\dim U > 0$. Wegen $\dim V = \dim U + \dim W$ ist also stets $\dim W < \dim V$. Dies führt zu den drei Fällen:

1) $\dim U = \dim V$, $W = \{0_V\}$.

2) $\dim U < \dim V$, $\dim W < \dim V$ und $m \nmid \dim U$.

3) $\dim U < \dim V$, $\dim W < \dim V$ und $m \nmid \dim W$.

Im Fall 1) ist $\mathrm{Kern}(\varphi) = V$, also $\varphi_1 = \lambda \mathrm{id}_V$. Da $\mathcal{P}(K, m)$ gilt, gibt es nach Lemma 25.3 einen Eigenwert von φ_2 in K. Sei $0_V \neq w \in V$ ein zugehöriger Eigenvektor von φ_2. Dann ist automatisch auch $\varphi_1(w) = \lambda w$; also ist w ein gemeinsamer Eigenvektor von φ_1 und φ_2.
Im Fall 2) können wir Induktion auf $\varphi_1|_U$ und $\varphi_2|_U$ anwenden; es gibt also einen gemeinsamen Eigenvektor von $\varphi_1|_U$ und $\varphi_2|_U$, und dieser ist natürlich auch ein gemeinsamer Eigenvektor von φ_1 und φ_2. Der Fall 3) geht analog mit $\varphi_1|_W$ und $\varphi_2|_W$. □

Nach diesen Vorbereitungen erhalten wir nun den ersten Fortschritt gegenüber Lemma 25.1. Zunächst noch eine Bezeichnung: Ist $g \in \mathbb{C}[X]$ beliebig, so sei $\overline{g} \in \mathbb{C}[X]$ das Polynom, das entsteht, wenn man komplexe Konjugation auf alle Koeffizienten von g anwendet. Man sieht sofort: $\overline{g(z)} = \overline{g}(\overline{z})$ für alle $z \in \mathbb{C}$; außerdem $\overline{g_1 * g_2} = \overline{g}_1 * \overline{g}_2$ für alle $g_1, g_2 \in \mathbb{C}[X]$.

Lemma 25.8. *Sei* $f \in \mathbb{C}[X]$ *nicht-konstant mit* $2 \nmid \mathrm{Grad}(f)$. *Dann hat* f *eine Nullstelle in* $\mathbb{C}$.

Beweis. (Der folgende Beweis ist etwas anders als im Artikel von Derksen.) Wir können annehmen, dass f normiert ist. Sei $n := \mathrm{Grad}(f) \geqslant 1$ und $g := f * \overline{f} \in \mathbb{C}[X]$ mit $\mathrm{Grad}(g) = 2n$, wobei n ungerade ist. Mit obigen Regeln folgt $\overline{g} = \overline{f * \overline{f}} = \overline{f} * \overline{\overline{f}} = \overline{f} * f = g$, also $g \in \mathbb{R}[X]$.

Wir zeigen zuerst, dass g eine Nullstelle in $\mathbb{C}$ besitzt. Sei $A = A_g \in M_{2n}(\mathbb{R})$ die Begleitmatrix von g. Wir bilden $\varphi_1, \varphi_2 \in \mathrm{End}\big(\mathrm{Sym}_{2n}(\mathbb{R})\big)$ wie in Beispiel 25.4. Nun ist $\dim \mathrm{Sym}_{2n}(\mathbb{R}) = 2n(2n+1)/2 = n(2n+1)$ ungerade. Da die Aussage $\mathcal{P}(\mathbb{R}, 2)$ gilt, gibt es nach Lemma 25.7 einen gemeinsamen Eigenvektor von φ_1 und φ_2 in $\mathrm{Sym}_{2n}(\mathbb{R})$. Damit ist die Voraussetzung von Lemma 25.6 erfüllt. Es gibt also eine Matrix $0_{n\times n} \neq S \in \mathrm{Sym}_n(\mathbb{R})$ und $\lambda, \mu \in \mathbb{R}$ mit $(A^2 - \lambda A + \mu I_n) \cdot S = 0_{n\times n}$. Mit der üblichen Lösungsformel für quadratische Gleichungen (Stichwort: quadratische Ergänzung) gibt es $z_1, z_2 \in \mathbb{C}$ mit $X^2 - \lambda X + \mu = (X - z_1) * (X - z_2) \in \mathbb{C}[X]$. Dann gilt auch $(A - z_1 I_n) \cdot (A - z_2 I_n) \cdot S = 0_{n\times n}$, nun in $M_n(\mathbb{C})$. Wäre $\det(A - z_i I_n) \neq 0$ für $i = 1, 2$, so wären $A - z_i I_n$ invertierbar für $i = 1, 2$ und damit $S = 0_{n\times n}$, Widerspruch. Also ist $\det(A - z I_n) = 0$ für ein $z \in \{z_1, z_2\}$ und damit z ein Eigenwert von A, d.h., $z \in \mathbb{C}$ ist eine Nullstelle von $g = (-1)^n \chi_A$.

Nun ist $0 = g(z) = (f * \overline{f})(z) = f(z)\overline{f}(z)$. Also folgt $f(z) = 0$ oder $\overline{f}(z) = 0$; im letzteren Fall ist auch $f(\overline{z}) = \overline{\overline{f}(z)} = 0$. Also hat f auf jeden Fall eine Nullstelle in $\mathbb{C}$. □

Satz 25.9. *Die Aussage* $\mathcal{P}(\mathbb{C}, 2^d)$ *gilt für alle* $d \in \mathbb{N}$. *Insbesondere folgt der Fundamentalsatz der Algebra.*

Beweis. Wir benutzen eine vollständige Induktion nach d. Der Induktionsanfang $d = 1$ ist genau die Aussage von Lemma 25.8. Sei nun $d \geqslant 1$ und nehmen wir an, dass $\mathcal{P}(\mathbb{C}, 2^d)$ bereits gezeigt sei. Sei $f \in \mathbb{C}[X]$ nicht-konstant, normiert mit $2^{d+1} \nmid n := \operatorname{Grad}(f)$. Wir müssen zeigen, dass f eine Nullstelle in $\mathbb{C}$ hat. Dazu: Es gilt $n = 2^e \tilde{n}$ mit $0 \leqslant e \leqslant d$ und $\tilde{n} \geqslant 1$ ungerade. Für $e = 0$ ist $n = \tilde{n}$ ungerade, also wissen wir bereits nach Lemma 25.8, dass f eine Nullstelle in $\mathbb{C}$ hat. Wir können also $1 \leqslant e \leqslant d$ annehmen. Sei $A = A_f \in M_n(\mathbb{C})$ die Begleitmatrix zu f. Wir bilden $\varphi_1, \varphi_2 \in \operatorname{End}\big(\operatorname{Sym}_n(\mathbb{C})\big)$ wie in Beispiel 25.4. Nun ist

$$\dim \operatorname{Sym}_n(\mathbb{C}) = n(n+1)/2 = 2^{e-1}\tilde{n}(2^e\tilde{n}+1).$$

Wegen $1 \leqslant e \leqslant d$ ist also $2^d \nmid \dim \operatorname{Sym}_n(\mathbb{C})$. Da $\mathcal{P}(\mathbb{C}, 2^d)$ gilt, gibt es nach Lemma 25.7 einen gemeinsamen Eigenvektor von φ_1 und φ_2 in $\operatorname{Sym}_n(\mathbb{C})$. Nach Lemma 25.6 gibt es eine Matrix $0_{n\times n} \neq S \in \operatorname{Sym}_n(\mathbb{C})$ und $\lambda, \mu \in \mathbb{C}$ mit $(A^2 - \lambda A + \mu I_n)\cdot S = 0_{n\times n}$. Nach Kapitel II, Bemerkung 10.3, gibt es $z_1, z_2 \in \mathbb{C}$ mit $X^2 - \lambda X + \mu = (X - z_1) * (X - z_2) \in \mathbb{C}[X]$. Wie im obigen Beweis folgt, dass z_1 oder z_2 eine Nullstelle von $f = (-1)^n \chi_A$ ist. Damit gilt $\mathcal{P}(\mathbb{C}, 2^{d+1})$.

Zum Fundamentalsatz: Sei $\underline{0} \neq f \in \mathbb{C}[X]$ normiert, mit $n := \operatorname{Grad}(f) \geqslant 1$. Dann gibt es ein $d \in \mathbb{N}$ mit $2^d \nmid n$. Da $\mathcal{P}(\mathbb{C}, 2^d)$ gilt, hat also hat f eine Nullstelle in $\mathbb{C}$. □

Wir halten noch die folgende Konsequenz für Polynome in $\mathbb{R}[X]$ fest: Man kann diese stets in Faktoren vom Grad 1 oder 2 zerlegen.

Lemma 25.10. *Sei $f \in \mathbb{R}[X]$ nicht-konstant und normiert. Dann gibt es $\alpha_i, \beta_i \in \mathbb{R}$ mit $4\beta_i > \alpha_i^2$ (für $1 \leqslant i \leqslant r$) und $\gamma_j \in \mathbb{R}$ (für $1 \leqslant j \leqslant s$) mit $2r + s = \operatorname{Grad}(f)$ und*

$$f = (X^2 + \alpha_1 X + \beta_1) * \ldots * (X^2 + \alpha_r X + \beta_r) * (X - \gamma_1) * \ldots (X - \gamma_s).$$

Beweis. Wir benutzen vollständige Induktion nach $n = \operatorname{Grad}(f) \geqslant 1$. Ist $n = 1$, so ist nichts zu zeigen. Sei nun $n \geqslant 2$. Nach Satz 25.9 gibt es eine Nullstelle $\lambda \in \mathbb{C}$; es folgt $f = (X - \lambda) * g_1$ mit $g_1 \in \mathbb{C}[X]$. Ist $\lambda \in \mathbb{R}$, so ist auch $g_1 \in \mathbb{R}[X]$ und nach Induktion hat g_1 eine Faktorisierung wie gewünscht. Sei nun $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Wegen $\overline{f} = f$ folgt $0 = \overline{f(\lambda)} = \overline{f}(\overline{\lambda}) = f(\overline{\lambda}) = (\overline{\lambda} - \lambda)g_1(\overline{\lambda})$. Wegen $\lambda \neq \overline{\lambda}$ folgt $g_1(\overline{\lambda}) = 0$ und damit $g_1 = (X - \overline{\lambda}) * g_2$ mit $g_2 \in \mathbb{C}[X]$. Insgesamt ergibt dies $f = f_1 * g_2$ mit $f_1 := (X - \lambda) * (X - \overline{\lambda}) = X^2 + \alpha X + \beta$, wobei $\alpha = -(\lambda + \overline{\lambda}) \in \mathbb{R}$ und $\beta = \lambda\overline{\lambda} \in \mathbb{R}$. (Beachte: Weil $f_1 = X^2 + \alpha X + \beta \in \mathbb{R}[X]$ keine reelle Nullstelle hat, ist $4\beta > \alpha^2$.) Nun ist $f_1 * g_2 = f = \overline{f} = \overline{f}_1 * \overline{g}_2 = f_1 * \overline{g}_2$. Mit Kürzen von f_1 folgt $g_2 = \overline{g}_2 \in \mathbb{R}[X]$, und wir können wieder Induktion auf g_2 anwenden. □

Ab hier Woche 4

Kapitel VI: Normalformen von Matrizen

Am Ende des ersten Semesters haben wir das Normalformen-Problem für Matrizen über einem Körper K formuliert: Ist $A \in M_n(K)$ gegeben, so wird eine invertierbare Matrix $T \in M_n(K)$ gesucht, so dass $A' := T^{-1} \cdot A \cdot T$ eine "möglichst einfache" Gestalt besitzt. Dort wurde auch bereits der Spezialfall der "diagonalisierbaren" Matrizen behandelt. Nun ist allerdings nicht jede Matrix diagonalisierbar; Beispiele für $n = 2$:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (K \text{ beliebig}) \qquad \text{und} \qquad A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (K = \mathbb{Q} \text{ oder } \mathbb{R}).$$

Die erste Matrix hat 0 als einzigen Eigenwert, aber der Eigenraum ist 1-dimensional und A damit nicht diagonalisierbar; die zweite Matrix hat charakteristisches Polynom $\chi_A = X^2 + 1$, also gibt es überhaupt keine Eigenwerte. (Zur Erinnerung: A ist diagonalisierbar genau dann, wenn es eine Basis von K^n gibt, die aus Eigenvektoren von A besteht.) Ziel dieses Kapitels ist es, auch für derartige Fälle eine Lösung des Normalformen-Problems zu entwickeln. Dies zerfällt in zwei größere Teil-Probleme: Einerseits der Fall, in dem χ_A in Linearfaktoren zerfällt, und andererseits der allgemeine Fall (ohne irgendwelche Voraussetzungen an χ_A).

26. ***Zerfallende Polynome und zerfallende Matrizen***

Bevor wir zu Normalformen kommen, stellen wir noch einige weitere Aussagen zu Polynomen bereit, die auch von allgemeinem Interesse sind. In Kapitel IV, §20, haben wir bereits den Begriff eines zerfallenden Polynoms in $K[X]$ definiert. Zur Erinnerung: Sei $\underline{0} \neq f \in K[X]$ mit $d := \text{Grad}(f) \geqslant 0$. Wir bezeichnen mit $Z(f) := \{\lambda \in K \mid f(\lambda) = 0\}$ die Menge der Nullstellen von f. Dann heißt f ein ***zerfallendes Polynom***, wenn es $0 \neq c \in K$ und $z_1, \ldots, z_d \in K$ gibt mit $f = c(X - z_1) * \ldots * (X - z_d)$. (Es gilt also $Z(f) = \{z_1, \ldots, z_d\}$, aber die z_i müssen hier nicht verschieden sein; für $d = 0$ ist $f = c \in K$ konstant und wird als zerfallend deklariert.)

Lemma 26.1. *Sei* $\underline{0} \neq f \in K[X]$ *und* $f = g * h$ *mit* $g, h \in K[X]$. *Ist* f *zerfallend, so sind auch* g *und* h *zerfallend.*

Beweis. (Vollständige Induktion nach $d := \text{Grad}(f) \geqslant 0$.) Ist $d = 0$, so sind f, g, h konstant und es ist nichts zu zeigen. Sei nun $d \geqslant 1$ und $\underline{0} \neq f \in K[X]$ zerfallend mit $d = \text{Grad}(f)$. Schreibe $f = c(X - z_1) * \ldots * (X - z_d)$ wie oben. Dann ist $0 = f(z_1) = g(z_1)h(z_1)$, also $g(z_1) = 0$ oder $h(z_1) = 0$. Nehmen wir an, es sei $g(z_1) = 0$. Nach dem Horner-Schema ist $g = (X - z_1) * g_1$ mit einem $g_1 \in K[X]$. Andererseits ist $f = (X - z_1) * f_1$ mit $f_1 := c(X - z_2) * \ldots * (X - z_d)$. In der Gleichung $(X - z_1) * f_1 = f = g * h = (X - z_1) * g_1 * h$ können wir auf beiden Seiten einen Faktor $X - z_1$ kürzen und erhalten $f_1 := g_1 * h$. Nun ist f_1 zerfallend mit $\text{Grad}(f_1) = d - 1$ (nach

Definition), also folgt mit Induktion, dass g_1 und h zerfallend sind, also auch $g = (X - z_1) * g_1$. Das Argument ist analog für den Fall $h(z_1) = 0$. $\square$

Lemma 26.2. *Gegeben seien* $\underline{0} \neq f \in K[X]$ *und* $\underline{0} \neq g \in K[X]$. *Ist* $Z(f) \cap Z(g) = \varnothing$ *und* f *zerfallend, so gilt* $\mathrm{ggT}(f, g) = 1$.

Beweis. Sei $h := \mathrm{ggT}(f, g)$ und nehmen wir an, es sei $\mathrm{Grad}(h) \geqslant 1$. Wegen $h \mid f$ ist auch h zerfallend (siehe Lemma 26.1). Sei $\lambda \in K$ mit $h(\lambda) = 0$. Wegen $h \mid f$ ist dann auch $f(\lambda) = 0$, und ebenso folgt $g(\lambda) = 0$. Also ist $\lambda \in Z(f) \cap Z(g)$, Widerspruch. $\square$

Sei $\underline{0} \neq f \in K[X]$ beliebig. Für $\lambda \in K$ sei $m_f(\lambda)$ das größte $j \in \mathbb{N}_0$ mit $(X - \lambda)^j \mid f$. Dann heißt $m_f(\lambda)$ die ***Vielfachheit*** von λ (als Nullstelle von f). Es gilt

$$\lambda \in Z(f) \quad \Leftrightarrow \quad m_f(\lambda) \geqslant 1 \quad \Leftrightarrow \quad X - \lambda \mid f.$$

(Die erste Äquivalenz gilt nach Kapitel II, Bemerkung 9.10.) Ist $\lambda \in Z(f)$, so gilt $f = (X - \lambda)^{m_f(\lambda)} * g$ mit $\underline{0} \neq g \in K[X]$ und $g(\lambda) \neq 0$.

Sei nun $\underline{0} \neq f \in K[X]$ zerfallend und $\mathrm{Grad}(f) \geqslant 1$. Es gibt also $0 \neq c \in K$ und $z_1, \ldots, z_d \in K$ mit $f = c(X - z_1) * \ldots * (X - z_d)$, wobei $d := \mathrm{Grad}(f) \geqslant 1$. Fassen wir gleiche Terme zusammen und schreiben $Z(f) = \{\lambda_1, \ldots, \lambda_r\}$ mit $r \geqslant 1$ (wobei die λ_i paarweise verschieden sind), so erhalten wir eine Darstellung $f = c(X - \lambda_1)^{l_1} * \ldots * (X - \lambda_r)^{l_r}$ mit $l_i \in \mathbb{N}$ für alle i.

Satz 26.3. *Sei* $f = c(X - \lambda_1)^{l_1} * \ldots * (X - \lambda_r)^{l_r}$ *wie oben. Dann folgt automatisch* $l_i = m_f(\lambda_i)$ *für* $1 \leqslant i \leqslant r$; *d.h., die Exponenten* l_i *sind eindeutig bestimmt.*

Beweis. Für jedes i ist $(X - \lambda_i)^{l_i} \mid f$, also $l_i \leqslant m_f(\lambda_i)$. Annahme, es gibt ein i mit $l_i < m_f(\lambda_i)$. Sei $g \in K[X]$ das Produkt von $c \in K$ und allen Faktoren $(X - \lambda_j)^{l_j}$ mit $j \neq i$; dann ist $f = (X - \lambda_i)^{l_i} * g$ und $g(\lambda_i) \neq 0$. Wegen $(X - \lambda_i)^{m_f(\lambda_i)} \mid f$ ist auch $f = (X - \lambda_i)^{m_f(\lambda_i)} * h$ mit einem $h \in K[X]$ und $h(\lambda_i) \neq 0$. In der Gleichung $(X - \lambda_i)^{l_i} * g = f = (X - \lambda_i)^{m_f(\lambda_i)} * h$ können wir $(X - \lambda_i)^{l_i}$ auf beiden Seiten kürzen. Also ist $g = (X - \lambda_i)^{m_f(\lambda_i) - l_i} * h$. Wegen $l_i < m_f(\lambda_i)$ ist λ_i eine Nullstelle der rechten Seite und damit auch $g(\lambda_i) = 0$, Widerspruch. $\square$

Der obige Satz ist ein spezieller Fall einer viel stärkeren Aussage, die den Hauptsatz der elementaren Arithmetik für natürliche Zahlen (siehe Kapitel I, §3) auf Polynome in $K[X]$ verallgemeinert. Dies wird in der Algebra-Vorlesung weiterbehandelt.

Folgerung 26.4. *Sei* $\underline{0} \neq f \in K[X]$ *zerfallend und* $g \in K[X]$ *mit* $g \mid f$. *Schreiben wir* f *wie in Satz 26.3, also* $f = c(X - \lambda_1)^{l_1} * \ldots * (X - \lambda_r)^{l_r}$, *wobei* $0 \neq c \in K$ *und* $l_i \in \mathbb{N}$ *für alle* i, *so gilt* $g = c'(X - \lambda_1)^{l'_1} * \ldots * (X - \lambda_r)^{l'_r}$ *mit* $0 \neq c' \in K$ *und* $0 \leqslant l'_i \leqslant l_i$ *für alle* i.

Beweis. Da f zerfallend ist, ist auch g zerfallend; siehe Lemma 26.1. Eine Nullstelle von g ist auch eine Nullstelle von f, also gilt $Z(g) \subseteq Z(f)$. Daher können wir schreiben $g =$

$c'(X-\lambda_1)^{l'_1} * \ldots * (X-\lambda_r)^{l'_r}$ mit $0 \neq c' \in K$ und $l'_i \geqslant 0$ für $i = 1, \ldots, r$. Für jedes i ist $(X-\lambda_i)^{l'_i}$ ein Teiler von g, also auch von f und damit $l'_i \leqslant m_f(\lambda_i) = l_i$. □

Wir unternehmen nun erste Schritt hin zu einer allgemeinen "Normalform" von Matrizen. Als Spezialfälle erhalten wir teilweise neue Beweise für die Kriterien zur Diagonalisierbarkeit von Matrizen, die wir bereits am Ende des 1. Semesters diskutiert haben.

Bemerkung 26.5. Seien $A, A_1 \in M_n(K)$ ähnliche Matrizen; es gibt also eine invertierbare Matrix $T \in M_n(K)$ mit $A_1 := T^{-1} \cdot A \cdot T$. Nach Lemma 24.3 gilt $\chi_A = \chi_{A_1}$. Eine analoge Aussage gilt auch für die Minimalpolynome. Dazu: Zunächst gilt $A_1^i = (T^{-1} \cdot A \cdot T)^i = T^{-1} \cdot A^i \cdot T$ für alle $i \in \mathbb{N}_0$ (siehe die Rechnung in Kapitel IV, Beispiel 20.8). Daraus folgt sofort $f(A_1) = T^{-1} \cdot f(A) \cdot T$ für alle $f \in K[X]$, d.h., $f(A)$, $f(A_1)$ sind ähnlich für alle $f \in K[X]$. Insbesondere ist $f(A_1) = 0_{n\times n} \Leftrightarrow f(A) = 0_{n\times n}$. Dies zeigt, dass $\mu_A = \mu_{A_1}$ gilt.

Lemma 26.6 (***Zerlegungslemma***). *Sei* $A \in M_n(K)$ *und* $\underline{0} \neq f \in K[X]$ *nicht-konstant mit* $f(A) = 0_{n\times n}$. *Es gelte* $f = f_1 * f_2$ *wobei* $\underline{0} \neq f_i \in K[X]$ *und* $\mathrm{ggT}(f_1, f_2) = 1$. *Dann gibt es eine invertierbare Matrix* $T \in M_n(K)$, *so dass* $A' := T^{-1} \cdot A \cdot T$ *eine Blockdiagonalgestalt der folgenden Form hat:*

$$A' = \left[\begin{array}{c|c} A_1 & 0_{l_1 \times l_2} \\ \hline 0_{l_2 \times l_1} & A_2 \end{array}\right] \quad \textit{wobei } n = l_1 + l_2, \;\; A_i \in M_{l_i}(K) \;\; \textit{und} \;\; f_i(A_i) = 0_{l_i \times l_i}.$$

Beweis. Wir betrachten die Teilräume $U_i := N\big(f_i(A)\big) = \{v \in K^n \mid f_i(A) \cdot v = 0_n\} \subseteq K^n$ für $i = 1, 2$. Sei B_i eine Basis von U_i. Wir werden nun zeigen, dass $B := B_1 \cup B_2$ eine Basis von K^n ist (wobei die Vereinigung disjunkt ist), und dass die Behauptung gilt mit der invertierbaren Matrix $T \in M_n(K)$, deren Spalten durch die Vektoren in B gegeben sind.
Wir teilen dies auf in 5 Schritte wie folgt.

1. Schritt: Sei $i \in \{1, 2\}$ und $g \in K[X]$; wir behaupten, dass $g(A) \cdot v \in U_i$ für alle $v \in U_i$ gilt. Dazu: Sei $v \in U_i$, also $f_i(A) \cdot v = 0_v$. Wir müssen zeigen, dass auch $f_i(A) \cdot (g(A) \cdot v) = 0_n$ gilt. Mit $f_i(A) \cdot g(A) = (f_i * g)(A) = (g * f_i)(A) = g(A) \cdot f_i(A)$ erhalten wir in der Tat $f_i(A) \cdot (g(A) \cdot v) = (f_i(A) \cdot g(A)) \cdot v = (g(A) \cdot f_i(A)) \cdot v = g(A) \cdot (f_i(A) \cdot v) = g(A) \cdot 0_n = 0_n$.

2. Schritt: Wir behaupten, dass $U_1 \cap U_2 = \{0_n\}$ gilt. Insbesondere ist also $B_1 \cap B_2 = \varnothing$.
Dazu: Nach dem Lemma von Beźout für Polynome (siehe Kapitel III, Satz 15.2) gibt es $g_1, g_2 \in K[X]$ mit $1 = \mathrm{ggT}(f_1, f_2) = g_2 * f_1 + g_1 * f_2$. Wir setzen A ein und erhalten $I_n = g_2(A) \cdot f_1(A) + g_1(A) \cdot f_2(A)$. Ist also $v \in U_1 \cap U_2$, d.h. $f_i(A) \cdot v = 0_n$ für $i = 1, 2$, so folgt

$$v = I_n \cdot v = g_2(A) \cdot (f_1(A) \cdot v) + g_1(A) \cdot (f_2(A) \cdot v) = g_2(A) \cdot 0_n + g_1(A) \cdot 0_n = 0_n.$$

Also ist $U_1 \cap U_2 = \{0_n\}$.

3. Schritt: Wir behaupten, dass $K^n = U_1 + U_2$ gilt.

Dazu: Sei $v \in K^n$ beliebig. Wir setzen $w_1 := f_2(A)\cdot v \in K^n$ und $w_2 := f_1(A)\cdot v \in K^n$. Dann ist $f_1(A)\cdot w_1 = f_1(A)\cdot(f_2(A)\cdot v) = (f_1(A)\cdot f_2(A))\cdot v = (f_1 * f_2)(A)\cdot v = f(A)\cdot v = 0_{n\times n}\cdot v = 0_n$, also $w_1 \in U_1$. Analog sieht man $f_2(A)\cdot w_2 = 0_n$, also $w_2 \in U_2$. Wie im 2. Schritt schreibe wieder $1 = g_2 * f_1 + g_1 * f_2$ mit $g_1, g_2 \in K[X]$. Dann folgt

$$v = I_n \cdot v = g_2(A)\cdot(f_1(A)\cdot v) + g_1(A)\cdot(f_2(A)\cdot v) = g_2(A)\cdot w_2 + g_1(A)\cdot w_1.$$

Nach dem 1. Schritt gilt $u_i := g_i(A)\cdot w_i \in U_i$ für $i = 1, 2$. Also ist $v = u_1 + u_2 \in U_1 + U_2$.

4. Schritt: Wir behaupten, dass $B = B_1 \cup B_2$ eine Basis von K^n ist.

Dazu: Nach Kapitel IV, Satz 19.10, gilt $\dim(U_1+U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$. Nach dem 3. Schritt ist die linke Seite gleich $n = \dim K^n$; nach dem 2. Schritt ist die rechte Seite gleich $\dim U_1 + \dim U_2 = |B_1| + |B_2| = |B_1 \cup B_2| = |B|$. Also folgt $|B| = n$. Nach Kapitel IV, Satz 17.6, genügt es nun noch zeigen, dass $K^n = \langle B\rangle_K$ gilt. Sei dazu $v \in K^n$ beliebig. Nach dem 3. Schritt gibt es $u_i \in U_i$ mit $v = u_1 + u_2$. Schreiben wir jeweils u_i als Linearkombination der Vektoren in B_i, so ist v insgesamt eine Linearkombination der Vektoren in $B = B_1 \cup B_2$.

5. Schritt: Sei $T \in M_n(K)$ die invertierbare Matrix mit Spalten gegeben durch die Vektoren in B. Wie behaupten, dass dann $A' := T^{-1}\cdot A \cdot T$ die gewünschte Gestalt hat.

Dazu betrachte die lineare Abbildung $\varphi_A\colon K^n \to K^n$, $x \mapsto A\cdot x$. Dann ist A die Matrix von φ_A bezüglich der Standardbasis von K^n. Nach Kapitel IV, Satz 20.6, ist A' die Matrix von φ_A bezüglich der Basis B. Um die Gestalt von A' zu bestimmen, müssen wir also φ_A auf ein $v \in B$ anwenden und dann den Koordinatenvektor $M_B(\varphi_A(v)) \in K^n$ betrachten. Sei zuerst $v \in B_1$. Nach dem 1. Schritt ist $A\cdot v$ wieder in U_1 enthalten, also eine Linearkombination der Vektoren in B_1; alle Komponenten von $M_B(\varphi_A(v))$ zu Basisvektoren in B_2 sind gleich 0. Eine analoge Aussage gilt entsprechend auch für $v \in B_2$. Damit hat A' eine Blockdiagonalgestalt:

$$A' = \left[\begin{array}{c|c} A_1 & 0_{l_1\times l_2} \\ \hline 0_{l_2\times l_1} & A_2 \end{array}\right] \quad \text{wobei } \ l_i = \dim U_i, \ \ n = l_1 + l_2 \ \text{ und } \ A_i \in M_{l_i}(K).$$

Wir müssen noch zeigen, dass $f_i(A_i) = 0_{l_i\times l_i}$ gilt. Nach Übungen können wir mit Blockdiagonalmatrizen "kästchenweise" rechnen. Wir erhalten

$$f_1(A') = \left[\begin{array}{c|c} f_1(A_1) & 0_{l_1\times l_2} \\ \hline 0_{l_2\times l_1} & f_1(A_2) \end{array}\right] \qquad \text{und} \qquad f_2(A') = \left[\begin{array}{c|c} f_2(A_1) & 0_{l_1\times l_2} \\ \hline 0_{l_2\times l_1} & f_2(A_2) \end{array}\right].$$

Sei $e_j \in K^n$ ein Standard-Basisvektor; dann ist $f_1(A')\cdot e_j$ die j-te Spalte von $f_1(A')$. Es genügt also zu zeigen, dass $f_1(A')\cdot e_j = 0_n$ gilt für $j = 1, \ldots, l_1$, sowie $f_2(A')\cdot e_j = 0_n$ für $j = l_1+1, \ldots, n$. Nun beachte: Es gilt $B_1 = \{T\cdot e_j \mid j = 1, \ldots, l_1\} \subseteq U_1$ und daher

$(f_1(A)\cdot T)\cdot e_j = f_1(A)\cdot(T\cdot e_j) = 0_n$ für $j = 1, \ldots, l_1$. Nach Bemerkung 26.5 ist andererseits $f_1(A') = T^{-1}\cdot f_1(A)\cdot T$, also folgt $f_1(A')\cdot e_j = T^{-1}\cdot\big((f_1(A)\cdot T)\cdot e_j\big) = T^{-1}\cdot 0_n = 0_n$ für $j = 1, \ldots, l_1$, wie gewünscht. Völlig analog zeigt man auch $f_2(A_2) = 0_{l_2\times l_2}$. □

Ab hier Woche 5

Bemerkung 26.7. Ist im obigen Lemma $f = \mu_A$ (und sind f_1, f_2 normiert), so behaupten wir, dass auch $f_i = \mu_{A_i}$ für $i = 1, 2$ gilt. Dazu: Wegen $f_i(A_i) = 0_{l_i\times l_i}$ gilt sicherlich $\mu_{A_i} \mid f_i$ für $i = 1, 2$. Wir müssen also nur noch zeigen, dass $\mathrm{Grad}(\mu_{A_i}) = \mathrm{Grad}(f_i)$ für $i = 1, 2$ gilt. Sei $g := \mu_{A_1} * \mu_{A_2} \in K[X]$. Mit "Kästchen-Rechnen" erhalten wir $g(A') = \left[\begin{array}{c|c} g(A_1) & 0_{l_1\times 2} \\ \hline 0_{l_2\times l_1} & g(A_2)\end{array}\right]$. Wegen $\mu_{A_i} \mid g$ folgt $g(A_i) = 0_{l_i\times l_i}$ für $i = 1, 2$; also $g(A') = 0_{n\times n}$ und damit $f = \mu_A = \mu_{A'} \mid g$; siehe Bemerkung 26.5 für die erste Gleichheit. Es folgt $\mathrm{Grad}(f) \leqslant \mathrm{Grad}(g) = \mathrm{Grad}(\mu_{A_1}) + \mathrm{Grad}(\mu_{A_2}) \leqslant \mathrm{Grad}(f_1) + \mathrm{Grad}(f_2) = \mathrm{Grad}(f)$, und damit $\mathrm{Grad}(\mu_{A_i}) = \mathrm{Grad}(f_i)$ für $i = 1, 2$.

Beispiel 26.8. Sei $K = \mathbb{F}_2 = \{0, 1\}$ (mit $1 + 1 = 0$) und $A := \begin{bmatrix} 1&1&0&1&1\\ 1&1&1&1&1\\ 0&0&0&1&1\\ 0&1&0&0&1\\ 0&1&0&1&0\end{bmatrix} \in M_5(\mathbb{F}_2)$. Man rechnet nach, dass $\mu_A = X^3 * (X + 1)$ gilt.
Also können wir Lemma 26.6 anwenden, indem wir $\mu_A = f_1 * f_2$ schreiben, mit $f_1 = X^3$ und $f_2 = X + 1$. Wie im obigen Beweis setze $U_i := N(f_i(A))$ für $i = 1, 2$. Wir finden:

$$U_1 = N(A^3) = \left\langle \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\\1\end{bmatrix} \right\rangle_{\mathbb{F}_2}, \qquad U_2 = N(A + I_5) = \left\langle \begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\\1\end{bmatrix} \right\rangle_{\mathbb{F}_2}.$$

Ist $T \in M_5(\mathbb{F}_2)$ die invertierbare Matrix, deren Spalten die obigen 5 Spaltenvektoren sind, so erhalten wir eine Blockdiagonalgestalt:

$$T^{-1}\cdot A\cdot T = \left[\begin{array}{ccc|cc} 1&1&1&0&0\\ 0&0&0&0&0\\ 1&0&1&0&0\\ \hline 0&0&0&1&0\\ 0&0&0&0&1\end{array}\right] = \left[\begin{array}{c|c} A_1 & 0\\ \hline 0 & A_2\end{array}\right] \quad \text{mit} \quad A_1 = \begin{bmatrix}1&1&1\\0&0&0\\1&0&1\end{bmatrix}, \quad A_2 = \begin{bmatrix}1&0\\0&1\end{bmatrix},$$

wobei $\mu_{A_1} = X^3$ und $\mu_{A_2} = X + 1$ (wie man auch direkt nachprüfen kann).

Eine Matrix $A \in M_n(K)$ heißt ***zerfallend*** wenn $\chi_A \in K[X]$ ein zerfallendes Polynom ist. Zur Erinnerung: Ist $A \in M_n(K)$, so ist $Z(\mu_A)$ genau die Menge der Eigenwerte von A; siehe Kapitel III, §14; nach Lemma 24.4 gilt auch $Z(\mu_A) = Z(\chi_A)$.
Ist $\lambda \in K$ ein Eigenwert von A und $n_\lambda \geqslant 1$ die Vielfachheit von λ als Nullstelle von χ_A, so heißt n_λ die ***algebraische Vielfachheit*** von λ.

Satz 26.9 (Hauptraumzerlegung). *Sei $A \in M_n(K)$ zerfallend, $Z(\mu_A) = \{\lambda_1, \ldots, \lambda_r\}$ und*

$$\chi_A = (-1)^n (X - \lambda_1)^{n_1} * \ldots * (X - \lambda_r)^{n_r}, \qquad \mu_A = (X - \lambda_1)^{m_1} * \ldots * (X - \lambda_r)^{m_r};$$

wegen $\mu_A \mid \chi_A$ ist hier $n_i \geqslant m_i \geqslant 1$ für alle i. Dann gibt es eine invertierbare Matrix $T \in M_n(K)$ so dass $A' := T^{-1}\cdot A\cdot T$ eine Blockdiagonalgestalt der folgenden Form hat:

$$A' = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A_r \end{bmatrix} \quad \textit{wobei } A_i \in M_{n_i}(K),\ n = n_1 + \ldots + n_r \textit{ und } \mu_{A_i} = (X - \lambda_i)^{m_i}.$$

Beweis. (Vollständige Induktion nach r.) Ist $r = 1$, so gilt die Behauptung mit $T = I_n$ (es ist nichts zu zeigen). Sei nun $r > 1$ und setze $f_1 := (X - \lambda_1)^{m_1}$, $f_2 := (X - \lambda_2)^{m_2} * \ldots * (X - \lambda_r)^{m_r}$. Dann gilt $\mu_A = f_1 * f_2$. Nach Lemma 26.2 ist $\mathrm{ggT}(f_1, f_2) = 1$. Also gibt es eine invertierbare Matrix $T_1 \in M_n(K)$, so dass $\tilde{A} := T_1^{-1} \cdot A \cdot T_1$ eine Blockdiagonalgestalt wie in Lemma 26.6 hat, mit Diagonalblöcken $\tilde{A}_i \in M_{l_i}(K)$, wobei $n = l_1 + l_2$ und $f_i(\tilde{A}_i) = 0_{l_i \times l_i}$ für $i = 1, 2$. (Hier ist $l_i = \dim U_i$, wobei $U_i = N(f_i(A))$.) Nach Bemerkung 26.7 gilt $\mu_{\tilde{A}_i} = f_i$ für $i = 1, 2$. Wir bestimmen nun $\chi_{\tilde{A}_i}$ für $i = 1, 2$. Mit Satz 23.4 und Lemma 24.3 erhalten wir

$$\chi_A = \chi_{\tilde{A}} = \det\left(\left[\begin{array}{c|c} \tilde{A}_1 - XI_{l_1} & 0 \\ \hline 0 & \tilde{A}_2 - XI_{l_2} \end{array}\right]\right) = \det(\tilde{A}_1 - XI_{l_1}) * \det(\tilde{A}_2 - XI_{l_2}) = \chi_{\tilde{A}_1} * \chi_{\tilde{A}_2}.$$

Nun ist $Z(\chi_{\tilde{A}_1}) = Z(\mu_{\tilde{A}_1}) = Z(f_1) = \{\lambda_1\}$ und $Z(\chi_{\tilde{A}_2}) = Z(\mu_{\tilde{A}_2}) = Z(f_2) = \{\lambda_2, \ldots, \lambda_r\}$. Wegen $\chi_A = \chi_{\tilde{A}_1} * \chi_{\tilde{A}_2}$ und mit der Eindeutigkeit in Satz 26.3 folgt $\chi_{\tilde{A}_1} = (-1)^{n_1}(X - \lambda_1)^{n_1}$ und $\chi_{\tilde{A}_2} = (-1)^{n_2}(X - \lambda_2)^{n_2} * \ldots * (X - \lambda_r)^{n_r}$; insbesondere ist $l_1 = \dim U_1 = n_1$.
Nach Induktion gibt es eine invertierbare Matrix $T_2 \in M_{l_2}(K)$ so dass $A_2' := T_2^{-1} \cdot \tilde{A}_2 \cdot T_2$ eine Blockdiagonalgestalt der folgenden Form hat:

$$A_2' = \begin{bmatrix} A_2 & 0 & \dots & 0 \\ 0 & A_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A_r \end{bmatrix} \quad \text{wobei } A_i \in M_{n_i}(K),\ l_2 = n_2 + \ldots + n_r \text{ und } \mu_{A_i} = (X - \lambda_i)^{m_i}.$$

Setze nun $T := T_1 \cdot \tilde{T}_2$, mit der Blockdiagonalmatrix $\tilde{T}_2 := \left[\begin{array}{c|c} I_{l_1} & 0 \\ \hline 0 & T_2 \end{array}\right] \in M_n(K)$.

Mit "Kästchenrechnen" folgt, dass $\tilde{T}_2$ invertierbar ist, wobei $\tilde{T}_2^{-1} = \left[\begin{array}{c|c} I_{l_1} & 0 \\ \hline 0 & T_2^{-1} \end{array}\right]$; außerdem gilt (noch einmal "Kästchenrechnen"):

$$\begin{aligned} T^{-1} \cdot A \cdot T &= \tilde{T}_2^{-1} \cdot \left(T_1^{-1} \cdot A \cdot T_1\right) \cdot \tilde{T}_2 = \tilde{T}_2^{-1} \cdot \tilde{A} \cdot \tilde{T}_2 = \tilde{T}_2^{-1} \cdot \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline 0 & \tilde{A}_2 \end{array}\right] \cdot \tilde{T}_2 \\ &= \left[\begin{array}{c|c} I_{l_1} & 0 \\ \hline 0 & T_2^{-1} \end{array}\right] \cdot \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline 0 & \tilde{A}_2 \end{array}\right] \cdot \left[\begin{array}{c|c} I_{l_1} & 0 \\ \hline 0 & T_2 \end{array}\right] = \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline 0 & T_2^{-1} \cdot \tilde{A}_2 \cdot T_2 \end{array}\right] = \left[\begin{array}{c|c} \tilde{A}_1 & 0 \\ \hline 0 & A_2' \end{array}\right], \end{aligned}$$

und die Matrix auf der rechten Seite hat die gewünschte Blockdiagonalgestalt. □

Bemerkung 26.10. Sei $A \in M_n(K)$ zerfallend und $Z(\mu_A) = \{\lambda_1, \ldots, \lambda_r\}$; für $i = 1, \ldots, r$ seien $n_i \geqslant m_i \geqslant 1$ wie oben. Zur Erinnerung: Nach Kapitel IV, Definition 20.9, heißt $E_A(\lambda_i) := N(A - \lambda_i I_n) = \{x \in K^n \mid A \cdot x = \lambda_i x\}$ der zu λ_i gehörige ***Eigenraum*** und

$\dim E_A(\lambda_i) \geqslant 1$ die ***geometrische Vielfachheit*** von λ_i. Dann gilt offenbar $E_A(\lambda_i) \subseteq U_i := N\big((A-\lambda_i I_n)^{m_i}\big)$. Die Teilräume U_i werden auch als ***Haupträume*** oder ***verallgemeinerte Eigenräume*** bezeichnet. Im obigem Beweis haben wir $\dim U_i = n_i$ gezeigt, also folgt:

(a) Für jedes i gilt $\dim E_A(\lambda_i) \leqslant n_i =$ algebraische Vielfachheit von λ_i.

Nach Kapitel IV, Satz 20.13, ist A diagonalisierbar, wenn $\sum_{i=1}^r \dim E_A(\lambda_i) = n$ gilt. Da stets $n = n_1 + \ldots + n_r$ gilt, folgt also sofort mit (a):

(b) Ist $\dim E_A(\lambda_i) = n_i$ ("algebraische Vielfachheit = geometrische Vielfachheit") für alle i, so ist A diagonalisierbar.

Schließlich haben wir in Kapitel IV, Satz 20.15, gesehen, dass A diagonalisierbar ist, wenn $m_i = 1$ für alle i gilt (d.h., μ_A ist einfach-zerfallend). Dies folgt ebenfalls aus Satz 26.9: Ist nämlich $m_i = 1$, so ist $\mu_{A_i} = X - \lambda_i$, also $A_i = \lambda_i I_{n_i}$ für alle i.

Um das allgemeine Normalformen-Problem zu lösen, müssen wir aufgrund der obigen Ergebnisse noch die folgenden beiden Aufgabenstellungen behandeln:

- Finde eine Normalform für zerfallende Matrizen mit genau einem Eigenwert.
- Finde eine Normalform für allgemeine, nicht-zerfallende Matrizen.

Der gemeinsame Schlüssel zur Lösung dieser beiden Probleme folgt im nächsten Abschnitt.

27. *Das lokale Minimalpolynom eines Vektors*

Sei K ein Körper und $K[X]$ der Polynomring über K in der Unbestimmten X. In Kapitel III, §14, haben wir das Minimalpolynom $\mu_A \in K[X]$ einer Matrix $A \in M_n(K)$ eingeführt. Wir betrachten nun eine Verfeinerung dieses Polynoms.

Satz 27.1. *Sei* $A \in M_n(K)$ *und* $v \in K^n$. *Dann gibt es ein eindeutiges normiertes Polynom* $\underline{0} \neq f_0 \in K[X]$ *kleinsten Grades mit* $f_0(A) \cdot v = 0_n$. *Dieses Polynom wird auch mit* $\mu_{A,v} = f_0$ *bezeichnet und heißt das* ***lokale Minimalpolynom*** *von* v *bezüglich* A. *Es gilt* $\mu_{A,v} \mid g$ *für alle* $g \in K[X]$ *mit* $g(A) \cdot v = 0_n$; *insbesondere ist* $\mu_{A,v} \mid \mu_A$.

Beweis. Dies ist völlig analog zu Satz 14.5 und Folgerung 15.3 in Kapitel III. Ist $v = 0_n$, so gilt die Aussage mit $f_0 := 1$. Sei nun $v \neq 0_n$ und setze $v_i := A^i \cdot v$ für $i \in \mathbb{N}_0$. Wegen $\dim K^n = n$ ist das Tupel $(v_0, v_1, \ldots, v_n)$ linear abhängig. Nach Kapitel IV, Lemma 17.5, gibt es ein $d \in \{1, \ldots, n\}$ mit $v_d \in \langle v_0, \ldots, v_{d-1}\rangle_K$; sei d minimal mit dieser Eigenschaft. Dann ist das Tupel $(v_0, \ldots, v_{d-1})$ linear unabhängig und es gibt $a_0, \ldots, a_{d-1} \in K$ mit $v_d = a_0 v_0 + \ldots + a_{d-1} v_{d-1}$. Setze nun $f_0 := X^d - a_{d-1}X^{d-1} - \ldots - a_1 X - a_0 \in K[X]$.
Dann ist f_0 normiert und

$$f_0(A)\cdot v = \big(A^d - a_{d-1}A^{d-1} - \ldots - a_1 A - a_0 I_n\big)\cdot v = v_d - a_{d-1}v_{d-1} - \ldots - a_1 v_1 - a_0 v_0 = 0_n.$$

Sei nun $g \in K[X]$ beliebig mit $g(A) \cdot v = 0_n$. Teilen mit Rest ergibt $g = q * f_0 + r$ mit

$q, r \in K[X]$, wobei $r = \underline{0}$ oder $r \neq \underline{0}$ und $\mathrm{Grad}(r) < d$ gilt. Nehmen wir an, es sei $r \neq \underline{0}$. Dann folgt $g(A) = q(A) \cdot f_0(A) + r(A)$ und damit $0_n = g(A) \cdot v = q(A) \cdot f_0(A) \cdot v + r(A) \cdot v = r(A) \cdot v$. Sei $d' = \mathrm{Grad}(r) \geqslant 1$ und $r = b_0 + b_1 X + \ldots + b_{d'} X^{d'}$ mit $b_i \in K$. Dann ist $0_n = r(A) \cdot v = b_0 v_0 + b_1 v_1 + \ldots + b_{d'} v_{d'}$. Wegen $r \neq \underline{0}$ ist mindestens eines der b_i ungleich 0 und damit das Tupel $(v_0, v_1, \ldots, v_{d'})$ linear abhängig. Also muss $d' \geqslant d$ gelten, Widerspruch zu $d' = \mathrm{Grad}(r) < d$. Also war die Annahme falsch, d.h., es ist $r = \underline{0}$ und damit $f_0 \mid g$. Ist nun $g \neq \underline{0}$ und $\mathrm{Grad}(g) = d$, so folgt $g = cf_0$ mit einem $0 \neq c \in K$. Ist g außerdem normiert, so muss $c = 1$ gelten; also ist f_0 eindeutig bestimmt. □

Der obige Beweis zeigt, wie man $\mu_{A,v}$ berechnet: Bilde so lange Produkte $A \cdot v, A^2 \cdot v, A^3 \cdot v, \ldots$ bis man ein $d \geqslant 1$ findet, so dass $A^d \cdot v \in K^n$ eine Linearkombination der vorherigen Vektoren $v, A \cdot v, A^2 \cdot v, \ldots, A^{d-1} \cdot v$ ist; aus den Koeffizienten dieser Linearkombination erhält man dann $\mu_{A,v}$. Dieses Verfahren ist völlig analog zu dem für die Berechnung des Minimalpolynoms μ_A, aber einfacher weil man nur Vektoren in K^n auf lineare Abhängigkeit testen muss, und nicht ganze Matrizen in $M_n(K)$. (Dadurch haben die zu betrachtenden linearen Gleichungssysteme viel weniger Unbekannte.)

Beispiel 27.2. (a) Sei $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \in M_3(K)$. Wir bestimmen μ_{A,e_i} für $i = 1, 2, 3$ (wobei $e_i \in K^3$ die Standard-Vektoren sind). Es gilt $A \cdot e_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \notin \langle e_1 \rangle_K$, $A^2 \cdot e_1 = 2e_1$, also ist $\mu_{A,e_1} = X^2 - 2$. Weiterhin ist $A \cdot e_2 = e_2$, also $\mu_{A,e_2} = X - 1$. Analog findet man auch $\mu_{A,e_3} = X^2 - 2$. Betrachte schließlich noch $v := e_1 + e_2$. Es gilt

$$A \cdot v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \notin \langle v \rangle_K, \quad A^2 \cdot v = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \notin \langle v, A \cdot v \rangle_K, \quad A^3 \cdot v = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = A^2 \cdot v + 2A \cdot v - 2v,$$

also ist $\mu_{A,v} = X^3 - X^2 - 2X + 2 = (X - 1) * (X^2 - 2)$. In diesem Fall ist dies gleich μ_A. (Und insbesondere können die Grade der lokalen Minimalpolynome durchaus verschieden sein.)

(b) Sei $f = X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0 \in K[X]$ und $A_f \in M_n(K)$ die zugehörige Begleitmatrix; siehe Beispiel 24.6. Es gilt $\mu_{A_f} = f$. Sei $\{e_1, \ldots, e_n\}$ die Standardbasis von K^n. Es gilt $A^i \cdot e_1 = e_{i+1}$ für $0 \leqslant i \leqslant n - 1$ (wie wir bereits in Kapitel IV, Lemma 15.9, nachgerechnet haben). Also sind $e_1, A_f \cdot e_1, \ldots, A_f^{n-1} \cdot e_1$ linear unabhängig und damit $\mathrm{Grad}(\mu_{A_f,e_1}) \geqslant n$. Andererseits ist $\mu_{A_f,e_1} \mid \mu_{A_f} = f$, also $\mathrm{Grad}(\mu_{A_f,e_1}) \leqslant n$. Damit folgt $\mu_{A_f,e_1} = \mu_{A_f} = f$.

Lemma 27.3. *Sei* $A \in M_n(K)$ *und* $v \in K^n$. *Wir setzen* $U_v := \langle g(A) \cdot v \mid g \in K[X] \rangle_K \subseteq K^n$ *und* $d := \mathrm{Grad}(\mu_{A,v})$. *Dann gilt* $U_v = \langle A^i \cdot v \mid i = 0, 1, \ldots, d - 1 \rangle_K$ *und* $\dim U_v = d$. *Außerdem ist* $\mu_{A,u} \mid \mu_{A,v}$ *für alle* $u \in U_v$.

Beweis. Ist $v = 0_n$, so ist $\mu_{A,v} = 1$ und $U_v = \{0_n\}$; also sind die Aussagen klar. Sei nun $v \neq 0$ und $d := \mathrm{Grad}(\mu_{A,v}) \geqslant 1$. Wie im obigen Beweis setze $v_i := A^i \cdot v$ für $i \in \mathbb{N}_0$. Sei $U := \langle v_0, v_1, \ldots, v_{d-1}\rangle_K \subseteq K^n$. Wegen $v_i \in U_v$ (für alle $i \in \mathbb{N}_0$) ist $U \subseteq U_v$. Weil das Tupel $(v_0, v_1, \ldots, v_{d-1})$ linear unabhängig ist, folgt außerdem $\dim U_v \geqslant d = \dim U$.

Sei nun $u \in U_v$ beliebig, also $u = g(A) \cdot v$ mit $g \in K[X]$. Teilen mit Rest ergibt $g = q * \mu_{A,v} + r$ mit $q, r \in K[X]$, wobei $r = \underline{0}$ oder $r \neq \underline{0}$ und $\mathrm{Grad}(r) < d$ gilt. Dann folgt $g(A) = q(A) \cdot \mu_{A,v}(A) + r(A)$; wegen $\mu_{A,v}(A) \cdot v = 0_n$ also $u = g(A) \cdot v = r(A) \cdot v$. Nun ist $r = b_0 + b_1X + \ldots + b_{d-1}X^{d-1}$ mit $b_i \in K$. Also folgt $u = b_0v + b_1(A\cdot v) + \ldots + b_{d-1}(A^{d-1}\cdot v) = b_0v_0 + b_1v_1 + \ldots + b_{d-1}v_{d-1} \in U$. Damit gilt auch $U_v \subseteq U$, und folglich $U_v = U$..

Schließlich gilt $\mu_{A,v}(A) \cdot u = \mu_{A,v}(A) \cdot (g(A) \cdot v) = (\mu_{A,v}(A) \cdot g(A)) \cdot v = (\mu_{A,v} * g)(A) \cdot v = (g * \mu_{A,v})(A) \cdot v = g(A) \cdot (\mu_{A,v}(A) \cdot v) = g(A) \cdot 0_n = 0_n$, also folgt $\mu_{A,u} \mid \mu_{A,v}$. □

Unser Ziel ist es nun zu zeigen, dass es zu jeder Matrix $A \in M_n(K)$ stets einen Vektor $v \in K^n$ gibt mit $\mu_{A,v} = \mu_A$. Dies ist das Hauptergebnis über lokale Minimalpolynome und wird eines der entscheidenden Hilfsmittel zur Lösung des allgemeinen Normalformen-Problems sein.

Definition 27.4. Sei $A \in M_n(K)$ und $v \in K^n$. Dann heißt v ein ***maximaler Vektor*** (bezüglich A), wenn $\mathrm{Grad}(\mu_{A,v})$ so groß wie möglich ist, also:

$$\mathrm{Grad}(\mu_{A,v}) = \max\{\mathrm{Grad}(\mu_{A,v'}) \mid v' \in K^n\}.$$

Beachte: Wegen $\mu_{A,v'} \mid \mu_A$ für alle $v' \in K^n$ ist die Menge $\{\mathrm{Grad}(\mu_{A,v'}) \mid v' \in K^n\}$ nach oben beschränkt (nämlich durch $\mathrm{Grad}(\mu_A)$); also ist es klar, dass es stets maximale Vektoren gibt. Außerdem: Gilt $\mathrm{Grad}(\mu_{A,v}) = \mathrm{Grad}(\mu_A)$, so ist v sicherlich ein maximaler Vektor.

Wir erinnern an die Definition der Begleitmatrix $A_f \in M_d(K)$ zu einem normierten Polynom $\underline{0} \neq f \in K[X]$ mit $d = \mathrm{Grad}(f) \geqslant 1$; siehe Beispiel 24.6. Es gilt $\mu_{A_f} = f$ und $\chi_{A_f} = (-1)^d f$, was im Folgenden mehrfach verwendet wird.

Lemma 27.5 (Jacob, 1973)**.** *Sei $A \in M_n(K)$ und $v \in K^n$ ein maximaler Vektor bezüglich A. Dann gibt es eine invertierbare Matrix $T_1 \in M_n(K)$, so dass $A_1 := T_1^{-1} \cdot A \cdot T_1$ eine Blockgestalt der folgenden Form hat:*

$$A_1 = \left[\begin{array}{c|c} A_f & 0 \\ \hline 0 & A' \end{array}\right] \quad \textit{wobei} \;\; f := \mu_{A,v} \in K[X], \;\; d := \mathrm{Grad}(f) \geqslant 1 \;\; \textit{und} \;\; A' \in M_{n-d}(K).$$

(Ist $\mathrm{Grad}(f) = n$, so gilt $A_1 = A_f$, d.h., A ist ähnlich zu A_f.)

Beweis. Setze $v_i := A^{i-1} \cdot v$ für $i = 1, \ldots, d$. Dann ist $U_v = \langle v_1, \ldots, v_d\rangle_K$ und $\dim U_v = d$; siehe Lemma 27.3. Wir können $v_{d+1}, \ldots, v_n \in K^n$ finden, so dass $\{v_1, \ldots, v_d, v_{d+1}, \ldots, v_n\}$ eine Basis von K^n ist. Sei $\varepsilon_d \colon K^n \to K$ die lineare Abbildung mit $\varepsilon_d(v_d) = 1$ und $\varepsilon_d(v_i) = 0$ für alle $i \neq d$. Ist $w \in K^n$ beliebig und $w = s_1v_1 + \ldots + s_nv_n$ mit $s_i \in K$, so ist also $s_d = \varepsilon_d(w)$.

Die geniale Idee des Beweises besteht in der folgenden Definition:

$$W := \{w \in K^n \mid \varepsilon_d(A^{j-1} \cdot w) = 0 \text{ für } j = 1, \ldots, d\}.$$

Wir werden nun zeigen, dass W ein Teilraum ist mit $\dim W = n - d$. Ist $\{w_1, \ldots, w_{n-d}\}$ eine Basis von W, so wird $B := \{v_1, \ldots, v_d, w_1, \ldots, w_{n-d}\}$ eine Basis von K^n sein. Schließlich werden wir zeigen, dass die invertierbare Matrix $T_1 \in M_n(K)$ mit Spalten gegeben durch die Vektoren in B die gewünschten Eigenschaften hat. Wir teilen dies auf in 6 Schritte wie folgt.

1. Schritt: Wir behaupten, dass $A \cdot w \in W$ für alle $w \in W$ gilt.
Dazu: Die Aussage ist klar für $w = 0_n$. Sei nun $0_n \neq w \in W$. Wir müssen zeigen, dass $\varepsilon_d(A^j \cdot w) = \varepsilon_d(A^{j-1} \cdot (A \cdot w)) = 0$ für $j = 1, \ldots, d$ gilt. Für $j < d$ ist dies klar, da $w \in W$. Sei nun $j = d$ und $d' := \mathrm{Grad}(\mu_{A,w}) \geqslant 1$. Da v ein maximaler Vektor ist, gilt $d' \leqslant d$. Nun ist $A^d \cdot w \in U_w$ wobei $U_w = \langle w, A \cdot w, \ldots, A^{d'-1} \cdot w\rangle_K$; siehe Lemma 27.3. Wir können also $A^d \cdot w$ als Linearkombination der Vektoren $A^i \cdot w$ für $i = 0, 1, \ldots, d' - 1$ schreiben. Wegen $d' \leqslant d$ und $\varepsilon_d(A^i \cdot w) = 0$ für $i = 0, 1, \ldots, d-1$ ist dann auch $\varepsilon_d(A^d \cdot w) = 0$, also $A \cdot w \in W$.

2. Schritt: Wir behaupten, dass W ein Teilraum von K^n ist mit $\dim W \geqslant n - d$.
Dazu definiere $\varphi\colon K^n \to K^{1\times d}$ durch $\varphi(w) := \big(\varepsilon_d(w), \varepsilon_d(A \cdot w), \ldots, \varepsilon_d(A^{d-1} \cdot w)\big)$. Weil ε_d linear ist, folgt sofort, dass φ linear ist. Nun ist $W = \mathrm{Kern}(\varphi)$, also W ein Teilraum. Mit der Kern-Bild-Dimensionsformel folgt außerdem $n = \dim W + \dim \mathrm{Bild}(\varphi)$; wegen $\mathrm{Bild}(\varphi) \subseteq K^{1\times d}$ ist $\dim \mathrm{Bild}(\varphi) \leqslant d$, und damit $\dim W \geqslant n - d$, wie behauptet.

3. Schritt: Wir behaupten, dass $U_v \cap W = \{0_n\}$ gilt.
Sei dazu $w \in U_v \cap W$, also $w \in W$ und $w = s_1 v_1 + \ldots + s_d v_d$ mit $s_i \in K$. Wegen $w \in W$ ist $\varepsilon_d(A^{j-1} \cdot w) = 0$ für $j = 1, \ldots, d$. Für $j = 1$ bedeutet dies $s_d = \varepsilon_d(w) = 0$, also ist $w = s_1 v_1 + \ldots + s_{d-1} v_{d-1}$. Nun ist $A \cdot v_i = v_{i+1}$ für $1 \leqslant i \leqslant d - 1$. Ist $d > 1$, so erhalten wir $A \cdot w = s_1 v_2 + \ldots + s_{d-1} v_d$ und $s_{d-1} = \varepsilon_d(A \cdot w) = 0$. Danach betrachten wir $A^2 \cdot w, A^3 \cdot w, \ldots$ und erhalten analog $s_{d-2} = \ldots = s_1 = 0$, d.h., es ist $w = 0_n$ wie behauptet.

4. Schritt: Wir behaupten, dass $\dim W = n - d$ und $K^n = U_v + W$ gilt.
Dazu: Nach Kapitel IV, Satz 19.10, gilt $\dim U_v + \dim W = \dim(U_v + W) + \dim(U_v \cap W)$. Wegen $U_v \cap W = \{0_n\}$, $\dim U_v = d$ und $\dim W \geqslant n - d$ ist also $\dim(U_v + W) \geqslant d + (n - d) = n$. Damit folgt $K^n = U_v + W$ und $\dim W = n - d$.

5. Schritt: Sei $\{w_1, \ldots, w_{n-d}\}$ eine Basis von W. Behauptung: $B := \{v_1, \ldots, v_d, w_1, \ldots, w_{n-d}\}$ ist eine Basis von K^n.
Dazu: Wegen $|B| \leqslant n$ müssen wir nur noch zeigen, dass B ein Erzeugendensystem von K^n ist. Sei $v \in K^n$ beliebig. Wegen $K^n = U_v + W$ ist $v = u + w$ mit $u \in U_v$ und $w \in W$. Schreiben wir u als Linearkombination von $v_1, \ldots, v_d$ und w als Linearkombination von $w_1, \ldots, w_{n-d}$, so ist ingesamt $v \in \langle B\rangle_K$.

6. Schritt: Sei $T_1 \in M_n(K)$ die invertierbare Matrix mit Spalten gegeben durch die Vektoren in B. Wir behaupten, dass dann $A_1 := T_1^{-1} \cdot A \cdot T_1$ die gewünschte Gestalt hat.
Dazu betrachte die lineare Abbildung $\varphi_A \colon K^n \to K^n$, $x \mapsto A \cdot x$. Dann ist A die Matrix von φ_A bezüglich der Standardbasis von K^n. Nach Kapitel IV, Satz 20.1, ist A_1 die Matrix von φ_A bezüglich der Basis B. Um die Gestalt von A_1 zu bestimmen, müssen wir also φ_A auf ein $v \in B$ anwenden und dann den Koordinatenvektor $M_B(\varphi_A(v)) \in K^n$ betrachten. Für $1 \leqslant i \leqslant d-1$ ist $\varphi_A(v_i) = A \cdot v_i = A \cdot A^{i-1} \cdot v = A^i \cdot v = v_{i+1}$. Also ist $M_B(\varphi_A(v_i)) = e_{i+1}$ für $i = 1, \ldots, d-1$, und das entspricht genau der gewünschten Form. Für $i = d$ ist $\varphi_A(v_d) = A \cdot v_d = A^d \cdot v = -a_0 v_1 - a_1 v_2 - \ldots - a_{d-1} v_d$, wobei $\mu_{A,v} = a_0 + a_1 X + \ldots + a_{d-1} X^{d-1} + X^d$. Also sind die Komponenten von $M_B(\varphi_A(v_d))$ gegeben durch $-a_0, -a_1, \ldots, -a_{d-1}, 0, \ldots, 0$, ebenfalls wie gewünscht. Betrachte nun die Basisvektoren w_j. Nach dem 1. Schritt ist $A \cdot w_j$ wieder in W enthalten, also eine Linearkombination von $w_1, \ldots, w_{n-d}$. Also sind die ersten d Komponenten in $M_B(\varphi_A(w_j))$ alle gleich 0, wie gewünscht. □

Ab hier Woche 6

Folgerung 27.6. *Sei* $A \in M_n(K)$ *und* $v \in K^n$ *ein maximaler Vektor (bezüglich* A*). Dann gilt* $\mu_{A,v} = \mu_A$.

Beweis. Wegen $\mu_{A,v} \mid \mu_A$ müssen wir nur zeigen, dass $\mu_{A,v}(A) = 0_{n\times n}$ gilt, also $\mu_{A,v}(A) \cdot v' = 0_n$ für alle $v' \in K^n$. Seien $U_v, W \subseteq K^n$ wie im obigen Beweis. Sei $v' \in K^n$ beliebig. Wegen $K^n = U_v + W$ ist $v' = u + w$ mit $u \in U_v$ und $w \in W$. Dann müssen wir zeigen, dass $\mu_{A,v}(A) \cdot u = 0_n$ und $\mu_{A,v}(A) \cdot w = 0_n$ gilt. Für $u \in U_v$ gilt dies nach Lemma 27.3. Für $w \in W$ betrachten wir den Vektor $v + w \in K^n$ und setzen $f := \mu_{A,v+w} \in K[X]$. Dann gilt $0_n = f(A) \cdot (v+w) = f(A) \cdot v + f(A) \cdot w$, also $f(A) \cdot w = -f(A) \cdot v$. Nach Schritt 1 im obigen Beweis ist $A \cdot w' \in W$ für alle $w' \in W$. Daraus folgt $A^i \cdot w' \in W$ für alle $w' \in W$ und alle $i \in \mathbb{N}_0$, also schließlich auch $g(A) \cdot w' \in W$ für alle $w' \in W$ und alle $g \in K[X]$. Insbesondere ist $f(A) \cdot w \in W$. Andererseits ist $f(A) \cdot v \in U_v$. Also ist der Vektor $f(A) \cdot w = -f(A) \cdot v$ sowohl in W als auch in U_v enthalten. Wegen $U_v \cap W = \{0_n\}$ muss damit $f(A) \cdot w = 0_n$ und $f(A) \cdot v = 0_n$ gelten. Es folgt $\mu_{A,v} \mid f$. Da v ein maximaler Vektor ist, gilt andererseits $\mathrm{Grad}(\mu_{A,v}) \geqslant \mathrm{Grad}(\mu_{A,v+w}) = \mathrm{Grad}(f)$. Also folgt $f = \mu_{A,v}$. Schließlich erhalten wir damit $\mu_{A,v}(A) \cdot w = f(A) \cdot w = 0_n$, wie gewünscht. □

Folgerung 27.7. *Sei* $A \in M_n(K)$ *und* $f := \mu_A$. *Gilt* $\mathrm{Grad}(f) = n$, *so ist* A *ähnlich zu* A_f.

Beweis. Sei $v \in K^n$ ein maximaler Vektor (bezüglich A). Nach Folgerung 27.6 gilt $\mu_{A,v} = \mu_A = f$ und damit $\mathrm{Grad}(\mu_{A,v}) = n$. Also folgt die Behauptung aus Lemma 27.5. □

Bemerkung 27.8. Um einen maximalen Vektor bezüglich $A \in M_n(K)$ zu finden, kann man wie folgt vorgehen: Betrachte die Menge S aller normierten Polynome $\underline{0} \neq f \in K[X]$ mit $f \mid \mu_A$

und Grad$(f) <$ Grad(μ_A). Bestimme dann die Teilräume $U_f := N(f(A)) \subseteq K^n$ für $f \in S$. Ist $v \in U_f$ für ein $f \in S$, so gilt $\mu_{A,v} \mid f$ und damit $\mu_{A,v} \neq \mu_A$. Nach Folgerung 27.6 muss es also ein $v_0 \in K^n$ geben mit $v_0 \notin \bigcup_{f \in S} U_f$; für ein solches v_0 gilt dann $f(A) \cdot v_0 \neq 0_n$ für alle $f \in S$. Andererseits ist $\mu_{A,v_0} \mid \mu_A$, also bleibt nur $\mu_{A,v_0} = \mu_A$ übrig.

Beachte allerdings: Es gilt zwar stets, dass $|S|$ endlich ist[2], aber im Allgemeinen kann es sehr schwierig sein, die Menge S explizit zu bestimmen. In konkreten Beispielen beginnt man mit (leicht zu bestimmenden) Teilmengen $S' \subseteq S$ und testet einfach für diverse $v \in K^n \setminus \left(\bigcup_{f \in S'} U_f\right)$, ob bereits $\mu_{A,v} = \mu_A$ gilt. Falls dies noch nicht zum Ziel führt, vergrößert man die Teilmenge S', testet weiter und so fort.

Beispiel 27.9. Sei $A = \begin{bmatrix} 1 & 2 & -2 \\ 0 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix} \in M_3(\mathbb{Q})$. Mit dem Verfahren in Kapitel III, §14, berechnen wir $\mu_A = X^3 - X = X * (X-1) * (X+1)$. Nach Folgerung 26.4 ist hier

$$S = \{1, X, X-1, X+1, X*(X-1), X*(X+1), (X-1)*(X+1)\}.$$

Nun beachte: Sind $f, g \in S$ mit $f \mid g$, so gilt offenbar $U_f \subseteq U_g$. Also genügt es, die Polynome in $S' = \{X*(X-1), X*(X+1), (X-1)*(X+1)\} \subseteq S$ zu betrachten. Wir erhalten:

$$A \cdot (A - I_3) = \begin{bmatrix} -2 & 0 & 4 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{bmatrix}, \qquad \text{also} \qquad U_{X*(X-1)} = \left\{ \begin{bmatrix} 2t \\ s \\ t \end{bmatrix} \,\middle|\, s, t \in \mathbb{Q} \right\}.$$

$$A \cdot (A + I_3) = \begin{bmatrix} 0 & 4 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 0 \end{bmatrix}, \qquad \text{also} \qquad U_{X*(X+1)} = \left\{ \begin{bmatrix} s \\ 0 \\ t \end{bmatrix} \,\middle|\, s, t \in \mathbb{Q} \right\}.$$

$$(A - I_3) * (A + I_3) = \begin{bmatrix} -2 & 2 & 2 \\ 0 & 0 & 0 \\ -1 & 1 & 1 \end{bmatrix}, \qquad \text{also} \qquad U_{(X-1)*(X-1)} = \left\{ \begin{bmatrix} s+t \\ s \\ t \end{bmatrix} \,\middle|\, s, t \in \mathbb{Q} \right\}.$$

Alle Vektoren, die nicht in der Vereinigung der obigen drei Teilräume liegen, sind also maximale Vektoren bezüglich A. Wir sehen damit zum Beispiel, dass $v := \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{Q}^3$ ein maximaler Vektor ist (was man dann auch direkt nachprüfen kann).

Es sei abschließend bemerkt, dass es tatsächlich einen effizienten Algorithmus zur Bestimmung eines maximalen Vektors für $A \in M_n(K)$ gibt, bei dem man nicht die Kenntnis aller Teiler von μ_A benötigt. Siehe:
K. Bongartz, A direct approach to the rational normal form (2014), `https://arxiv.org/abs/1410.1683`;
M. Geck, On Jacob's construction of the rational canonical form of a matrix, Electron. J. Linear Algebra **36** (2020), 177–182; `https://doi.org/10.13001/ela.2020.5055` (siehe dort §4).

[2]Für $|K| < \infty$ ist dies klar. Ist $K \subseteq \mathbb{C}$, so kann man wie folgt argumentieren: Fasse μ_A als Polynom in $\mathbb{C}[X]$ auf und bilde die Menge S' aller normierten $\underline{0} \neq f \in \mathbb{C}[X]$ mit $f \mid \mu_A$ und Grad$(f) <$ Grad(μ_A). Da alle nicht-konstanten Polynome in $\mathbb{C}[X]$ zerfallend sind, folgt sofort aus Folgerung 26.4, dass $|S'|$ endlich ist. Nun ist S die Menge aller $f \in S'$, so dass alle Koeffizienten von f bereits in K sind; also ist auch $|S|$ endlich.

28. *Die Frobenius–Normalform*

In diesem Abschnitt stellen wir die Frobenius–Normalform einer Matrix $A \in M_n(K)$ vor. Im Gegensatz zu anderen Normalformen (zum Beispiel der später zu behandelnden Jordan–Normalform) funktioniert diese über beliebigen Körpern K. Der Beweis beruht auf den Aussagen zu lokalen Minimalpolynomen und maximalen Vektoren aus dem letzten Abschnitt.

Wir beginnen mit einigen Beispielen zu Lemma 27.5 (Jacobs Lemma). Dazu ist es hilfreich, die ersten Schritte des Beweises zu wiederholen und diese etwas anders zu formulieren.

Bemerkung 28.1. Sei $A \in M_n(K)$ und $0_n \neq v \in K^n$ ein maximaler Vektor bezüglich A. Sei $d := \mathrm{Grad}(\mu_{A,v}) \geqslant 1$ und $\{v_1, \dots, v_n\}$ eine Basis von K^n wie im Beweis von Lemma 27.5, wobei $v_i := A^{i-1} \cdot v$ für $i = 1, \dots, d$. Der entscheidende Schritt ist dann, den Teilraum $W := \{w \in K^n \mid \varepsilon_d(A^{j-1} \cdot w) = 0 \text{ für } j = 1, \dots, d\}$ zu bestimmen, wobei $\varepsilon_d \colon K^n \to K$ linear ist mit $\varepsilon_d(v_d) = 1$ und $\varepsilon_d(v_i) = 0$ für $i \neq d$. Überlegen wir uns, wie man dies geschickt macht. Sei $P_0 \in M_n(K)$ die invertierbare Matrix mit Spalten $v_1, \dots, v_n$. Sei $Z_d \in K^{1\times n}$ die d-te Zeile von P_0^{-1}; also $Z_d = e_d^{tr} \cdot P_0^{-1}$. Dann gilt:

$$W = \{w \in K^n \mid C \cdot w = 0_d\} \quad \text{wobei} \quad \begin{cases} C \in K^{d\times n} \text{ die Matrix ist, deren Zeilen durch} \\ Z_d \cdot A^{j-1} \in K^{1\times n} \text{ für } j = 1, \dots, d \text{ gegeben sind.} \end{cases}$$

Denn: Sei $w \in K^n$ und schreibe $w = s_1 v_1 + \ldots + s_n v_n$ mit $s_i \in K$. Nach Definition ist $\varepsilon_d(w) = s_d$. Wegen $P_0 \cdot e_i = v_i$ für alle i folgt $P_0^{-1} \cdot w = s_1 e_1 + \ldots + s_n e_n$, also ist s_d die d-te Komponente von $P_0^{-1} \cdot w$, und damit $\varepsilon_d(w) = s_d = e_d^{tr} \cdot (P_0^{-1} \cdot w) = (e_d^{tr} \cdot P_0^{-1}) \cdot w = Z_d \cdot w$. Es folgt $\varepsilon_d(A^{j-1} \cdot w) = (Z_d \cdot A^{j-1}) \cdot w$ für $j = 1, \dots, d$. Da $(Z_d \cdot A^{j-1}) \cdot w$ die j-te Zeile von $C \cdot w$ ist, gilt also in der Tat $W = \{w \in K^n \mid C \cdot w = 0_d\}$. □

Schließlich beachte: Um die Zeile Z_d von P_0^{-1} zu bestimmen, muss man nicht komplett P_0 invertieren. Da die Spalten von P_0 durch $v_1, \dots, v_n$ gegeben sind, ist Z_d eindeutig bestimmt durch die linearen Gleichungen $Z_d \cdot v_d = 1$ und $Z_d \cdot v_i = 0_n$ für $i \neq d$.

Beispiel 28.2. (a) Sei $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \in M_3(K)$. In Beispiel 27.2(a) haben wir gesehen, dass $f := \mu_A = (X-1) * (X^2-2)$ gilt; sei $d := \mathrm{Grad}(f) = 3$. Es gilt $f = \mu_{A,v}$ für $v = e_1 + e_2$. Setze $v_i := A^{i-1} \cdot v$ für $i = 1, 2, 3$; dann ist $U_v = \langle v_1, v_2, v_3 \rangle_K = K^3$. Ist $T_1 \in M_3(K)$ die invertierbare Matrix mit Spalten v_1, v_2, v_3, so gilt $A_1 = T_1^{-1} \cdot A \cdot T_1 = A_f$. (Hier ist also $W = \{0_3\}$.)

(b) Sei $A = \begin{bmatrix} 5 & -1 & -1 & -13 \\ 2 & 0 & -1 & -7 \\ -4 & 2 & 3 & 14 \\ 2 & -1 & -1 & -6 \end{bmatrix} \in M_4(\mathbb{Q})$.

Wie in Kapitel II, §14, bestimmen wir $f := \mu_A = X^3 - X^2 - X + 1$; sei $d := \mathrm{Grad}(f) = 3$. Sei nun $v_1 := e_2$. Wir berechnen:

$$v_2 := A \cdot e_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \\ -1 \end{bmatrix} \notin \langle v_1 \rangle_{\mathbb{Q}}, \qquad v_3 := A^2 \cdot e_2 = \begin{bmatrix} 6 \\ 3 \\ -4 \\ 2 \end{bmatrix} \notin \langle v_1, v_2 \rangle_{\mathbb{Q}}.$$

Also ist $\mathrm{Grad}(\mu_{A,e_2}) \geqslant 3$ und damit $\mu_{A,e_2} = \mu_A$, d.h., $v_1 = e_2$ ist ein maximaler Vektor. (Wenn man durch Probieren nicht gleich einen solchen Vektor findet, so muss man wie in Beispiel 27.9 vorgehen.) Wählen wir außerdem $v_4 := e_3$, so ist $\{v_1, v_2, v_3, v_4\}$ eine Basis von $\mathbb{Q}^4$. Sei $P_0 \in M_4(\mathbb{Q})$ die invertierbare Matrix mit Spalten v_1, v_2, v_3, v_4. Wir gehen nun wie in Bemerkung 28.1 vor. Sei $Z_3 \in \mathbb{Q}^{1\times 4}$ die dritte Zeile von P_0^{-1}. Die Gleichungen $Z_3 \cdot v_3 = 1$ und $Z_3 \cdot v_i = 0_4$ für $i \neq 3$ haben die eindeutige Lösung $Z_3 = [1/4 \;\; 0 \;\; 0 \;\; -1/4]$. Damit gilt

$$W = \{w \in \mathbb{Q}^4 \mid C \cdot w = 0_3\} \qquad \text{mit} \qquad C = \begin{bmatrix} 1/4 & 0 & 0 & -1/4 \\ 3/4 & 0 & 0 & -7/4 \\ 1/4 & 1 & 1 & 3/4 \end{bmatrix} \in \mathbb{Q}^{3\times 4}.$$

(Die erste Zeile von C ist Z_3, die zweite ist $Z_3 \cdot A$ und die dritte ist $Z_3 \cdot A^2 = (Z_3 \cdot A) \cdot A$.) Mit dem Gauß–Verfahren finden wir $\dim W = 1$ und $W = \langle e_2 - e_3 \rangle_{\mathbb{Q}}$. Ist $T_1 \in M_4(\mathbb{Q})$ die invertierbare Matrix mit Spalten $v_1, v_2, v_3, e_2 - e_3$, so ist

$$A_1 = T_1^{-1} \cdot A \cdot T_1 = \left[\begin{array}{c|c} A_f & 0 \\ \hline 0 & 1 \end{array}\right] \qquad \text{(also } A' = [1] \in M_1(\mathbb{Q}) \text{ wie in Lemma 27.5).}$$

Durch eine wiederholte Anwendung des Lemmas von Jacob gelangen wir zur Frobenius-Normalform einer Matrix. Zunächst noch einige nützliche Bezeichnungen.

Bemerkung 28.3. Gegeben seien Matrizen $A_i \in M_{n_i}(K)$ für $i = 1, \ldots, r$, wobei $n_i \geqslant 1$ für alle i; sei $n := n_1 + \ldots + n_r$. Wir bezeichnen dann mit $R := \mathcal{R}(A_1, \ldots, A_r) \in M_n(K)$ die Blockdiagonalmatrix mit Diagonalblöcken $A_1, \ldots, A_r$, also

$$R = \mathcal{R}(A_1, \ldots, A_r) = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & A_r \end{bmatrix} \in M_n(K).$$

Im Folgenden werden wir mehrfach die folgenden Regeln benutzen.

(a) Mit "Kästchenrechnen" folgt leicht $R^i = \mathcal{R}\big(A_1^i, \ldots, A_r^i\big)$ für alle $i \in \mathbb{N}_0$ und dann auch $f(R) = \mathcal{R}\big(f(A_1), \ldots, f(A_r)\big)$ für alle $f \in K[X]$.

Seien außerdem $A_i' \in M_{n_i}(K)$ für $i = 1, \ldots, r$ und $R' := \mathcal{R}(A_1', \ldots, A_r') \in M_n(K)$. Seien A_i, A_i' ähnlich für $i = 1, \ldots, r$. D.h., es gibt invertierbare Matrizen $T_i \in M_{n_i}(K)$ mit $A_i' = T_i^{-1} \cdot A_i \cdot T_i$ für alle i. Wiederum mit "Kästchenrechnen" folgt:

(b) Die Matrix $\mathsf{T} := \mathcal{R}(\mathsf{T}_1, \dots, \mathsf{T}_r) \in \mathsf{M}_n(\mathsf{K})$ ist invertierbar mit $\mathsf{T}^{-1} = \mathcal{R}(\mathsf{T}_1^{-1}, \dots, \mathsf{T}_r^{-1})$; außerdem gilt $\mathsf{R}' = \mathsf{T}^{-1} \cdot \mathsf{R} \cdot \mathsf{T}$, also sind insgesamt R und R' ähnlich.

Satz 28.4 (Frobenius, 1879). *Sei* $A \in M_n(K)$ *beliebig. Dann gibt es eine invertierbare Matrix* $T \in M_n(K)$ *und nicht-konstante, normierte Polynome* $f_1, \dots, f_r \in K[X]$ *(für ein* $r \geqslant 1$*) mit* $\tilde{A} := T^{-1} \cdot A \cdot T = \mathcal{R}(A_{f_1}, \dots, A_{f_r})$, *wobei* $f_1 = \mu_A$ *und* $f_{i+1} \mid f_i$ *für* $i = 1, \dots, r-1$.

Beweis. (Vollständige Induktion nach n.) Ist $n = 1$ und $A = [a_{11}] \in M_1(K)$, so ist $A = A_f$ mit $f := X - a_{11} \in K[X]$. Sei nun $n > 1$. Sei $f_1 := \mu_A$ und $d := \mathrm{Grad}(\mu_A) \geqslant 1$. Sei $v \in K^n$ ein maximaler Vektor für A. Nach Folgerung 27.6 gilt dann $f_1 = \mu_A = \mu_{A,v}$. Sei $T_1 \in M_n(K)$ eine invertierbare Matrix wie in Lemma 27.5, und setze $A_1 := T_1^{-1} \cdot A \cdot T_1 \in M_n(K)$. Ist $d = n$, so gilt $A_1 = A_{f_1}$ und wir sind fertig. Sei nun $d < n$. Dann hat A_1 eine Blockdiagonalgestalt der folgenden Form:

$$A_1 = \mathcal{R}(A_{f_1}, A') \qquad \text{mit} \qquad A' \in M_{n-d}(K).$$

Nach Bemerkung 26.5 gilt $\mu_{A_1} = \mu_A = f_1$. Zusammen mit Bemerkung 28.3(a) folgt $0_{n\times n} = f_1(A_1) = \mathcal{R}(f_1(A_{f_1}), f_1(A'))$, also ist $f_1(A') = 0_{(n-d)\times(n-d)}$ und damit $\mu_{A'} \mid f_1$. Nach Induktion gibt es eine invertierbare Matrix $T_2 \in M_{n-d}(K)$, so dass $A_2 := T_2^{-1} \cdot A' \cdot T_2 \in M_{n-d}(K)$ eine Blockdiagonalgestalt der folgenden Form hat:

$$A_2 = \mathcal{R}(A_{f_2}, \dots, A_{f_r}) \qquad \begin{array}{l} \text{wobei die } f_i \in K[X] \text{ normiert sind mit } \mathrm{Grad}(f_i) \geqslant 1, \\ \text{sowie } f_2 = \mu_{A'} \text{ und } f_{i+1} \mid f_i \text{ für } i = 2, \dots, r-1 \text{ gilt.} \end{array}$$

Setze nun $\tilde{T}_2 := \mathcal{R}(I_d, T_2) \in M_n(K)$ und $T := T_1 \cdot \tilde{T}_2$. Wegen $A_1 = T_1^{-1} \cdot A \cdot T_1$ und mit Bemerkung 28.3(b) hat dann

$$T^{-1} \cdot A \cdot T = \tilde{T}_2^{-1} \cdot A_1 \cdot \tilde{T}_2 = \mathcal{R}(I_d, T_2^{-1}) \cdot \mathcal{R}(A_{f_1}, A') \cdot \mathcal{R}(I_d, T_2) = \mathcal{R}(A_{f_1}, A_2) = \mathcal{R}(A_{f_1}, \dots, A_{f_r})$$

die gewünschte Form. □

Bemerkung 28.5. In Kapitel IV, §20, hatten wir angekündigt, dass jede Matrix $A \in M_n(K)$ ähnlich ist zu einer Matrix mit höchstens $2n - 1$ Einträgen ungleich 0. Wir sehen nun, dass die Blockdiagonalmatrix $\tilde{A}$ in Satz 28.4 diese Eigenschaft hat. (Dies folgt sofort aus: Ist $f \in K[X]$ ein normiertes Polynom vom Grad $d \geqslant 1$, so hat die Begleitmatrix $A_f \in M_d(K)$ höchstens $2d - 1$ Einträge ungleich 0.)

Folgerung 28.6. *Mit den Bezeichnungen in Satz 28.4 gilt* $\chi_A = (-1)^n f_1 * \dots * f_r$ *und* $\chi_A \mid \mu_A^r$. *Insbesondere ist* χ_A *genau dann zerfallend, wenn* μ_A *zerfallend ist.*

Beweis. Nach Lemma 24.3 gilt $\chi_A = \chi_{\tilde{A}}$. Nun hat auch $\tilde{A} - XI_n$ eine Blockdiagonalgestalt, mit Diagonalblöcken $A_{f_i} - XI_{d_i}$ wobei $d_i = \mathrm{Grad}(f_i) \geqslant 1$ für $i = 1, \dots, r$. Mit Satz 23.4 folgt

$$\chi_{\tilde{A}} = \det(\tilde{A} - XI_n) = \det(A_{f_1} - XI_{d_1}) * \dots * \det(A_{f_r} - XI_{d_r}) = \chi_{A_{f_1}} * \dots * \chi_{A_{f_r}}.$$

Nach Beispiel 24.6 gilt $\chi_{A_{f_i}} = (-1)^{d_i} f_i$ für alle i, also $\chi_A = \chi_{\tilde{A}} = (-1)^n f_1 * \dots * f_r$. Weiterhin ist $f_i \mid f_1 = \mu_A$ für alle i. Also ist μ_A^r ein Vielfaches von $f_1 * \dots * f_r$ und damit $\chi_A \mid \mu_A^r$.

Schließlich: Ist χ_A zerfallend, so ist wegen $\mu_A \mid \chi_A$ auch zerfallend. Ist umgekehrt μ_A zerfallend, so auch μ_A^r und dann wegen $\chi_A \mid \mu_A^r$ auch χ_A. □

Beispiel 28.7. Eine Matrix $A \in M_n(K)$ heißt ***nilpotent***, wenn es ein $m \in \mathbb{N}$ gibt mit $A^m = 0_{n\times n}$. Es gelten die folgenden Äquivalenzen:

$$A \text{ nilpotent} \quad\Leftrightarrow\quad \mu_A = X^d \text{ mit } d \in \mathbb{N},\ d \leqslant n \quad\Leftrightarrow\quad \chi_A = (-1)^n X^n.$$

Denn: Sei zuerst A nilpotent, also $A^m = 0_{n\times n}$ mit $m \in \mathbb{N}$. Dann ist $\mu_A \mid X^m$; also $\mu_A = X^d$ mit einem $d \in \mathbb{N}$ (siehe Folgerung 26.4). Hier ist $d = \mathrm{Grad}(\mu_A) \leqslant n$. Sei nun $\mu_A = X^d$ mit $d \in \mathbb{N}$. Nach Folgerung 28.6 gibt es ein $r \geqslant 1$ mit $\chi_A \mid \mu_A^r = X^{dr}$. Also ist auch χ_A eine Potenz von X und damit $\chi_A = (-1)^n X^n$. Sei schließlich $\chi_A = (-1)^n X^n$. Mit dem Satz von Cayley–Hamilton folgt $A^n = (-1)^n \chi_A(A) = 0_{n\times n}$, also ist A nilpotent. □

Ist A nilpotent, so haben die Polynome $f_i \in K[X]$ in Satz 28.4 die Form $f_i = X^{d_i}$ mit $d_i \in \mathbb{N}$, wobei $n \geqslant d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r \geqslant 1$ und $n = d_1 + \ldots + d_r$. Mehr dazu im nächsten Abschnitt.

Folgerung 28.8. *Sei $A \in M_n(K)$. Dann ist A ähnlich*[3] *zur transponierten Matrix A^{tr}.*

Beweis. Zunächst bemerken wir, dass A und A^{tr} das gleiche Minimalpolynom haben. Denn: Es gilt $(A^{tr})^i = (A^i)^{tr}$ für alle $i \in \mathbb{N}_0$ und damit auch $f(A^{tr}) = f(A)^{tr}$ für alle $f \in K[X]$. Insbesondere ist $f(A) = 0_{n\times n} \Leftrightarrow f(A^{tr}) = 0_{n\times n}$. Also folgt $\mu_{A^{tr}} = \mu_A$.
Sei nun $T \in M_n(K)$ invertierbar und $\tilde{A} := T^{-1} \cdot A \cdot T = \mathcal{R}(A_1, \ldots, A_r)$ wie in Satz 28.4, mit $A_i := A_{f_i}$ zu normierten Polynomen $f_i \in K[X]$. Sei $n_i := \mathrm{Grad}(f_i) \geqslant 1$ für $i = 1, \ldots, r$; also $A_i \in M_{n_i}(K)$. Wie wir gerade gesehen haben, ist $\mu_{A_i^{tr}} = \mu_{A_i}$ für alle i, insbesondere also $\mathrm{Grad}(\mu_{A_i^{tr}}) = n_i$. Nach Folgerung 27.7 ist A_i^{tr} ähnlich zu A_i. Es gibt also invertierbare $C_i \in M_{n_i}(K)$ mit $C_i^{-1} \cdot A_i \cdot C_i = A_i^{tr}$. Dann ist $C := \mathcal{R}(C_1, \ldots, C_r) \in M_n(K)$ invertierbar und es gilt $C^{-1} \cdot \tilde{A} \cdot C = \tilde{A}^{tr}$; siehe Bemerkung 28.3. Es folgt $A^{tr} = P^{-1} \cdot A \cdot P$ mit $P := T \cdot C \cdot T^{tr}$. □

Bemerkung 28.9. Sei $f = X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0 \in K[X]$. Im obigen Beweis wurde benutzt, dass A_f und A_f^{tr} ähnlich sind.
Man kann sogar eine invertierbare Matrix $C_0 \in M_n(K)$ explizit angeben mit $A_f \cdot C_0 = C_0 \cdot A_f^{tr}$, nämlich:

$$C_0 := \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_{n-1} & 1 \\ a_2 & a_3 & \ldots & a_{n-1} & 1 & 0 \\ a_3 & \vdots & \iddots & \iddots & \iddots & \vdots \\ \vdots & a_{n-1} & 1 & 0 & \ldots & 0 \\ a_{n-1} & 1 & \iddots & & & \vdots \\ 1 & 0 & \ldots & & & 0 \end{bmatrix} \in M_n(K).$$

(Beweis selbst oder Übung; siehe auch §2.6, Exercise 5, im Buch von Petersen.)

Beachte $C_0 = C_0^{tr}$. Damit kann man auch die Fußnote unten beweisen, siehe Übungen.

[3] Es gibt sogar eine invertierbare symmetrische Matrix $T \in M_n(K)$ mit $T^{-1} \cdot A \cdot T = A^{tr}$, siehe:
O. Taussky and H. Zassenhaus, On the similarity transformation between a matrix and its transpose, Pacific J. Math. **9** (1959), 893–896.

Satz 28.10. *Gegeben seien* $A, A' \in M_n(K)$. *Sei* A *ähnlich zu* $\tilde{A} = \mathcal{R}(A_1, \ldots, A_r)$ *und* A' *ähnlich zu* $\tilde{A}' = \mathcal{R}(A'_1, \ldots, A'_{r'})$, *wobei* $A_i = A_{f_i}$ *und* $A'_j = A_{f'_j}$ *mit nicht-konstanten, normierten Polynomen* $f_i, f'_j \in K[X]$ *gelte, die jeweils die Bedingungen in Satz 28.4 erfüllen. Dann sind die folgenden Aussagen äquivalent.*

(a) A *und* A' *sind ähnlich.*

(b) $\mathrm{Rang}(f(A)) = \mathrm{Rang}(f(A'))$ *für alle normierten* $\underline{0} \neq f \in K[X]$ *mit* $f \mid \chi_A$ *oder* $f \mid \chi_{A'}$.

(c) $r = r'$ *und* $f_i = f'_i$ *für* $i = 1, \ldots, r$.

(Beachte: Wie früher bemerkt, muss man in (b) nur endlich viele Polynome f testen.)

Beweis. Zunächst einige Vorbetrachtungen. Seien $R, R' \in M_n(K)$ ähnliche Matrizen. Dann gilt $\mathrm{Rang}(R) = \mathrm{Rang}(R')$; siehe Kapitel IV, Bemerkung 19.14. (Dies wird im Folgenden mehrfach benutzt.) Ist $f \in K[X]$ beliebig, so sind $f(R)$ und $f(R')$ ebenfalls ähnlich (siehe Bemerkung 26.5); also gilt auch $\mathrm{Rang}(f(R)) = \mathrm{Rang}(f(R'))$. Es folgt

$(*_1)$ $\mathrm{Rang}(f(A)) = \mathrm{Rang}(f(\tilde{A}))$ und $\mathrm{Rang}(f(A')) = \mathrm{Rang}(f(\tilde{A}'))$ für alle $f \in K[X]$.

Sei schließlich $R = \mathcal{R}(B_1, \ldots, B_r) \in M_n(K)$ eine Blockdiagonalmatrix wie in Bemerkung 28.3, mit $B_i \in M_{n_i}(K)$ wobei $n = n_1 + \ldots + n_r$. Mit "Kästchenrechnen" folgt dann leicht:

$(*_2)$ $\mathrm{Rang}(R) = \mathrm{Rang}(B_1) + \ldots + \mathrm{Rang}(B_r)$.

Nun zum eigentlichen Beweis.

"(a) $\Rightarrow$ (b)" Seien A, A' ähnlich. Wie oben bemerkt sind $f(A), f(A')$ ähnlich für $f \in K[X]$, und ähnliche Matrizen haben den gleichen Rang. Also gilt (b) (sogar für alle $f \in K[X]$).

"(b) $\Rightarrow$ (c)" Zuerst zeigen wir $f_1 = f'_1$. Dazu: Wegen $f_1 = \mu_A$ gilt $f_1(A) = 0_{n\times n}$. Nun ist $f_1 \mid \chi_A$, also gilt $\mathrm{Rang}(f_1(A')) = \mathrm{Rang}(f_1(A)) = 0$ nach Voraussetzung (b), d.h., $f_1(A') = 0_{n\times n}$ und damit $f'_1 = \mu_{A'} \mid f_1$. Völlig analog sieht man auch $f_1 \mid f'_1$, also schließlich $f_1 = f'_1$.
Wir setzen $r_0 := \min\{r, r'\}$. Sei nun $2 \leqslant k \leqslant r_0$ und bereits gezeigt, dass $f_i = f'_i$ für $i = 1, \ldots, k-1$ gilt. Betrachte dann $f_k(\tilde{A})$ und $f_k(\tilde{A}')$. Wegen $f_k \mid \chi_A$ gilt Voraussetzung (b) für f_k; mit $(*_1)$ folgt $\mathrm{Rang}(f_k(\tilde{A})) = \mathrm{Rang}(f_k(\tilde{A}'))$. Da $f_i = f'_i$ für $1 \leqslant i \leqslant k-1$ gilt, stimmen die ersten $k-1$ Diagonalblöcke in $\tilde{A}$ und $\tilde{A}'$ überein; also auch in $f_k(\tilde{A})$ und $f_k(\tilde{A}')$ (siehe Bemerkung 28.3(a)). Die weiteren Diagonalblöcke sind in $f_k(\tilde{A})$ alle gleich Null (weil $\mu_{A_i} = f_i \mid f_k$ für $i \geqslant k$). Wir haben also folgende Situation:

$f_k(\tilde{A})$ hat Diagonalblöcke $f_k(A_1), \ldots, f_k(A_{k-1})$ sowie $r - k + 1$ weitere Null-Blöcke;

$f_k(\tilde{A}')$ hat Diagonalblöcke $f_k(A_1), \ldots, f_k(A_{k-1}), f_k(A'_k), f_k(A'_{k+1}), \ldots, f_k(A'_{r'})$.

Sei $m_i := \mathrm{Rang}(f_k(A_i))$ für $i = 1, \ldots, r$, und $m'_i := \mathrm{Rang}(f_k(A'_i))$ für $i = 1, \ldots, r'$. Dann gilt

$\mathrm{Rang}(f_k(\tilde{A})) = m_1 + \ldots + m_{k-1}$ und

$\mathrm{Rang}(f_k(\tilde{A}')) = m_1 + \ldots + m_{k-1} + m'_k + m'_{k+1} + \ldots + m'_{r'}$;

siehe $(*_2)$. Wegen $\mathrm{Rang}(f_k(\tilde{A})) = \mathrm{Rang}(f_k(\tilde{A}'))$ muss also $m'_k = m'_{k+1} = \ldots = 0$ gelten, d.h., die Diagonalblöcke $f_k(A'_k), f_k(A'_{k+1}), \ldots$ in $\tilde{A}'$ sind alle gleich Null. Insbesondere ist

$f'_k = \mu_{A'_k} \mid f_k$. Mit einem völlig analogen Argument zeigt man $f_k \mid f'_k$. Also gilt $f_k = f'_k$. Indem wir dieses Argument nacheinander für $k = 1, 2, \ldots, r$ ausführen, erhalten wir $f_k = f'_k$ für $k = 1, \ldots, r_0$. Da die Summe der Grade der f_k gleich n ist, und ebenso die Summe der Grade der f'_k, folgt dann auch $r_0 = r = r'$.

“(c) $\Rightarrow$ (a)” Nach (c) gilt $\tilde{A} = \tilde{A}'$. Da $A, \tilde{A}$ sowie $A', \tilde{A}'$ jeweils ähnlich sind, folgt also (a). □

Folgerung 28.11. *Sei $A \in M_n(K)$. Dann sind die Polynome $f_1, \ldots, f_r \in K[X]$ in Satz 28.4 eindeutig bestimmt und heißen die* ***Invariantenteiler*** *von A. Folglich heißt die Matrix $\tilde{A}$ in Satz 28.4 die* ***Frobenius–Normalform*** *(oder auch* ***rationale Normalform****) von A.*

Beweis. Dies folgt sofort aus Satz 28.10, angewandt mit $A = A'$. □

Es gibt mehrere Varianten für die Beweise zur Frobenius–Normalform; siehe etwa §9.3 im Buch von Serre; oder Kapitel 12, §7, im Buch von Artin; oder Chapter 6 im Buch von Blyth–Robertson; oder den am Ende von §27 zitierten Artikel von Bongartz. Jede dieser Varianten hat ihre Vor- und Nachteile, und in konkreten Beispielen kann eine Variante effizienter sein als die andere. Der Zugang mittels Lemma 27.5 (Jacob) erscheint als sehr elementar und elegant; für den Beweis von Satz 28.10 folgen wir dem Artikel von Bongartz.
Das obige Verfahren eignet sich gut zum Programmieren; siehe `https://github.com/geckmf/NoFoMa` für eine Implementierung in GAP. Probieren Sie es einfach mal aus (auch für große Matrizen)!

Ab hier Woche 7

29. *Die Jordan–Normalform einer zerfallenden Matrix*

Eine Matrix $A \in M_n(K)$ heißt eine ***zerfallende Matrix***, wenn das charakteristische Polynom $\chi_A \in K[X]$ zerfallend ist. In Satz 26.9 haben wir bereits eine erste Annäherung an eine Normalform für derartige Matrizen gezeigt. Die Jordan–Normalform stellt eine erhebliche Verfeinerung dieser Aussage dar. Diese ist nützlich in vielen Anwendungen, etwa zum Lösen linearer Differentialgleichungssysteme mit konstanten Koeffizienten (siehe Vorlesung Analysis oder Kapitel 4, §8, im Buch von Artin).
Die Jordan–Normalform setzt sich aus den wie folgt definierten ***Jordan–Blöcken*** zusammen:

$$J_n(\lambda) := \begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 \ldots & & 0 & \lambda & 1 \\ 0 & & \ldots & 0 & \lambda \end{bmatrix} \in M_n(K) \qquad \text{wobei} \qquad \lambda \in K.$$

(Dies ist also eine obere Dreiecksmatrix mit λ auf der Diagonalen und 1 direkt über der Diagonalen; alle anderen Einträge sind 0.) Da $J_n(\lambda)$ eine obere Dreiecksmatrix ist, folgt

$$\chi_{J_n(\lambda)} = \det(J_n(\lambda) - XI_n) = (\lambda - X)^n \qquad \text{und} \qquad J_n(\lambda) \text{ ist zerfallend.}$$

Beachte: Wegen $\chi_{J_n(\lambda)} = (\lambda - X)^n$ ist λ der einzige Eigenwert von $J_n(\lambda)$. Andererseits hat

$J_n(\lambda) - \lambda I_n$ Stufenform mit $n-1$ Stufen, also ist $\mathrm{Rang}\big(J_n(\lambda) - \lambda I_n\big) = n-1$. Damit ist die geometrische Vielfachheit von λ gleich 1. Es gilt also: $J_n(\lambda)$ diagonalisierbar $\Leftrightarrow$ $n = 1$.

Wir behandeln nun zuerst den Fall einer nilpotenten Matrix — nach Beispiel 28.7 ist eine derartige Matrix zerfallend und die Frobenius–Normalform nimmt eine besonders einfache Form an. Bis auf Transponieren ist dies genau die Jordan–Normalform.

Bemerkung 29.1. Betrachte einen Jordan–Block $J_n(0)$ zum Eigenwert $\lambda = 0$. Wegen $\chi_{J_n(0)} = (-1)^n X^n$ ist $J_n(0)$ nilpotent mit $J_n(0)^n = 0_{n\times n}$. Wir beobachten nun, dass $J_n(0)^{\mathrm{tr}} = A_f$ gilt, wobei $f = X^n$. (Die letzte Spalte von A_f besteht für $f = X^n$ nur aus Nullen.) Sei nun $R_n \in M_n(K)$ die Matrix mit Spalten $e_n, e_{n-1}, \dots, e_1$; also

$$R_1 := [\,1\,], \quad R_2 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad R_3 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad R_4 := \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \dots$$

Dann sieht man sofort, dass $R_n^{-1} \cdot J_n(0) \cdot R_n = J_n(0)^{\mathrm{tr}} = A_f$ gilt. Beachte auch $R_n = R_n^{-1}$.

Satz 29.2 (Jordan–Normalform nilpotenter Matrizen)**.** *Sei $A \in M_n(K)$ nilpotent; sei $d \geqslant 1$ minimal mit $A^d = 0_{n\times n}$; also $\mu_A = X^d$. Dann gibt es eine invertierbare Matrix $P \in M_n(K)$ sowie $d_1, \dots, d_r \in \mathbb{N}$ (für $r \geqslant 1$) mit $n = d_1 + \ldots + d_r$ und $d = d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r \geqslant 1$, so dass $P^{-1} \cdot A \cdot P = \tilde{J}_{d_1,\dots,d_r}(A) := \mathcal{R}\big(J_{d_1}(0), \dots, J_{d_r}(0)\big)$ gilt.*

Beweis. Sei $\tilde{A} = \mathcal{R}(A_{f_1}, \dots, A_{f_r})$ die Frobenius–Normalform von A. Sei $d_i := \mathrm{Grad}(f_i)$ für alle i; wegen $f_{i+1} \mid f_i$ für $i = 1, \dots, r-1$ ist dann $n = d_1 + \ldots + d_r$ und $d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r$. Es gilt $f_1 = \mu_A = X^d$, also $d_1 = d$. Für $i \geqslant 2$ ist $f_i \mid f_1 = X^d$; nach Folgerung 26.4 ist auch f_i eine Potenz von X, also $f_i = X^{d_i}$. Nach Bemerkung 29.1 ist jedes A_{f_i} ähnlich zu $J_{d_i}(0)$. Nach Bemerkung 28.3(b) ist dann $\tilde{A}$ insgesamt ähnlich zu $\tilde{J}_{d_1,\dots,d_r}(A)$. $\square$

Folgerung 29.3. *Sei $A \in M_n(K)$; es gelte $\mu_A = (X-\lambda)^d$ mit $\lambda \in K$, $d \geqslant 1$. Dann gibt es eine invertierbare Matrix $P \in M_n(K)$ sowie $d_1, \dots, d_r \in \mathbb{N}$ (für $r \geqslant 1$) mit $n = d_1 + \ldots + d_r$ und $d = d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r \geqslant 1$, so dass $P^{-1} \cdot A \cdot P = \tilde{J}^{\lambda}_{d_1,\dots,d_r}(A) := \mathcal{R}\big(J_{d_1}(\lambda), \dots, J_{d_r}(\lambda)\big)$.*

Beweis. Wegen $(A - \lambda I_n)^d = 0_{n\times n}$ ist $A_0 := A - \lambda I_n$ nilpotent, mit $\mu_{A_0} = X^d$. Nach Satz 29.2 gibt es eine invertierbare Matrix $P \in M_n(K)$ mit $\tilde{J}(A_0) = P^{-1} \cdot A \cdot P = \mathcal{R}(J_{d_1}(0), \dots, J_{d_r}(0))$, wobei $n \geqslant d = d_1 \geqslant d_2 \geqslant \ldots \geqslant d_r \geqslant 1$. Dann ist $P^{-1} \cdot A \cdot P = P^{-1} \cdot (A_0 + \lambda I_n) \cdot P = P^{-1} \cdot A \cdot P + \lambda I_n = \mathcal{R}\big(J_{d_1}(0), \dots, J_{d_r}(0)\big) + \lambda I_n = \mathcal{R}\big(J_{d_1}(\lambda), \dots, J_{d_r}(\lambda)\big)$. $\square$

Bemerkung 29.4. Ein Algorithmus zur Berechnung der Frobenius–Normalform einer beliebigen Matrix $A \in M_n(K)$ wurde im letzten Abschnitt diskutiert, basierend auf dem Lemma von Jacob. Für A nilpotent vereinfacht sich dieser Algorithmus erheblich.

Dazu beachte: Bei diesem Algorithmus ist es entscheidend, ein $v \in K^n$ zu finden mit $\mu_{A,v} =$

μ_A; wie bereits bemerkt, ist dies im Allgemeinen nicht ganz einfach. Ist A nilpotent, so kann man allerdings sehr leicht einen solchen Vektor finden. Sei nämlich $d \geqslant 1$ minimal mit $A^d = 0_{n\times n}$; dann ist $\mu_A = X^d$ (siehe Beispiel 28.7). Wegen $A^{d-1} \neq 0_{n\times n}$ gibt es ein $j \in \{1,\dots,n\}$ mit $A^{d-1}\cdot e_j \neq 0_n$. Nun ist $\mu_{A,e_j} \mid \mu_A = X^d$, also $\mu_{A,e_j} = X^{d'}$ mit $0 \leqslant d' \leqslant d$ (siehe Folgerung 26.4). Wegen $A^{d-1}\cdot e_j \neq 0_n$ ist also $\mu_{A,e_j} = X^d = \mu_A$.

Beispiel 29.5. Sei $K = \mathbb{F}_2 = \{0,1\}$ (mit $1+1=0$) und $A := \begin{bmatrix} 1&1&0&0&1&1\\ 1&1&1&1&1&1\\ 0&1&0&0&1&0\\ 0&1&0&0&1&0\\ 1&1&1&1&1&1\\ 1&0&1&1&0&1 \end{bmatrix} \in M_6(\mathbb{F}_2)$.

Dann ist $A^2 \neq 0_{6\times 6}$, $A^3 = 0_{6\times 6}$, also A nilpotent mit $\mu_A = X^3$.
Hier ist die 2. Spalte von A^2 ungleich 0_6 (siehe unten), also setzen wir $v := e_2 \in \mathbb{F}_2^6$; dann ist $\mu_{A,v} = \mu_A = X^3$. Wir wenden die Prozedur aus dem Lemma von Jacob an (siehe letzter Abschnitt). Setze $v_i := A^{i-1}\cdot v$ für $i = 1,2,3$:

$$A^2 = \begin{bmatrix} 0&1&1&1&1&0\\ 0&1&1&1&1&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ 0&1&1&1&1&0\\ 0&1&1&1&1&0 \end{bmatrix} \quad\rightsquigarrow\quad v_1 := \begin{bmatrix}0\\1\\0\\0\\0\\0\end{bmatrix},\quad v_2 = \begin{bmatrix}1\\1\\1\\1\\1\\0\end{bmatrix},\quad v_3 = \begin{bmatrix}1\\1\\0\\0\\1\\1\end{bmatrix}.$$

Wir sehen, dass v_1, v_2, v_3 zusammen mit den Standard-Vektoren e_4, e_5, e_6 eine Basis B_0 von $\mathbb{F}_2^6$ bilden. Sei $P_0 \in M_6(\mathbb{F}_2)$ die invertierbare Matrix mit Spalten $v_1, v_2, v_3, e_4, e_5, e_6$. Wie in Bemerkung 28.1 müssen wir dann die dritte Zeile Z_3 von P_0^{-1} berechnen, sowie $Z_3\cdot A$ und $Z_3\cdot A^2$. Diese drei Zeilenvektoren bilden die Matrix $C \in \mathbb{F}_2^{3\times 6}$, die hier wie folgt gegeben ist:

$$C = \begin{bmatrix} 1&0&1&0&0&0\\ 1&0&0&0&0&1\\ 0&1&1&1&1&0 \end{bmatrix} \quad\rightsquigarrow\quad W = \{w \in \mathbb{F}_2^6 \mid C\cdot w = 0_3\}.$$

Mit dem Gauß–Verfahren finden wir, dass die folgenden Vektoren eine Basis von W bilden:

$$w_1 = \begin{bmatrix}0\\1\\0\\1\\0\\0\end{bmatrix},\quad w_2 = \begin{bmatrix}0\\1\\0\\0\\1\\0\end{bmatrix},\quad w_3 = \begin{bmatrix}1\\1\\1\\0\\0\\1\end{bmatrix}.$$

Sei $P_1 \in M_6(\mathbb{F}_2)$ die invertierbare Matrix mit Spalten $v_3, v_2, v_1, w_1, w_2, w_3$. Dann ist

$$P_1^{-1}\cdot A_0\cdot P_1 = \left[\begin{array}{ccc|ccc} 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&0&0&0\\ \hline 0&0&0&1&0&1\\ 0&0&0&0&0&0\\ 0&0&0&1&0&1 \end{array}\right] = \left[\begin{array}{c|c} J_3(0) & 0\\ \hline 0 & A'' \end{array}\right] \qquad \text{mit} \qquad A'' = \begin{bmatrix}1&0&1\\0&0&0\\1&0&1\end{bmatrix}.$$

(Hätten wir für P_1 die Reihenfolge $v_1, v_2, v_3, w_1, w_2, w_3$ genommen, so wäre der erste Diagonalblock nicht $J_3(0)$ sondern gleich $A_{X^3} = J_3(0)^{\mathrm{tr}}$ gewesen; siehe Bemerkung 29.1.) Nun müssen wir analog die Matrix A'' behandeln; wir geben nur das Ergebnis an:

$$P_2^{-1} \cdot A'' \cdot P_2 = \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} J_2(0) & 0 \\ \hline 0 & J_1(0) \end{array}\right] \qquad \text{mit} \qquad P_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Damit erhalten wir schließlich die Jordan–Normalform von A:

$$\tilde{J}(A) = P^{-1} \cdot A \cdot P = \mathcal{R}\big(J_3(0), J_2(0), J_1(0)\big) \qquad \text{mit} \qquad P := P_1 \cdot \mathcal{R}(I_3, P_2).$$

Indem wir Satz 26.9 (Hauptraumzerlegung) und Folgerung 29.3 kombinieren, erhalten wir:

Satz 29.6 (Jordan–Normalform, 1870)**.** *Sei $A \in M_n(K)$ zerfallend. Dann gibt es eine invertierbare Matrix $T \in M_n(K)$, so dass $\hat{A} := T^{-1} \cdot A \cdot T$ eine Blockdiagonalgestalt hat mit Jordan–Blöcken $J_d(\lambda)$ entlang der Diagonalen (für diverse $\lambda \in Z(\mu_A)$ und $d \in \mathbb{N}$). Insbesondere ist $\hat{A}$ eine obere Dreiecksmatrix, also A “**trigonalisierbar**”. Außerdem gilt:*

- *Ist $\lambda \in Z(\mu_A)$ und $m_\lambda \geqslant 1$ die Vielfachheit von λ als Nullstelle von μ_A, so gilt $d \leqslant m_\lambda$ für jeden Jordan-Block $J_d(\lambda)$ in $\hat{A}$, und es gibt mindestens einen Block $J_{m_\lambda}(\lambda)$ in $\hat{A}$;*
- *Die Gesamtzahl der Jordan-Blöcke zum Eigenwert λ in $\hat{A}$ ist gleich der Dimension des Eigenraums $E_A(\lambda)$.*

Beweis. Sei $Z(\chi_A) = Z(\mu_A) = \{\lambda_1, \ldots, \lambda_r\}$. Für jedes i sei wie zuvor $n_i \geqslant 1$ die Vielfachheit von λ_i als Nullstelle von χ_A und $m_i \geqslant 1$ die Vielfachheit von λ_i als Nullstelle von μ_A. Nach Satz 26.9 gibt es eine invertierbare Matrix $T_1 \in M_n(K)$ mit $A' := T_1^{-1} \cdot A \cdot T_1 = \mathcal{R}(A_1, \ldots, A_r)$, wobei $A_i \in M_{n_i}(K)$ und $\mu_{A_i} = (X - \lambda_i)^{m_i}$ für alle i. Nach Folgerung 29.3 gibt es zu jedem i eine invertierbare Matrix $P_i \in M_{n_i}(K)$ mit

$$P_i^{-1} \cdot A_i \cdot P_i = \tilde{J}^{\lambda_i}_{d_{i1},\ldots,d_{ir_i}}(A_i) \qquad \text{wobei} \qquad m_i = d_{i1} \geqslant d_{i2} \geqslant \ldots \geqslant d_{ir_i}.$$

Sei $T_2 := \mathcal{R}(P_1, \ldots, P_r) \in M_n(K)$. Nach Bemerkung 28.3(b) ist T_2 invertierbar mit $T_2^{-1} = \mathcal{R}(P_1^{-1}, \ldots, P_r^{-1})$. Sei $T := T_1 \cdot T_2$. Mit “Kästchenrechnen” folgt

$$T^{-1} \cdot A \cdot T = T_2^{-1} \cdot (T_1^{-1} \cdot A \cdot T_1) \cdot T_2) = T_2^{-1} \cdot A' \cdot T_2 = \mathcal{R}\big(P_1^{-1} \cdot A_1 \cdot P_1, \ldots, P_r^{-1} \cdot A_r \cdot P_r\big),$$

und die Matrix auf der rechten Seite ist eine Blockdiagonalmatrix der gewünschten Form. Zur Anzahl der Jordanblöcke zum Eigenwert λ: Es gilt $\dim E_A(\lambda) = n - \mathrm{Rang}(A - \lambda I_n) = n - \mathrm{Rang}(\hat{A} - \lambda I_n)$, und $\mathrm{Rang}(\hat{A} - \lambda I_n)$ kann blockweise berechnet werden (siehe $(*_2)$ im Beweis von Satz 28.10). Für $d \in \mathbb{N}$ und $\mu \in K$ ist schließlich $\mathrm{Rang}(J_d(\mu) - \lambda I_d)$ gleich d (falls $\mu \neq \lambda$) bzw. gleich $d - 1$ (falls $\lambda = \mu$). Also liefert ein Block $J_d(\mu)$ für $\lambda \neq \mu$ den Beitrag d zum Rang von $\hat{A}$, und ein Block $J_d(\lambda)$ den Beitrag $d - 1$ zum Rang von $\hat{A}$. Folglich ist $\mathrm{Rang}(\hat{A} - \lambda I_n)$ gleich $n-$ Anzahl der Jordan-Blöcke zum Eigenwert λ. □

Beispiel 29.7. Betrachte noch einmal die Matrix $A \in M_5(\mathbb{F}_2)$ in Beispiel 26.8, mit $\mu_A = X^3 * (X + 1)$. Wir haben dort gesehen, dass A ähnlich ist zu $\mathcal{R}(A_1, A_2)$, wobei

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (\text{mit } \mu_{A_1} = X^3), \qquad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\text{mit } \mu_{A_2} = X+1).$$

Die Normalform von A_1 muss einen Block $J_3(0)$ enthalten, also ist A_1 ähnlich zu $J_3(0)$. Folglich ist die Jordan–Normalform von A gegegen durch $\mathcal{R}(J_3(0), J_1(1), J_1(1))$.

Beispiel 29.8. Gegeben sei eine Matrix $A \in M_{10}(\mathbb{C})$ mit $\chi_A = (X-3)^7 * (X+1)^3$ und $\mu_A = (X-3)^5 * (X+1)^2$. Welches sind die Möglichkeiten für die Jordan–Normalform von A ? Nun, es gibt zwei Eigenwerte $\lambda_1 = 3$ und $\lambda_2 = -1$ (mit $n_1 = 7$, $m_1 = 5$ und $n_2 = 3$, $m_2 = 2$). Nach Satz 26.9 ist also A ähnlich zu $A' = \mathcal{R}(A_1, A_2)$ wobei $A_1 \in M_7(\mathbb{C})$, $A_2 \in M_3(\mathbb{C})$, mit $\mu_{A_1} = (X-3)^5$ und $\mu_{A_2} = (X+1)^2$. In der Normalform von A_1 muss es einen Block $J_5(3)$ geben. Wegen $A_1 \in M_7(\mathbb{C})$ gibt es also 2 Möglichkeiten: Entweder ein weiterer Block $J_2(3)$, oder zwei weitere Blöcke $J_1(3), J_1(3)$. Analog: In der Normalform von A_2 muss es einen Block $J_2(-1)$ geben. Wegen $A_2 \in M_3(\mathbb{C})$ gibt es nur eine Möglichkeit: ein weiterer Block $J_1(-1)$. Insgesamt gibt es also genau 2 Möglichkeiten für die Jordan–Normalform von A:

$$\mathcal{R}(J_5(3), J_2(3), J_2(-1), J_2(-1)) \qquad \text{oder} \qquad \mathcal{R}(J_5(3), J_1(3), J_1(3), J_2(-1), J_2(-1)).$$

Satz 29.9. *Sei* $A \in M_n(K)$ *zerfallend. Dann gibt es eine diagonalisierbare Matrix* $D \in M_n(K)$ *und eine nilpotente Matrix* $N \in M_n(K)$ *mit* $A = D + N$ *und* $D \cdot N = N \cdot D$.

Beweis. Wir benötigen hier nicht die volle Jordan–Normalform, sondern nur die Hauptraumzerlegung. Sei $T \in M_n(K)$ eine invertierbare Matrix wie in Satz 26.9; es gilt also $A' = T^{-1} \cdot A \cdot T = \mathcal{R}(A_1, \ldots, A_r)$, wobei $A_i \in M_{n_i}(K)$ und $\mu_{A_i} = (X-\lambda_i)^{m_i}$ für alle i. Sei $N_i := A_i - \lambda_i I_{n_i} \in M_{n_i}(K)$; wegen $\mu_{A_i} = (X-\lambda_i)^{m_i}$ ist dann $N_i^{m_i} = 0_{n_i \times n_i}$, also N_i nilpotent.

Wir setzen: $\quad D' := \mathcal{R}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r}) \in M_n(K) \quad$ und $\quad N' := \mathcal{R}(N_1, \ldots, N_r) \in M_n(K)$.

Wegen $A_i = \lambda_i I_{n_i} + N_i$ für $i = 1, \ldots, r$ folgt dann $A' = D' + N'$. Nun gilt offenbar $(\lambda_i I_{n_i}) \cdot N_i = N_i \cdot (\lambda_i I_{n_i})$ für alle i. Mit "Kästchenrechnen" folgt daher auch $D' \cdot N' = N' \cdot D'$. Wir setzen nun $D := T \cdot D' \cdot T^{-1}$ und $N := T \cdot N' \cdot T$. Dann ist D ähnlich zu D', also diagonalisierbar. Mit $m := \max\{m_1, \ldots, m_r\}$ gilt $N_i^m = 0_{n_i \times n_i}$ für alle i, also folgt mit "Kästchenrechnen" auch $N'^m = 0_{n \times n}$, und schließlich $N^m = (T \cdot N' \cdot T^{-1})^m = T \cdot N'^m \cdot T^{-1} = 0_{n \times n}$; also ist N nilpotent. Weiterhin gilt $A = T \cdot A' \cdot T^{-1} = T \cdot (D' + N') \cdot T^{-1} = T \cdot D' \cdot T^{-1} + T \cdot N' \cdot T^{-1} = D + N$, und $D \cdot N = (T \cdot D' \cdot T^{-1}) \cdot (T \cdot N' \cdot T^{-1}) = T \cdot (D' \cdot N') \cdot T^{-1} = T \cdot (N' \cdot D') \cdot T^{-1} = (T \cdot N' \cdot T^{-1}) \cdot (T \cdot D' \cdot T^{-1}) = N \cdot D$. Also erfüllen D, N die gewünschten Bedingungen. □

Beispiel 29.10. Sei $A = \begin{bmatrix} 1 & 3 & 4 & 4 \\ -5 & -11 & -16 & -14 \\ 1 & 2 & 2 & 3 \\ 2 & 5 & 9 & 6 \end{bmatrix} \in M_4(\mathbb{R})$, mit $\mu_A = (X+1)^2 * (X^2 - 2)$.

Also ist A zerfallend, mit Eigenwerten $\lambda_1 = -1$, $\lambda_2 = \sqrt{2}$ und $\lambda_3 = -\sqrt{2}$. In der Jordan–Normalform muss es Blöcke $J_2(-1)$ und $J_1(\pm\sqrt{2})$ geben; also ist die Jordan–Normalform gegeben durch

$$A = \left[\begin{array}{cc|c|c} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ \hline 0 & 0 & \sqrt{2} & 0 \\ \hline 0 & 0 & 0 & -\sqrt{2} \end{array}\right] \in M_4(\mathbb{R}).$$

Gehen wir die Prozedur im Beweis von Satz 29.9 durch, so erhalten wir die Zerlegung $A = D + N$ mit

$$D = \begin{bmatrix} -1 & 1 & 2 & 2 \\ -1 & -7 & -12 & -10 \\ 0 & 1 & 1 & 2 \\ 1 & 4 & 8 & 5 \end{bmatrix} \quad \text{und} \quad N = \begin{bmatrix} 2 & 2 & 2 & 2 \\ -4 & -4 & -4 & -4 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

wobei $\mu_D = (X+1) * (X^2 - 2)$ und $N^2 = 0_{4\times 4}$. (Beweis selbst oder Übung.) Es ist hier bemerkenswert, dass wieder $D, N \in M_4(\mathbb{Q})$ gilt, obwohl es Eigenwerte in $\mathbb{R} \setminus \mathbb{Q}$ gibt!

Bemerkung 29.11. In Folgerung 29.9 kann man sogar zeigen, dass D und N eindeutig bestimmt sind. Die Zerlegung $A = D + N$ heißt ***Jordan–Chevalley–Zerlegung*** von A; siehe zum Beispiel Kapitel 8, §5, im Buch von Koecher für weitere Details, oder auch den etwas ausführlicheren Artikel (mit historischen Bemerkungen):

D. Couty, J. Esterle, R. Zarouf,
Décomposition effective de Jordan–Chevalley, Gazette Math. **129** (2011), 29–49.

In diesem Artikel wird außerdem gezeigt, dass man D, N auf eine völlig andere Weise als im Beweis von Satz 29.9 bestimmen kann, sogar ohne die Eigenwerte von A zu kennen!

Schließlich sei noch bemerkt, dass es auch mehrere Varianten für den Beweis von Satz 29.2 zur Normalform von nilpotenten Matrizen gibt; siehe zum Beispiel:

Kapitel 9, §5* im Buch von Koecher; oder §8.D im Buch von Axler;

oder die folgenden Artikel (alphabetisch geordnet):

K. Bongartz, A direct approach to the rational normal form (2014), online verfügbar auf https://arxiv.org/abs/1410.1683. (Theorem 2 am Ende des Artikels.)

A. Storjohann, An $O(n^3)$ algorithm for the Frobenius normal form, ISSAC 1998; online verfügbar auf https://doi.org/10.1145/281508.281570.

M. Wildon, A short proof of the existence of Jordan normal form; online verfügbar auf http://www.ma.rhul.ac.uk/~uvah099/Maths/JNFfinal.pdf;

Es wäre daher ein interessantes Thema für eine Bachelor-Arbeit, die Effizienz der Algorithmen, die sich aus den verschiedenen möglichen Zugängen zur Frobenius– oder Jordan–Normalfall ergeben, systematisch zu testen und zu untersuchen, inklusive konkreter Implementierungen in GAP oder einem anderen Computer-Algebra-System, etwa dem neu entstehenden OSCAR; siehe https://oscar.computeralgebra.de/.

Kapitel VII: Bilinearformen

In Kapitel IV, §18, haben wir das Standard–Skalarprodukt $\langle\ ,\ \rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ betrachtet, mit $\langle x, y\rangle = x_1y_1 + \ldots + x_ny_n$, wobei $x_1, \ldots, x_n \in \mathbb{R}$ die Komponenten des Spaltenvektors $x \in \mathbb{R}^n$ sind und $y_1, \ldots, y_n \in \mathbb{R}$ die Komponenten des Spaltenvektors $y \in \mathbb{R}^n$. Wir konnten damit auch die Norm $\|x\|$ eines Vektors $x \in \mathbb{R}^n$ definieren und haben ein paar interessante Anwendungen gesehen (etwa die Gaußsche Methode der kleinsten Quadrate). Eine analoge Definition von $\langle\ ,\ \rangle\colon K^n \times K^n \to K$ ist natürlich auch für beliebige Körper K möglich. Dies führt auf den allgemeinen Begriff einer "Bilinearform" auf einem K-Vektorraum V.

30. *Orthogonalität*

Sei K ein beliebiger Körper und V ein K-Vektorraum. Eine Abbildung $\beta\colon V \times V \to K$ heißt eine ***Bilinearform***, wenn β linear in beiden Argumenten ist, also:

$$\beta(s.v + t.v', w) = s\beta(v, w) + t\beta(v', w) \quad \text{und} \quad \beta(v, s.w + t.w') = s\beta(v, w) + t\beta(v, w')$$

für alle $v, v', w, w' \in V$ und alle $s, t \in K$. Ist β eine Bilinearform, so heißt β ***symmetrisch***, wenn $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$ gilt.

Beispiel 30.1. (a) Sei $V = K^n$ und $\beta_0\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) := x_1y_1 + \ldots + x_ny_n$ für alle $x_i, y_j \in K$.

Dann rechnet man sofort nach, dass β_0 eine symmetrische Bilinearform ist. Wir bezeichnen dieses β_0 dann auch als ***Standard-Skalarprodukt*** auf $V = K^n$.

(b) Sei $V = K^n$ und $A = [a_{ij}] \in M_n(K)$ eine fest vorgegebene Matrix. Dann definieren wir

$$\beta_A\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) := \sum_{i,j=1}^{n} a_{ij}x_iy_j \qquad \text{für alle } x_i, y_j \in K.$$

(Die Bilinearform in (a) erhält man mit $A = I_n$.) Mit Hilfe der Definition des Matrixproduktes können wir dies auch schreiben als $\beta_A(v, w) = v^{tr} \cdot A \cdot w$, wobei das Ergebnis eine 1×1-Matrix der Form $[a]$ mit $a \in K$ ist, wofür wir nach unseren Konventionen einfach nur $a \in K$ schreiben. Mit den Regeln für das Matrixprodukt folgt sofort, dass β_A bilinear ist. Mit den allgemeinen Regeln für das Transponieren von Matrizen folgt außerdem: Ist A eine symmetrische Matrix, also $A = A^{tr}$, so ist β_A eine symmetrische Bilinearform.

Seien $e_1, \ldots, e_n \in K^n$ die Standardvektoren. Dann ist $A \cdot e_j$ die j-te Spalte von A und damit $\beta_A(e_i, e_j) = e_i^{tr} \cdot A \cdot e_j = a_{ij}$ für $1 \leqslant i, j \leqslant n$. Ist also β_A eine symmetrische Bilinearform, so folgt umgekehrt auch, dass $A = A^{tr}$ gelten muss.

Umgekehrt lassen sich Bilinearformen für $\dim V < \infty$ komplett durch Matrizen beschreiben.

Bemerkung 30.2. Sei $\beta\colon V \times V \to K$ bilinear und $1 \leqslant n := \dim V < \infty$. Sei $B := \{v_1, \ldots, v_n\}$ eine Basis von V und setze $g_{ij} := \beta(v_i, v_j)$ für alle i, j. Dann definiere die Matrix $G = G_B(\beta) := [g_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(K)$; diese heißt die ***Gram–Matrix*** von β bezüglich B. Wir können damit die Bilinearform $\beta_G\colon K^n \times K^n \to K$ wie oben definieren.
Sind $v, w \in V$ beliebig und $x := M_B(v) \in K^n$, $y := M_B(\beta) \in M_B(w) \in M_n(K)$ die zugehörigen Koordinatenvektoren (bezüglich B), so gilt $\beta(v, w) = \beta_G(x, y) = x^{tr} \cdot G \cdot y$.

[Denn: Seien $x_1, \ldots, x_n \in K$ die Komponenten von x und $y_1, \ldots, y_n \in K$ die Komponenten von y; also $v = \sum_{i=1}^n x_i.v_i$ und $w = \sum_{j=1}^n y_j.v_j$. Mit der Bilinearität von β folgt

$$\begin{aligned}\beta(v, w) &= \beta\big(\textstyle\sum_{i=1}^n x_i.v_i, w\big) = \sum_{i=1}^n x_i\beta(v_i, w) = \sum_{i=1}^n x_i\beta\big(v_i, \sum_{j=1}^n y_j.v_j\big)\\ &= \textstyle\sum_{i=1}^n \sum_{j=1}^n x_i y_j \beta(v_i, v_j) = \sum_{i=1}^n x_i\big(\sum_{j=1}^n g_{ij} y_j\big).\end{aligned}$$

Hier ist $\sum_{j=1}^n g_{ij} y_j$ die i-te Komponente von $G \cdot y \in K^n$; also ist das Ergebnis genau $x^{tr} \cdot G \cdot y$.].

Beachte auch: Ist $A \in M_n(K)$ und B die Standardbasis von K^n, so ist $G_B(\beta_A) = A$. (Denn es gilt $\beta_A(e_i, e_j) = e_i^{tr} \cdot A \cdot e_j = a_{ij}$ für $1 \leqslant i, j \leqslant n$; siehe oben.)

Lemma 30.3 (Basistransformation, vgl. Kapitel IV, Satz 20.1)**.** *Sei* $\beta\colon V \times V \to K$ *bilinear und* $1 \leqslant n := \dim V < \infty$. *Sind* $B = \{v_1, \ldots, v_n\}$ *und* $B' = \{v_1', \ldots, v_n'\}$ *Basen von* V, *so gilt*

$$G_{B'}(\beta) = T^{tr} \cdot G_B(\beta) \cdot T,$$

wobei $T = [t_{ij}] \in M_n(K)$ *die Basiswechselmatrix ist, also* $v_j' = \sum_{i=1}^n t_{ij}.v_i$ *für* $j = 1, \ldots, n$.

Beweis. Sei $G := G_B(\beta) = [g_{ij}]$ und $G' := G_{B'}(\beta) = [g_{ij}']$. Dann ist $g_{ij}' = \beta(v_i', v_j') = \sum_{k=1}^n \sum_{l=1}^n r_{ki} t_{lj} \beta(v_k, v_l) = \sum_{k=1}^n \sum_{l=1}^n t_{ki} g_{kl} t_{lj}$. Nun ist $\sum_{l=1}^n g_{kl} t_{lj}$ der (k, j)-Eintrag von $G \cdot T$, also ist das Ergebnis der (i, j)-Eintrag von $T^{tr} \cdot G \cdot T$, wie behauptet. ☐

Definition 30.4. Sei $\beta\colon V \times V \to K$ eine Bilinearform. Dann heißt β ***reflexiv***, wenn für alle $v, w \in V$ gilt: $\beta(v, w) = 0 \Leftrightarrow \beta(w, v) = 0$. Ist β reflexiv, so heißen $v, w \in V$ ***orthogonal*** (in Zeichen $v \perp w$), wenn $\beta(v, w) = 0 = \beta(w, v)$ gilt. Für eine Teilmenge $X \subseteq V$ definiere dann

$$X^\perp := \{v \in V \mid \beta(x, v) = 0 \text{ für alle } x \in X\}.$$

Weil β linear im zweiten Argument ist, folgt sofort, dass $X^\perp \subseteq V$ ein Teilraum ist. Ist $X = V$, so heißt $V^\perp$ das ***Radikal*** von β; gilt $V^\perp = \{0_V\}$, so heißt β ***nicht-ausgeartet***.

Beispiel 30.5. Jede symmetrische Bilinearform ist natürlich reflexiv. Aber es gibt auch nicht-symmetrische Beispiele: Sei $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in M_2(K)$. Dann gilt $\beta_A\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = x_1 y_2 - x_2 y_1$ für alle $x_1, x_2, y_1, y_2 \in K$. Es folgt $\beta_A(v, v) = 0$ und $\beta_A(v, w) = -\beta_A(w, v)$ für alle $v, w \in K^2$. D.h., β_A ist nicht symmetrisch (jedenfalls wenn $1 + 1 \neq 0$ in K), aber reflexiv.

Wir erwähnen ohne Beweis: Ist $\beta\colon V \times V \to K$ eine reflexive Bilinearform, so ist entweder β symmetrisch oder es gilt $\beta(v, w) = -\beta(w, v)$ für alle $v, w \in V$.
(Siehe z.B. Hauptsatz 7.1.15(a) im Buch von Huppert–Willems für einen Beweis.)

Im Folgenden sei stets $1 \leqslant \dim V < \infty$ und $\beta\colon V \times V \to K$ eine reflexive Bilinearform.

Lemma 30.6. *Ist* $U \subseteq V$ *ein Teilraum, so gilt* $\dim U^\perp \geqslant \dim V - \dim U$*. Ist* β *außerdem nicht-ausgeartet, so gilt* $\dim U^\perp = \dim V - \dim U$*.*

Beweis. Sei $n := \dim V$ und $d := \dim U$. Sei $\{v_1, \ldots, v_d\}$ eine Basis von U. Da β linear im ersten Argument ist, folgt sofort: $U^\perp = \{v \in V \mid \beta(v_i, v) = 0 \text{ für } 1 \leqslant i \leqslant d\}$. Betrachte nun die Abbildung $\varphi\colon V \to K^{1\times d}$ mit $\varphi(v) := \big(\beta(v_1, v), \ldots, \beta(v_d, v)\big)$ für $v \in V$. Weil β linear im zweiten Argument ist, folgt sofort, dass φ linear ist. Es gilt dann $U^\perp = \mathrm{Kern}(\varphi)$. Mit der Kern-Bild-Dimensionsformel folgt außerdem $n = \dim U^\perp + \dim \mathrm{Bild}(\varphi)$; wegen $\mathrm{Bild}(\varphi) \subseteq K^{1\times d}$ ist $\dim \mathrm{Bild}(\varphi) \leqslant d$ und damit $\dim U^\perp \geqslant n - d$, wie behauptet. Nehmen wir an, es sei $\dim U^\perp > n - d$; wir müssen dann $V^\perp \neq \{0_V\}$ zeigen. Dazu: Nach dem Basisergänzungssatz können wir $v_{d+1}, \ldots, v_n \in V$ finden, so dass $\{v_1, \ldots, v_d, v_{d+1}, \ldots, v_n\}$ eine Basis von V ist. Betrachte dann den Teilraum $W := \langle v_{d+1}, \ldots, v_n\rangle_K \subseteq V$, mit $\dim W = n - d$. Analog zu oben gilt $\dim W^\perp \geqslant n - (n - d) = d$. Mit Kapitel IV, Satz 19.10, folgt

$$\dim(U^\perp + W^\perp) + \dim(U^\perp \cap W^\perp) = \dim U^\perp + \dim W^\perp > (n - d) + d = n.$$

Wegen $U^\perp + W^\perp \subseteq V$ ist $\dim(U^\perp + W^\perp) \leqslant n$, also folgt $\dim(U^\perp \cap W^\perp) > 0$. Sei $0_V \neq v \in U^\perp \cap W^\perp$. Dann ist $\beta(v, v_i) = 0$ für alle i, also auch $\beta(v, v') = 0$ für alle $v' \in V$ (wiederum weil β linear im zweiten Argument ist); also $v \in V^\perp$ und damit $V^\perp \neq \{0_V\}$. $\square$

Bemerkung 30.7. Sei $n = \dim V$ und $B = \{v_1, \ldots, v_n\}$ eine Basis von V. Sei $G = G_B(\beta) \in M_n(K)$ die zugehörige Gram–Matrix. Dann gilt:

$$\beta \text{ nicht-ausgeartet} \quad \Leftrightarrow \quad \det(G) \neq 0.$$

Denn: Sei $U := V$. Wie im obigen Beweis betrachte die lineare Abbildung $\varphi\colon V \to K^{1\times n}$ mit $\varphi(v) := \big(\beta(v_1, v), \ldots, \beta(v_n, v)\big)$ für $v \in V$. Es gilt dann $\mathrm{Kern}(\varphi) = V^\perp$. Sei $C = \{f_1, \ldots, f_n\}$ die Standardbasis von $K^{1\times n}$. Für $i, j \in \{1, \ldots, n\}$ ist die i-Komponente von $\varphi(v_j)$ (bezüglich C) gleich $\beta(v_i, v_j) = g_{ij}$; also ist $G = M^B_C(\varphi) \in M_n(K)$ die darstellende Matrix von φ bezüglich B, C. Also folgt: $\det(G) \neq 0 \;\Leftrightarrow\; G$ invertierbar $\;\Leftrightarrow\; V^\perp = \mathrm{Kern}(\varphi) = \{0_V\}$. $\square$

Ab hier Woche 8

Der folgende Satz beschreibt einen fundamentalen Zusammenhang zwischen der Bilinearform β und Endomorphismen $\varphi \in \mathrm{End}(V)$.

Satz 30.8. *Sei* β *nicht-ausgeartet. Ist* $\varphi \in \mathrm{End}(V)$*, so gibt es genau ein* $\varphi^* \in \mathrm{End}(V)$ *mit*

$$\beta\big(\varphi(v), w\big) = \beta\big(v, \varphi^*(w)\big) \qquad \textit{für alle } v, w \in V.$$

Diese Abbildung φ^* heißt die zu φ ***adjungierte Abbildung***.

Beweis. Sei $n = \dim V$ und $B = \{v_1, \ldots, v_n\}$ eine Basis von V. Sei $A := M_B(\varphi) \in M_n(K)$ die darstellende Matrix von φ und $G := G_B(\beta) \in M_n(K)$ die Gram–Matrix von β (beide bezüglich B). Da β nicht-ausgeartet ist, ist G invertierbar (siehe Bemerkung 30.7). Wir setzen

$$A^* := G^{-1} \cdot A^{tr} \cdot G \in M_n(K).$$

Nach Kapitel IV, Satz 19.19, ist die Abbildung $\mathrm{End}(V) \to M_n(K)$, $\psi \mapsto M_B(\psi)$, bijektiv. Also gibt es ein $\varphi^* \in \mathrm{End}(V)$ mit $M_B(\varphi^*) = A^*$. Seien nun $v, w \in V$ beliebig; seien $x := M_B(v) \in K^n$ und $y := M_B(w) \in K^n$ die zugehörigen Koordinatenvektoren. Dann ist $M_B(\varphi(v)) = A \cdot x \in K^n$ (siehe Kapitel IV, Lemma 19.17), also folgt mit Bemerkung 30.2:

$$\beta(\varphi(v), w) = \beta_G(A \cdot x, y) = (A \cdot x)^{tr} \cdot G \cdot y = x^{tr} \cdot A^{tr} \cdot G \cdot y.$$

Nun ist $A^{tr} \cdot G = G \cdot A^*$ und $M_B(\varphi^*(w)) = A^* \cdot y$; also folgt

$$\beta(\varphi(v), w) = x^{tr} \cdot G \cdot (A^* \cdot y) = \beta_G(x, A^* \cdot y) = \beta(v, \varphi^*(w)),$$

wie gewünscht. Schließlich zur Eindeutigkeit von φ^*. Sei auch $\psi \in \mathrm{End}(V)$ mit $\beta(\varphi(v), w) = \beta(v, \psi(w))$ für alle $v, w \in V$. Für festes $w \in V$ ist dann

$$\beta(v, \varphi^*(w) - \psi(w)) = \beta(v, \varphi^*(w)) - \beta(v, \psi(w)) = \beta(\varphi(v), w) - \beta(\varphi(v), w) = 0$$

für alle $v \in V$. Also ist $\varphi^*(w) - \psi(w) \in V^\perp = \{0_V\}$ und damit $\psi(w) = \varphi^*(w)$. □

Definition 30.9. Sei β nicht-ausgeartet und $\varphi \in \mathrm{End}(V)$. Dann heißt φ eine ***orthogonale Abbildung*** (bezüglich β), wenn $\beta\big(\varphi(v), \varphi(w)\big) = \beta(v, w)$ für alle $v, w \in V$ gilt.

Bemerkung 30.10. Sei $\varphi \in \mathrm{End}(V)$. Genau dann ist φ eine orthogonale Abbildung, wenn $\varphi^* \circ \varphi = \mathrm{id}_V$ gilt. Insbesondere ist in diesem Fall φ invertierbar, mit $\varphi^{-1} = \varphi^*$.

Denn: Sei zuerst $\varphi^* \circ \varphi = \mathrm{id}_V$. Dann folgt mit der Formel in Satz 30.8: $\beta\big(\varphi(v), \varphi(w)\big) = \beta\big(v, \varphi^*(\varphi(w))\big) = \beta\big(v, (\varphi^* \circ \varphi)(w)\big) = \beta(v, w)$ für alle $v, w \in V$. Sei umgekehrt φ eine orthogonale Abbildung. Sei $w \in V$ fest. Für alle $v \in V$ gilt dann $\beta(v, w) = \beta\big(\varphi(v), \varphi(w)\big) = \beta\big(v, (\varphi^* \circ \varphi)(w)\big)$, also $\beta\big(v, (\varphi^* \circ \varphi)(w) - w\big) = 0$, d.h., $(\varphi^* \circ \varphi)(w) - w \in V^\perp = \{0_V\}$. Da $w \in V$ beleibig war, gilt also $\varphi^* \circ \varphi = \mathrm{id}_V$. □

Satz 30.11 (***Orthogonale Gruppen***)**.** *Sei β nicht-ausgeartet. Sei $n = \dim V$ und $G \in M_n(K)$ die Gram–Matrix von β bezüglich einer Basis B von V. Dann sind*

$$O(V, \beta) := \{\varphi \in \mathrm{End}(V) \mid \varphi \textit{ orthogonale Abbildung}\} \qquad \textit{und}$$
$$O_n(G, K) := \{A \in M_n(K) \mid A^{tr} \cdot G \cdot A = G\}$$

Gruppen (die erste mit “$\circ$” als Verknüpfung; die zweite mit dem üblichen Matrixprodukt).

Beweis. Sei $\varphi \in O(V, \beta)$. Nach Bemerkung 30.10 gilt $\varphi^* \circ \varphi = \mathrm{id}_V$, also ist φ invertierbar mit $\varphi^{-1} = \varphi^*$; es folgt dann auch $\varphi \circ \varphi^* = \mathrm{id}_V$. Für alle $v, w \in V$ gilt nun $\beta\big(\varphi^{-1}(v), \varphi^{-1}(w)\big) = \beta\big(\varphi^*(v), \varphi^*(w)\big) = \beta\big(\varphi(\varphi^*(v)), w\big) = \beta\big((\varphi \circ \varphi^*)(v), w\big) = \beta(v, w)$, also $\varphi^{-1} \in O(V, \beta)$. Sei nun auch $\psi \in O(V, \beta)$. Dann ist $\beta\big((\varphi \circ \psi)(v), (\varphi \circ \psi)(w)\big) = \beta\big(\varphi(\psi(v)), \varphi(\psi(w))\big) = \beta\big(\psi(v), \psi(w)\big) = \beta(v, w)$ für alle $v, w \in V$, d.h., $\varphi \circ \psi \in O(V, \beta)$. Also ist $O(V, \beta)$ eine Gruppe. Der Beweis ist völlig analog für $O_n(G, K)$. □

Beispiel 30.12. Sei $V = K^n$ und β_0 das Standard-Skalarprodukt. Sei $A \in M_n(K)$ beliebig und $\varphi_A \colon K^n \to K^n$, $x \mapsto A \cdot x$, die zugehörige lineare Abbildung. Sei $B = \{e_1, \ldots, e_n\}$

die Standardbasis von K^n; dann ist $M_B(\varphi_A) = A$ und $G = M_B(\beta_0) = I_n$. Definieren wir $A^* \in M_n(K)$ wie im obigen Beweis, so gilt $A^* = G^{-1} \cdot A^{\mathrm{tr}} \cdot G = A^{\mathrm{tr}}$. Also ist $\varphi^*\colon K^n \to K^n$, $x \mapsto A^{\mathrm{tr}} \cdot x$. In diesem Fall ist $O_n(K) := O_n(I_n, K) = \{A \in M_n(K) \mid A^{\mathrm{tr}} \cdot A = I_n\}$.

Beispiel 30.13. Sei $K = \mathbb{R}$ und $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in O_2(\mathbb{R})$ (siehe Beispiel 30.12), also $A^{\mathrm{tr}} \cdot A = I_2$. Ausmultiplizieren ergibt die Bedingungen $a^2 + c^2 = 1$, $b^2 + d^2 = 1$ und $ab + cd = 0$.
1. Fall: Ist $a = 0$, so folgt $c = \pm 1$ und dann auch $d = 0$, $c = \pm 1$.
2. Fall: Sei $a \neq 0$. Aus $ab + cd = 0$ erhalten wir $b = -cd/a$ und dann $1 = b^2 + d^2 = c^2d^2/a^2 + d^2 = (a^2 + c^2)d^2/a^2 = d^2/a^2$, also $d = \pm a$ und dann auch $c = \mp b$. Also ist $A = \begin{bmatrix} a & b \\ \mp b & \pm a \end{bmatrix}$ mit $a^2 + b^2 = 1$. Umgekehrt sieht man sofort, dass eine Matrix dieser Form orthogonal ist. Der 1. Fall passt auch in dieses Schema, mit $a = 0$ und $b = \pm 1$. Schließlich beachte: Sind $a, b \in \mathbb{R}$ gegeben mit $a^2 + b^2 = 1$, so gibt es ein eindeutiges $\theta \in [0, 2\pi)$ mit $a = \cos(\theta)$ und $b = \sin(\theta)$. Also gilt:

$$A \in M_2(\mathbb{R}) \text{ orthogonal} \quad \Leftrightarrow \quad A = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\delta\sin(\theta) & \delta\cos(\theta) \end{bmatrix} \text{ mit } \delta = \pm 1, \;\; \theta \in [0, 2\pi).$$

Definition 30.14. Sei $n = \dim V$ und β symmetrisch. Sei $B = \{v_1, \ldots, v_n\}$ eine Basis von V. Dann heißt B eine ***Orthogonalbasis*** (bezüglich β), wenn $\beta(v_i, v_j) = 0$ für alle $i \neq j$ gilt. Ist zusätzlich noch $\beta(v_i, v_i) = 1$ für alle i, so heißt B eine ***Orthonormalbasis*** von V. In diesem Fall hat jedes $v \in V$ eine eindeutige Darstellung als $v = \beta(v, v_1)v_1 + \ldots + \beta(v, v_n)v_n$.

Satz 30.15 (Existenz von Orthogonalbasen). *Sei β symmetrisch und $1 + 1 \neq 0$ in K. Dann gibt es eine Orthogonalbasis von V bezüglich β.*

Beweis. (Vollständige Induktion nach $n := \dim V \geqslant 1$.) Für $n = 1$ ist nichts zu zeigen. Sei nun $n > 1$. Wir unterscheiden jetzt zwei Fälle:

1. Fall: Es gibt ein $0 \neq v_1 \in V$ mit $\beta(v_1, v_1) \neq 0$. Setze $U := \langle v_1 \rangle_K$. Dann ist $\dim U = 1$, also $\dim U^\perp \geqslant n - 1$ nach Lemma 30.6. Wegen $\beta(v_1, v_1) \neq 0$ ist $v_1 \notin U^\perp$, also $U^\perp \subsetneqq V$ und damit $\dim U^\perp = n - 1$. Sei $\beta'\colon U^\perp \times U^\perp \to K$ die Einschränkung von β. Dann ist natürlich auch β' eine symmetrische Bilinearform. Nach Induktion gibt es eine Orthogonalbasis $\{v_2, \ldots, v_n\}$ von $U^\perp$. Wegen $v_1 \notin U^\perp$ ist das Tupel $(v_1, v_2, \ldots, v_n)$ linear unabhängig, also ist $B := \{v_1, v_2, \ldots, v_n\}$ eine Basis von V. Wegen $v_2, \ldots, v_n \in U^\perp$ gilt $\beta(v_1, v_i) = 0$ für alle $i \geqslant 2$, also ist B insgesamt eine Orthogonalbasis.

2. Fall: Es gilt $\beta(v, v) = 0$ für alle $v \in V$. Wir behaupten, dass dann $\beta(v, w) = 0$ für alle $v, w \in V$ gilt. (In diesem Fall ist jede Basis von V eine Orthogonalbasis; also gilt der Satz auch in diesem Fall.) Seien nun $v, w \in V$ beliebig. Dann gilt

$$0 = \beta(v + w, v + w) = \beta(v, v) + \beta(v, w) + \beta(w, v) + \beta(w, w) = \beta(v, w) + \beta(w, v).$$

Wegen der Symmetrie von β folgt $0 = \beta(v,w) + \beta(v,w) = (1+1)\beta(v,w)$, und wegen $1+1 \neq 0$ in K dann $\beta(v,w) = 0$, wie behauptet. $\square$

Bemerkung 30.16. Der obige Beweis liefert auch ein Verfahren zur Bestimmung einer Orthogonalbasis. Sei $n = \dim V$ und $C := \{w_1, \dots, w_n\}$ eine beliebige Basis von V; bilde dann $A = [a_{ij}] := G_C(\beta) \in M_n(K)$. Ist $A = 0_{n\times n}$, so ist $\beta(v,w) = 0$ für alle $v, w \in V$, also C eine Orthogonalbasis. Sei nun $A \neq 0_{n\times n}$. Wir behaupten, dass wir dann ein $v_1 \in V$ mit $\beta(v_1, v_1) \neq 0$ finden können. Dazu: Seien $i, j \in \{1, \dots, n\}$ mit $\beta(w_i, w_j) = a_{ij} \neq 0$. Ist $i = j$ so ist $v_1 := w_i$ der gesuchte Vektor. Nehmen wir nun an, es sei $i \neq j$ und $\beta(w_k, w_k) = a_{kk} = 0$ für alle k. Wie im 2. Fall des obigen Beweises erhalten wir die Gleichung

$$\beta(w_i + w_j, w_i + w_j) = \beta(w_i, w_i) + (1+1)a_{ij} + \beta(w_j, w_j) = (1+1)a_{ij} \neq 0;$$

also ist hier $v_1 := w_i + w_j$ der gesuchte Vektor. Wir sind damit im 1. Fall des obigen Beweises. Sei $U := \langle v_1 \rangle_K$; dann ist $\dim U^\perp = n-1$. Wir wenden nun das obige Verfahren rekursiv auf die Einschränkung von β auf $U^\perp$ an und erhalten eine Orthogonalbasis $\{v_2, \dots, v_n\}$ von $U^\perp$. Wie im obigen Beweis ist dann $\{v_1, \dots, v_n\}$ eine Orthogonalbasis von V bezüglich β.

Beispiel 30.17. Sei $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \in M_3(K)$ und $\beta = \beta_A \colon K^3 \times K^3 \to K$.

Sei $2 := 1 + 1 \neq 0$ in K, und $C = \{e_1, e_2, e_3\}$ die Standardbasis von K^3. Es gilt $\beta(e_k, e_k) = 0$ für $k = 1, 2, 3$ aber $\beta(e_1, e_2) \neq 0$ (zum Beispiel). Setzen wir $v_1 := e_1 + e_2$, so gilt $\beta(v_1, v_1) = (1+1)a_{12} = 2 \neq 0$. Sei $U := \langle v_1 \rangle_K$. Wir bestimmen dann

$$U^\perp = \left\{ v = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in K^3 \;\middle|\; 0 = \beta(v_1, v) = [1\ 1\ 0] \cdot A \cdot v = a + b + c \right\} = \langle v_2, v_3 \rangle_K,$$

wobei $v_2 = e_1 - e_2$, $v_3 = e_2 - e_3$. Die Gram–Matrix der Einschränkung von β auf $U^\perp$ bezüglich der Basis $\{v_2, v_3\}$ ist dann gegeben durch $A' = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \in M_2(K)$. Hier sehen wir direkt, dass $v_3' := e_1 - e_3 = v_2 + v_3 \perp v_2$ gilt, also ist $\{v_2, v_3'\}$ eine Orthogonalbasis von $U^\perp$ und damit $\{v_1, v_2, v_3'\}$ eine Orthogonalbasis von K^3.

Beispiel 30.18. Auf die Voraussetzung "$1 + 1 \neq 0$" im obigen Satz kann nicht verzichtet werden. Sei nämlich $K = \mathbb{F}_2$ und $\beta_A \colon K^2 \times K^2 \to K$ mit A wie in Beispiel 30.5. Wegen $1 + 1 = 0$ ist dann β_A symmetrisch und es gilt $\beta_A(v, v) = 0$ für alle $v \in K^2$. Nehmen wir an, es gibt eine Orthogonalbasis $\{v_1, v_2\}$ von K^2. Dann gilt $\beta_A(v_i, v_i) = 0$ für $i = 1, 2$, und auch $\beta_A(v_1, v_2) = \beta_A(v_2, v_1) = 0$. Also ist die Gram–Matrix von β_A bezüglich der Basis $\{v_1, v_2\}$ die Null-Matrix. Damit folgt auch $\beta_A(v, w) = 0$ für alle $v, w \in K^2$, Widerspruch.

Beispiel 30.19. Sei $A \in M_n(K)$ eine Diagonalmatrix mit Diagonaleinträgen $d_1, \dots, d_n \in K$. Bilde dann die Bilinearform $\beta_A \colon K^n \times K^n \to K$. Diese ist symmetrisch und die Standardbasis $B = \{e_1, \dots, e_n\}$ ist eine Orthogonalbasis, mit $\beta(e_i, e_i) = d_i$ für alle i.

Sei nun $d_i \neq 0$ für alle i und nehmen wir an, es gibt $z_i \in K$ mit $z_i^2 = d_i$; setze damit $v_i := z_i^{-1}e_i$. Dann ist $B' = \{v_1, \dots, v_n\}$ auch eine Orthogonalbasis und es gilt $\beta(v_i, v_i) = \beta(z_i^{-1}e_i, z_i^{-1}e_i) = z_i^{-2}\beta(e_i, e_i) = z_i^{-2}d_i = 1$. Also ist B' eine Orthonormalbasis.

Wir sehen also, dass die Existenz einer Orthonormalbasis sehr stark vom Körper K abhängt. Ist $K = \mathbb{C}$, so gibt es stets $z_i \in K$ mit $z_i^2 = d_i$ wie oben. Bereits über $\mathbb{R}$ ist dies nicht immer möglich, zum Beispiel wenn $d_i = -1$ für ein i gilt.

Ab hier Woche 9

31. *Symmetrische Bilinearformen über* $\mathbb{R}$

In diesem Abschnitt sei V ein $\mathbb{R}$-Vektorraum und $\beta\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform, wobei $1 \leqslant \dim V < \infty$. Nach Satz 30.15 gibt es stets eine Orthogonalbasis B von V bezüglich β; die zugehörige Gram–Matrix $G_B(\beta)$ ist also eine Diagonalmatrix. Da $\mathbb{R}$ angeordnet ist, können wir nach den Diagonal-Einträgen von $G_B(\beta)$ schauen, die positiv oder negativ sind. Der folgende Satz gibt eine präzise Information dazu.

Satz 31.1 (Trägheitssatz von Sylvester). *Sei* $n := \dim V$ *und* $B = \{v_1, \dots, v_n\}$ *eine Orthogonalbasis von* V *bezüglich* β*; sei* $d_i := \beta(v_i, v_i) \in \mathbb{R}$ *für alle* i*. Sei* p *die Anzahl der* i *mit* $d_i > 0$ *und* q *die Anzahl der* i *mit* $d_i < 0$*. Dann sind* p, q *eindeutig durch* β *bestimmt, also unabhängig von der Wahl von* B*. — Das Paar* (p, q) *heißt die* ***Signatur*** *von* β*.*

Beweis. Wir sortieren B so, dass $d_i > 0$ für $i = 1, \dots, p$ und $d_i < 0$ für $i = p+1, \dots, p+q \leqslant n$ gilt. Sei $B' = \{w_1, \dots, w_n\}$ eine weitere Orthogonalbasis; sei $d_i' := \beta(w_i, w_i)$ für alle i; sei analog p' die Anzahl der i mit $d_i' > 0$ und q' die Anzahl der i mit $d_i' < 0$. Wiederum sei B' so sortiert, dass $d_i' > 0$ für $i = 1, \dots, p'$ und $d_i' < 0$ für $i = p'+1, \dots, p'+q' \leqslant n$ gilt. Wir müssen nun zeigen, dass $p = p'$ und $q = q'$ gilt. Betrachte dazu die Teilräume

$$U_1 := \langle v_1, \dots, v_p \rangle_{\mathbb{R}} \qquad \text{und} \qquad U_2 := \langle w_{p'+1}, \dots, w_n \rangle_{\mathbb{R}}.$$

Wir behaupten: $U_1 \cap U_2 = \{0_V\}$. Annahme, es gibt ein $0_V \neq v \in U_1 \cap U_2$. Schreibe $v = \sum_{i=1}^{p} x_i.v_i$ mit $x_i \in \mathbb{R}$. Dann gilt $\beta(v, v) = \sum_{i=1}^{p} x_i^2 d_i > 0$ (wegen $x_i \neq 0$ für mindestens ein i). Schreibe andererseits $v = \sum_{i=p'+1}^{n} y_i.v_i$ mit $y_i \in \mathbb{R}$. Da $d_i < 0$ für $p' \leqslant i \leqslant p' + q'$ und $d_i = 0$ für $i > p' + q'$ gilt, folgt $\beta(v, v) = \sum_{i=p'+1}^{n} y_i^2 d_i \leqslant 0$, Widerspruch.
Also ist $U_1 \cap U_2 = \{0_V\}$. Mit Kapitel IV, Satz 19.10, erhalten wir daraus

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 = p + n - p'.$$

Wegen $U_1 + U_2 \subseteq V$ ist $\dim(U_1 + U_2) \leqslant n$, also folgt $p + n - p' \leqslant n$ und damit $p \leqslant p'$. Indem man die Rollen von B und B' vertauscht, folgt mit einem völlig analogen Argument auch $p' \leqslant p$. Also gilt $p = p'$. Um auch $q = q'$ zu zeigen, setzen wir $\tilde{\beta}(v, w) := -\beta(v, w)$ für alle $v, w \in V$. Dann ist auch $\tilde{\beta}\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform, und B, B' sind natürlich weiterhin Orthogonalbasen bezüglich $\tilde{\beta}$. Definieren wir analog die Anzahlen $\tilde{p}, \tilde{q}$

und $\tilde{p}', \tilde{q}'$ für $\tilde{\beta}$ (bezüglich B und B'), so ist offensichtlich $\tilde{p} = q$, $\tilde{q} = p$, $\tilde{p}' = q'$, $\tilde{q}' = p'$. Nach dem vorherigen Argument gilt $\tilde{p} = \tilde{p}'$, also folgt auch $q = \tilde{p} = \tilde{p}' = q'$. □

Folgerung 31.2. *Sei (p, q) die Signatur von β.*
(a) *Es gilt $q = 0$ genau dann, wenn $\beta(v, v) \geqslant 0$ für alle $v \in V$ gilt. In diesem Fall bezeichnen wir β als* ***positiv-semidefinit****.*
(b) *Es gilt $p = \dim V$ und $q = 0$ genau dann, wenn $\beta(v, v) > 0$ für alle $0_V \neq v \in V$ gilt. In diesem Fall heißt β* ***positiv-definit*** *und das Paar (V, β) ein* ***Euklidischer Raum****; außerdem gibt es in diesem Fall eine Orthonormalbasis von V bezüglich β.*
(c) *Ist β positiv-semidefinit und nicht-ausgeartet, so ist β positiv-definit.*

Beweis. Sei $n = \dim V$. Nach Satz 30.15 gibt es eine Orthogonalbasis $B = \{v_1, \ldots, v_n\}$; setze $d_i := \beta(v_i, v_i)$ für alle i. Dann ist $G_B(\beta)$ die Diagonalmatrix mit Einträgen $d_1, \ldots, d_n$ auf der Diagonalen. Sei $v \in V$ und schreibe $v = x_1.v_1 + \ldots + x_n.v_n$ mit $x_i \in \mathbb{R}$. Dann gilt

$$\beta(v, v) = \textstyle\sum_{i=1}^n \sum_{j=1}^n x_i x_j \beta(v_i, v_j) = \sum_{i=1}^n x_i^2 \beta(v_i, v_i) = \sum_{i=1}^n x_i^2 d_i.$$

(a) Sei zuerst $q = 0$, also $d_i \geqslant 0$ für alle i. Dann ist mit obiger Formel auch $\beta(v, v) \geqslant 0$. Sei umgekehrt $\beta(v, v) \geqslant 0$. Insbesondere ist $d_i = \beta(v_i, v_i) \geqslant 0$ für alle i, also $q = 0$.
(b) Sei zuerst $p = n$ und $q = 0$. Wegen $p = \dim V$ ist dann $d_i > 0$ für alle i. Ist also $v \neq 0_V$, so ist $x_i \neq 0$ für mindestens ein i und damit $\beta(v, v) = \sum_{i=1}^n x_i^2 d_i > 0$. Sei umgekehrt $\beta(v, v) > 0$ für alle $0_V \neq v \in V$. Insbesondere ist $d_i = \beta(v_i, v_i) > 0$ für alle i, also $p = n$. Da jedes d_i eine Quadratwurzel in $\mathbb{R}$ hat, gibt es eine Orthonormalbasis wie in Beispiel 30.19.
(c) Da β nicht-ausgeartet ist, gilt $d_1 \cdots d_n = \det(G_B(\beta)) \neq 0$; siehe Bemerkung 30.7. Also ist $d_i \neq 0$ für alle i. Da β positiv-semidefinit ist, gilt außerdem $q = 0$ nach (a), also $d_i \geqslant 0$ für alle i. Also ist $d_i > 0$ für alle i und damit $p = n$, d.h., β ist positiv-definit nach (b). □

Beispiel 31.3. (a) Das Standard-Skalarprodukt $\beta_0 \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ ist positiv-definit, wie bereits zu Beginn von Kapitel IV, §18, bemerkt.
(b) In Einsteins Relativitätstheorie spielen gewisse $\beta \colon \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ mit Signatur $(3, 1)$ eine bedeutende Rolle ("Minkowski–Raum"); siehe z.B. §7.5 im Buch von Huppert–Willems.

Wir beschreiben nun eine allgemeine Konstrukton, um orthogonale Abbildungen in $O(V, \beta)$ zu erhalten (wenn β nicht-ausgeartet ist).

Beispiel 31.4. Sei β nicht-ausgeartet und ein Vektor $0_V \neq r \in V$ gegeben mit $\beta(r, r) \neq 0$. Dann definiere die Abbildung $\varphi_r \colon V \to V$ durch

$$\varphi_r(v) := v - 2\beta(r, r)^{-1}\beta(r, v)r \qquad \text{für alle } v \in V.$$

Man rechnet nach, dass $\varphi_r \in O(V, \beta)$ und $\varphi_r^2 = \mathrm{id}_V$ gilt; außerdem $\varphi_r(r) = -r$ und $\varphi_r(v) = v$ falls $v \perp r$ (siehe Übungen). Die Abbildung φ_r heißt ***Spiegelung*** mit Wurzel r.

Betrachten wir speziell den Euklidischen Raum $(\mathbb{R}^2, \beta_0)$; eine orthogonale Abbildung in $O(\mathbb{R}^2, \beta_0)$ wird dann (bezüglich der Standardbasis von $\mathbb{R}^2$) durch eine Matrix $A \in O_2(\mathbb{R})$ wie in Beispiel 30.13 dargestellt. Mit den dortigen Bezeichnungen sei $\delta = -1$. Dann gilt

$$\chi_A = (\cos(\theta) - X) * (-\cos(\theta) - X) - \sin(\theta)^2 = X^2 - \cos(\theta)^2 - \sin(\theta)^2 = X^2 - 1;$$

es gibt also die beiden Eigenwerte $\lambda_1 = 1$ und $\lambda_2 = -1$. Mit Hilfe der Additionstheoreme für sin und cos in Kapitel II, §10, findet man zugehörige Eigenvektoren (mit Norm 1):

$$v_1 = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix} \in \mathbb{R}^2 \qquad \text{und} \qquad v_2 = \begin{bmatrix} -\sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix} \in \mathbb{R}^2.$$

Dann ist $\{v_1, v_2\}$ sogar eine Orthonormalbasis von $\mathbb{R}^2$. Sei nun $v \in \mathbb{R}^2$ beliebig und schreibe $v = sv_1 + tv_2$ mit $s, t \in \mathbb{R}$. Dann ist $A \cdot v = A \cdot (sv_1 + tv_2) = sv_1 - tv_2$. Sei $\varphi_{v_2} \in O(\mathbb{R}^2, \beta_0)$ die Spiegelung mit Wurzel v_2. Wegen $v_1 \perp v_2$ gilt dann $\varphi_{v_2}(v) = s\varphi_{v_2}(v_1) + t\varphi_{v_2}(v_2) = sv_1 - tv_2 = A \cdot v$. Also wird φ_{v_2} durch die Matrix A beschrieben. (Ist dagegen $\delta = 1$, so überzeugt man sich, dass A die Drehung im Uhrzeigersinn um den Winkel θ darstellt.)

Folgerung 31.5. *Sei β positiv-definit. Gegeben seien $v, w \in V$ mit $\beta(v, v) = \beta(w, w)$. Dann gibt es ein $\varphi \in O(V, \beta)$ mit $\varphi(v) = w$.*

Beweis. Ist $v = w$, so können wir $\varphi = \mathrm{id}_V$ nehmen. Sei nun $v \neq w$ und $r := v - w \neq 0_V$. Da β positiv-definit ist, gilt $\beta(r, r) \neq 0$; wir können also die Spiegelung $\varphi_r \in O(V, \beta)$ bilden. Dann gilt $\varphi_r(v) = v - 2\beta(r, r)^{-1}\beta(r, v)r$. Nun ist $\beta(r, v) = \beta(v - w, v) = \beta(v, v) - \beta(w, v) = \beta(v, v) - \beta(v, w)$; außerdem $\beta(r, r) = \beta(v - w, v - w) = \beta(v, v) - 2\beta(v, w) + \beta(w, w) = 2\beta(v, v) - 2\beta(v, w) = 2\beta(r, v)$. Also ist $\varphi_r(v) = v - r = v - (v - w) = w$. $\square$

Um zu testen, ob unsere gegebene Bilinearform β positiv-definit ist, könnte man gemäß dem Verfahren in Bemerkung 30.16 eine Orthogonalbasis von V bestimmen und dann schauen, ob alle Diagonaleinträge der zugehörigen Gram–Matrix positiv sind. Unser nächstes Ziel ist ein alternatives Kriterium, welches oft nützlich ist.

Lemma 31.6. *Sei $G = G_B(\beta) \in M_n(\mathbb{R})$ die Gram–Matrix von β bezüglich einer Basis B von V. Ist β positiv-definitiv, so gilt $G = T^{\mathrm{tr}} \cdot T$ mit einer invertierbaren Matrix $T \in M_n(K)$ und damit $\det(G) = \det(T)^2 > 0$.*

Beweis. Nach Folgerung 31.2(b) gibt es eine Orthonormalbasis C von V bezüglich β. Dann ist $G_C(\beta) = I_n$ und damit $G = T^{\mathrm{tr}} \cdot T$ wobei $T \in M_n(\mathbb{R})$ die Basiswechselmatrix zwischen B und C ist; siehe Lemma 30.3. Wegen $\det(T^{\mathrm{tr}}) = \det(T)$ folgt also $\det G = \det(T)^2 > 0$. $\square$

Die obige Aussage lässt sich erheblich verstärken. Dazu benutzen wir die folgende Konstruktion zu beliebigen Matrizen, die auch von allgemeinem Interesse ist.

Sei K ein beliebiger Körper und $A = [a_{ij}]_{1 \leqslant i,j \in n} \in M_n(K)$. Setzen wir

$$A_k := \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{bmatrix} \in M_k(K) \qquad \text{für } k = 1, 2, \dots, n,$$

so heißen die Determinanten $\det(A_k)$ für $k = 1, \dots, n$ die ***Hauptminoren*** von A.

Satz 31.7 (LU-Zerlegung). *Sei $A \in M_n(K)$ so, dass alle Hauptminoren von A ungleich 0 sind. Dann gibt es eine untere Dreiecksmatrix $L \in M_n(K)$ mit 1 entlang der Diagonalen, und eine obere Dreiecksmatrix $U \in M_n(K)$ mit $A = L \cdot U$. Ist $K = \mathbb{R}$ und sind alle Hauptminoren von A positiv, so sind auch alle Diagonaleinträge von U positiv.*

(Hier steht L für "lower" und U für "upper".)

Beweis. (Vollständige Induktion nach n.) Ist $n = 1$, also $A = [a_{11}]$, so gilt die Aussage mit $L = [1]$ und $U = [a_{11}]$. Sei nun $n > 1$. Wir teilen A wie folgt in Blöcke auf:

$$A = \left[\begin{array}{c|c} A' & w \\ \hline v^{\mathrm{tr}} & a \end{array}\right] \quad \text{mit} \quad A' \in M_{n-1}(K),\ v, w \in K^{n-1} \ \text{ und } \ a \in K.$$

Nach Voraussetzung sind alle Hauptminoren von A' ungleich 0, also gibt es nach Induktion eine untere Dreiecksmatrix $L' \in M_{n-1}(K)$ mit 1 entlang der Diagonalen, und eine obere Dreiecksmatrix $U' \in M_{n-1}(K)$ mit $A' = L' \cdot U'$. Dann ist

$$A = L_1 \cdot \left[\begin{array}{c|c} I_{n-1} & y \\ \hline x^{\mathrm{tr}} & a \end{array}\right] \cdot U_1 \quad \text{mit} \quad L_1 := \left[\begin{array}{c|c} L' & 0_{n-1} \\ \hline 0_{n-1}^{\mathrm{tr}} & 1 \end{array}\right], \quad U_1 := \left[\begin{array}{c|c} U' & 0_{n-1} \\ \hline 0_{n-1}^{\mathrm{tr}} & 1 \end{array}\right]$$

wobei $x := (U'^{\mathrm{tr}})^{-1} \cdot v \in K^{n-1}$ und $y := (L')^{-1} \cdot w \in K^{n-1}$, wie man einfach mit "Kästchenrechnen" nachprüft. Ebenso rechnet man einfach nach, dass gilt:

$$\left[\begin{array}{c|c} I_{n-1} & y \\ \hline x^{\mathrm{tr}} & a \end{array}\right] = L_2 \cdot U_2 \quad \text{mit} \quad L_2 := \left[\begin{array}{c|c} I_{n-1} & 0_{n-1} \\ \hline x^{\mathrm{tr}} & 1 \end{array}\right], \quad U_2 := \left[\begin{array}{c|c} I_{n-1} & y \\ \hline 0_{n-1}^{\mathrm{tr}} & c \end{array}\right],$$

wobei $c := a - x^{\mathrm{tr}} \cdot y \in K$. Insgesamt also $A = L \cdot U$ mit $L := L_1 \cdot L_2$ und $U := U_2 \cdot U_1$. Beachte: Mit L_1, L_2 ist auch L eine untere Dreiecksmatrix mit 1 entlang der Diagonalen, und mit U_1, U_2 ist U eine obere Dreiecksmatrix. Sei schließlich $K = \mathbb{R}$ und alle Hauptminoren von A positiv. Dies gilt dann auch für A', also sind nach Induktion alle Diagonaleinträge von U' positiv. Damit sich auch alle Diagonaleinträge von U_1 positiv und $\det(U_1) = \det(U') > 0$. Weiterhin ist $\det(U_2) = c$, also $\det(U) = \det(U_2)\det(U_1) = c\det(U_1)$. Wegen $\det(L) = 1$ ergibt dies $0 < \det(A) = \det(L)c\det(U_1) = c\det(U_1)$ und damit $c > 0$, wie behauptet. □

Satz 31.8 (***Determinanten-Kriterium für Definitheit***). *Sei $G = G_B(\beta) \in M_n(\mathbb{R})$ die Gram–Matrix von β bezüglich einer beliebigen Basis B von V. Genau dann ist β positiv-definit, wenn alle Hauptminoren von G positiv (ungleich 0) sind.*

Beweis. Sei $n = \dim V$. Sei zuerst β positiv-definit. Sei $1 \leqslant k \leqslant n$. Schreiben wir $B = \{v_1, \dots, v_n\}$, so sei $U := \langle v_1, \dots, v_k \rangle_K$. Dann ist G_k die Gram–Matrix der Einschränkung von

β auf U bezüglich der Basis $\{v_1, \ldots, v_k\}$ von U. Diese Einschränkung ist natürlich weiterhin positiv-definit, also gilt $\det(G_k) > 0$ nach Lemma 31.6.

Seien nun umgekehrt alle Hauptminoren von G positiv. Betrachte dann eine LU-Zerlegung $G = L \cdot U$ wie in Satz 31.7, wobei alle Diagonaleinträge von U positiv sind. Wir setzen $T := (L^{-1})^{tr} \in M_n(\mathbb{R})$ und $D := T^{tr} \cdot G \cdot T = L^{-1} \cdot G \cdot T = U \cdot T$. Mit L ist auch L^{-1} eine untere Dreiecksmatrix mit 1 entlang der Diagonalen, also ist T eine obere Dreiecksmatrix mit 1 entlang der Diagonalen (siehe Übungen 1. Semester), und damit D eine obere Dreiecksmatrix mit den gleichen Diagonaleinträgen wie U. Nun ist G symmetrisch, also gilt auch $D^{tr} = (T^{tr} \cdot G \cdot T)^{tr} = T^{tr} \cdot G^{tr} \cdot (T^{tr})^{tr} = T^{tr} \cdot G \cdot T = D$. Da gleichzeitig D eine obere Dreiecksmatrix ist, muss D eine Diagonalmatrix sein; wie oben bereits bemerkt, sind die Diagonaleinträge von D die gleichen wie die von U, also alle positiv.

Damit folgt nun leicht, dass β positiv-definit ist. Denn sei $0_V \neq v \in V$ beliebig und $x = M_B(v) \in \mathbb{R}^n$ der Koordinatenvektor von v; sei $y := T^{-1} \cdot x \in \mathbb{R}^n$. Wegen $v \neq 0_V$ ist $x \neq 0_n$ und $y \neq 0_n$. Nach Bemerkung 30.2 gilt $\beta(v, v) = x^{tr} \cdot G \cdot x = y^{tr} \cdot T^{tr} \cdot G \cdot T \cdot y = y^{tr} \cdot D \cdot y = \sum_{i=1}^n y_i^2 d_i$ wobei y_i die Komponenten von y und d_i die Diagonaleinträge von D sind. Wegen $d_i > 0$ für alle i, und $y_i \neq 0$ für mindestens ein i, ist $\beta(v, v) > 0$. $\square$

32. *Der Spektralsatz (über $\mathbb{R}$)*

Der im Titel dieses Abschnittes genannte "Spektralsatz" ist einer der zentralen Sätze der Linearen Algebra. Wir formulieren und beweisen hier zunächst eine Version für Matrizen; danach folgt dann eine abstraktere Umformulierung für Endomorphismen.

Im Folgenden betrachten wir ausschließlich den Euklidischen Raum $(\mathbb{R}^n, \beta_0)$ wobei $\beta_0 = \langle\ ,\ \rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ das Standard-Skalarprodukt ist, also $\beta_0(x, y) = \langle x, y \rangle = x^{tr} \cdot y$ für alle $x, y \in \mathbb{R}^n$. Für $x \in \mathbb{R}^n$ ist dann $\|x\| = \sqrt{x^{tr} \cdot x}$ die übliche Euklidische Norm. Eine Matrix $T \in O_n(\mathbb{R})$ (siehe Beispiel 30.12) heißt hier einfach ***orthogonale Matrix***; es gilt also $T^{tr} \cdot T = I_n$, d.h., T ist invertierbar und die inverse Matrix ist auf besonders einfache Weise gegeben, nämlich durch $T^{-1} = T^{tr}$. (Es folgt dann auch $T \cdot T^{tr} = I_n$.)

Lemma 32.1. *Sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine beliebige lineare Abbildung und $T \in M_n(\mathbb{R})$ die Matrix mit $\varphi(x) = T \cdot x$ für alle $x \in \mathbb{R}^n$. Seien $v_1, \ldots, v_n \in \mathbb{R}^n$ die Spalten von T. Dann gilt $\varphi \in O(\mathbb{R}^n, \beta_0) \Leftrightarrow T \in O_n(\mathbb{R}) \Leftrightarrow \{v_1, \ldots, v_n\}$ ist eine Orthonormalbasis von $\mathbb{R}^n$ bezüglich $\langle\ ,\ \rangle$.*

Beweis. Zunächst beachte: Für $1 \leqslant i, j \leqslant n$ gilt $\varphi(e_i) = T \cdot e_i = v_i$ und

$$\langle v_i, v_j \rangle = \langle T \cdot e_i, T \cdot e_j \rangle = \text{Eintrag an der Stelle } (i, j) \text{ von } T^{tr} \cdot T.$$

Damit folgt bereits: $T^{tr} \cdot T = I_n \Leftrightarrow \{v_1, \ldots, v_n\}$ ist eine Orthonormalbasis, also die 2. Äquivalenz. Sei nun $\varphi \in O(\mathbb{R}^n, \beta_0)$. Dann gilt mit obigen Formeln $\delta_{ij} = \langle e_i, e_j \rangle = \langle \varphi(e_i), \varphi(e_j) \rangle =$

$\langle v_i, v_j\rangle$ = Eintrag an der Stelle (i,j) von $T^{tr} \cdot T$. Also folgt $T^{tr} \cdot T = I_n$, d.h., $T \in O_n(\mathbb{R})$. Ist umgekehrt $T \in O_n(\mathbb{R})$, so folgt $\langle \varphi(x), \varphi(y)\rangle = \langle T\cdot x, T\cdot y\rangle = x^{tr}\cdot (T^{tr}\cdot T)\cdot y = x^{tr}\cdot y = \langle x, y\rangle$ für alle $x, y \in \mathbb{R}^n$, und damit $\varphi \in O(\mathbb{R}^n, \beta_0)$. □

Beispiel 32.2. Sei $0_n \neq v_1 \in \mathbb{R}^n$ mit $\|v_1\| = 1$. Dann gibt es eine orthogonale Matrix $T \in M_n(\mathbb{R})$ deren erste Spalte durch v_1 gegeben ist.
Dazu: Nach dem Basisergänzungssatz gibt es $v_2, \ldots, v_n \in \mathbb{R}^n$, so dass $\{v_1, v_2, \ldots, v_n\}$ eine Basis von $\mathbb{R}^n$ ist. Mit dem Gram–Schmidt–Verfahren in Kapitel IV, Satz 18.3, erhalten wir eine Orthogonalbasis $\{w_1, w_2, \ldots, w_n\}$ von $\mathbb{R}^n$, wobei nach Konstruktion $w_1 = v_1$ gilt. Setze $v_i' := \|w_i\|^{-1} w_i$ für $i = 2, \ldots, n$. Dann ist $\{v_1', v_2', \ldots, v_n'\}$ eine Orthonormalbasis von $\mathbb{R}^n$. Nach Lemma 32.1 ist die Matrix $T \in M_n(\mathbb{R})$ mit Spalten $v_1', v_2', \ldots, v_n'$ orthogonal. □

Lemma 32.3. *Ist $A \in M_n(\mathbb{R})$ symmetrisch, so ist A zerfallend, d.h., das charakteristische Polynom $\chi_A \in \mathbb{R}[X]$ zerfällt vollständig in Linearfaktoren in $\mathbb{R}[X]$.*

Beweis. Wir fassen $A = [a_{ij}]_{1\leqslant i,j\leqslant n}$ als Matrix in $M_n(\mathbb{C})$ auf. Nach dem Fundamentalsatz der Algebra zerfällt χ_A in Linearfaktoren über $\mathbb{C}$. Wir zeigen, dass alle Nullstellen von χ_A bereits in $\mathbb{R}$ sind. Sei also $\lambda \in \mathbb{C}$ eine Nullstelle (= Eigenwert) von A, und $v \in \mathbb{C}^n$ ein zugehöriger Eigenvektor. Seien $x_1, \ldots, x_n \in \mathbb{C}$ die Komponenten von v. Dann gilt also $\sum_{l=1}^n a_{kl}x_l = \lambda x_k$ für $k = 1, \ldots, n$. Wegen $A^{tr} = A$ folgt damit:

$$\sum_{k,l=1}^n a_{kl}x_k\bar{x}_l = \sum_{l=1}^n \Big(\sum_{k=1}^n a_{kl}x_k\Big)\bar{x}_l = \sum_{l=1}^n \Big(\sum_{k=1}^n a_{lk}x_k\Big)\bar{x}_l = \sum_{l=1}^n (\lambda x_l)\bar{x}_l = \lambda \sum_{l=1}^n x_l\bar{x}_l.$$

Mit komplexer Konjugation und $A \in M_n(\mathbb{R})$ gilt andererseits $\sum_{l=1}^n a_{kl}\bar{x}_l = \sum_{l=1}^n \bar{a}_{kl}\bar{x}_l = \overline{\sum_{l=1}^n a_{kl}x_l} = \bar{\lambda}\bar{x}_k$ für $k = 1, \ldots, n$; also folgt:

$$\sum_{k,l=1}^n a_{kl}x_k\bar{x}_l = \sum_{k=1}^n x_k\Big(\sum_{l=1}^n a_{kl}\bar{x}_l\Big) = \sum_{k=1}^n x_k(\bar{\lambda}\bar{x}_k) = \bar{\lambda}\sum_{k=1}^n x_k\bar{x}_k.$$

Nun beachte, dass $\sum_{k=1}^n x_k\bar{x}_k = \sum_{l=1}^n x_l\bar{x}_l$ eine reelle Zahl ist, die wegen $v \neq 0_n$ ungleich 0 (sogar > 0) ist. Also folgt $\lambda = \bar{\lambda}$. □

Lemma 32.4. *Sei $A \in M_n(\mathbb{R})$ symmetrisch. Gegeben seien $r \geqslant 1$ paarweise verschiedene Eigenwerte $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ von A. Für $i = 1, \ldots, r$ sei $w_i \in E_A(\lambda_i)$. Dann gilt $\langle w_i, w_j\rangle = 0$ für $i \neq j$. Insbesondere: Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal zueinander.*

Beweis. Es gilt $A \cdot w_i = \lambda_i w_i$ für alle i. Sei nun $i \neq j$. Dann ist

$$\lambda_i\langle w_i, w_j\rangle = \langle \lambda_i w_i, w_j\rangle = \langle A\cdot w_i, w_j\rangle = (A\cdot w_i)^{tr}\cdot w_j = (w_i^{tr}\cdot A^{tr})\cdot w_j.$$

Da A symmetrisch ist, ist die rechte Seite gleich:

$$(w_i^{tr}\cdot A)\cdot w_j = w_i^{tr}\cdot (A\cdot w_j) = \langle w_i, A\cdot w_j\rangle = \langle w_i, \lambda_j w_j\rangle = \lambda_j\langle w_i, w_j\rangle.$$

Also gilt $(\lambda_i - \lambda_j)\langle w_1, w_2\rangle = 0$. Wegen $\lambda_i \neq \lambda_j$ folgt $\langle w_i, w_j\rangle = 0$. □

Satz 32.5 (***Spektralzerlegung über*** $\mathbb{R}$)**.** *Sei* $A \in M_n(\mathbb{R})$ *symmetrisch. Dann gibt es eine orthogonale Matrix* $T \in M_n(\mathbb{R})$ *und eine Diagonalmatrix* $D \in M_n(\mathbb{R})$ *mit* $T^{tr} \cdot A \cdot T = D$. *Insbesondere ist* A *diagonalisierbar, wobei die Eigenwerte die Diagonaleinträge von* D *sind.*

Ab hier Woche 10

Beweis.[4] Nach Lemma 32.3 ist A zerfallend; sei $Z(\mu_A) = \{\lambda_1, \ldots, \lambda_r\} \subseteq \mathbb{R}$. Wir zeigen nun, dass A diagonalisierbar ist. Dazu verwenden wir das Minimalpolynom–Kriterium in Kapitel IV, Satz 20.15 (siehe auch Bemerkung 26.10). Wir müssen also zeigen, dass jedes λ_i nur eine einfache Nullstelle von μ_A ist. Sei i fest und nehmen wir an, λ_i ist eine mehrfache Nullstelle; d.h., es gilt $(X - \lambda_i)^2 \mid \mu_A$. Schreibe $\mu_A = (X - \lambda_i)^2 * f$ mit $f \in \mathbb{R}[X]$, und setze $g := (X - \lambda_i) * f$. Dann ist $\mathrm{Grad}(g) < \mathrm{Grad}(\mu_A)$, also $g(A) \neq 0_{n\times n}$. Es gibt also ein $v \in \mathbb{R}^n$ mit $w := g(A) \cdot v \neq 0_n$. Dann folgt

$$(*) \qquad 0 \neq \langle w, w\rangle = w^{tr} \cdot w = (g(A) \cdot v)^{tr} \cdot (g(A) \cdot v) = v^{tr} \cdot (g(A)^{tr} \cdot g(A)) \cdot v.$$

Wegen $A^{tr} = A$ gilt auch $g(A)^{tr} = g(A^{tr}) = g(A)$, also $g(A)^{tr} \cdot g(A) = g(A)^2 = g^2(A)$. Nun ist $g^2 = (X-\lambda_i)^2 * f^2 = f * \mu_A$, also folgt $g^2(A) = f(A) \cdot \mu_A(A) = 0_{n\times n}$. Folglich ist die rechte Seite von $(*)$ gleich $v^{tr} \cdot \big(g^2(A) \cdot v\big) = v^{tr} \cdot (0_{n\times n} \cdot v) = 0$, Widerspruch.

Also ist in der Tat A diagonalisierbar. Sei nun $n_i := \dim E_A(\lambda_i)$ für $i = 1, \ldots, r$. Sei zunächst B_i irgendeine Basis von $E_A(\lambda_i)$. Wie in Kapitel IV, Definition 18.4, erklärt, kann man diese mit dem Gram–Schmidt–Verfahren in eine Orthonomalbasis C_i von $E_A(\lambda_i)$ überführen. Wie im Beweis von Kapitel IV, Satz 20.13, folgt, dass $n = n_1 + \ldots + n_r$ gilt und $C := C_1 \cup \ldots \cup C_r$ eine Basis von $\mathbb{R}^n$ ist. Nach Lemma 32.4 ist dies sogar insgesamt eine Orthonormalbasis von $\mathbb{R}^n$. Ist $T \in M_n(\mathbb{R})$ die invertierbare Matrix mit Spalten gegeben durch die Vektoren in C, so ist $T \in O_n(\mathbb{R})$ (siehe Lemma 32.1) und $D := T^{-1} \cdot A \cdot T$ eine Diagonalmatrix. $\square$

Beispiel 32.6. Set $A = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \in M_3(\mathbb{R})$; offenbar ist A symmetrisch. Wir berechnen $\chi_A = \det(A - XI_3) = \ldots = (X + 2) * (X - 1)^2$; es gibt also die beiden Eigenwerte $\lambda_1 = -2$ und $\lambda_2 = 1$. Wir finden $E_A(\lambda_1) = \langle v_1\rangle_{\mathbb{R}}$ und $E_A(\lambda_2) = \langle v_2, v_3\rangle_{\mathbb{R}}$, wobei

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad v_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Es gilt $\langle v_1, v_1\rangle = 3$, $\langle v_2, v_2\rangle = 2$ und $\langle v_2, v_3\rangle = -1$; die Basis $\{v_2, v_3\}$ ist also noch keine Orthogonalbasis des Eigenraums. Wir können nun das Gram–Schmidt–Verfahren anwenden, oder direkt wie folgt vorgehen: Gesucht sind $a, b \in \mathbb{R}$ so dass $0_3 \neq v_3' := (av_2 + bv_3) \perp v_2$ gilt, also $0 = \langle v_2, av_2 + bv_3\rangle = 2a - b$; eine Lösung $(a, b) \neq (0, 0)$ ist gegeben durch $a = 1$, $b = 2$, also $v_3' := 2v_3 + v_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$; dann folgt $\langle v_2, v_3'\rangle = 0$, $\langle v_3', v_3'\rangle = 6$ und $\{v_2, v_3'\}$ ist

[4] Dieser Beweis stammt aus dem Buch von Kaye–Wilson, siehe dort §13.3.

eine Orthogonalbasis des Eigenraums. Sei $T \in M_3(\mathbb{R})$ die Matrix mit Spalten gegeben durch $\sqrt{3}^{-1} v_1$, $\sqrt{2}^{-1} v_2$, $\sqrt{6}^{-1} v_3'$. Nach Lemma 32.1 ist T eine orthogonale Matrix; es gilt dann $T^{tr} \cdot A \cdot T =$ Diagonalmatrix mit Diagonaleinträgen $-2, 1, 1$.

Es folgt nun noch eine Umformulierung von Satz 32.5 in eine Aussage über Endomorphismen eines Euklidischen Raums. Wir erinnern an Bemerkung 25.2, wo wir Eigenwerte und Eigenvektoren für Endomorphismen definiert haben.

Folgerung 32.7 (Spektralzerlegung, 2. Version)**.** *Sei* (V, β) *ein Euklidischer Raum (siehe Folgerung 31.1) und* $\varphi \in \mathrm{End}(V)$ *selbst-adjungiert, d.h., es gilt* $\varphi^* = \varphi$*. Dann gibt es eine Orthonormalbasis von* V *(bezüglich* β*), die aus Eigenvektoren von* φ *besteht.*

Beweis. Nach Folgerung 31.2(b) gibt es eine Orthonormalbasis B von V (bezüglich β). Sei $A = M_B(\varphi) \in M_n(\mathbb{R})$ die darstellende Matrix von φ und $G = I_n$ die Gram-Matrix von β, wobei $n = \dim V$. Wie im Beweis von Satz 30.8 ist $M_B(\varphi^*) = G^{-1} \cdot A^{tr} \cdot G = A^{tr}$. Wegen $\varphi^* = \varphi$ ist die linke Seite auch gleich A, also A symmetrisch. Nach Satz 32.5 gibt es eine orthogonale Matrix $T \in M_n(\mathbb{R})$, so dass $D := T^{-1} \cdot A \cdot T \in M_n(\mathbb{R})$ eine Diagonalmatrix ist.

Sei $B := \{v_1, \ldots, v_n\}$. Schreiben wir $T = [t_{ij}] \in M_n(\mathbb{R})$, so setze $v_j' = \sum_{i=1}^n t_{ij}.v_i$ für $j = 1, \ldots, n$. Weil T invertierbar ist, ist auch $B' := \{v_1', \ldots, v_n'\}$ eine Basis von V. Nach Lemma 30.3 gilt $G_{B'}(\beta) = T^{tr} \cdot G_B(\beta) \cdot T = T^{tr} \cdot I_n \cdot T = I_n$, also ist auch B' eine Orthonormalbasis. Nach Kapitel IV, Satz 20.1, gilt $M_{B'}(\varphi) = T^{-1} \cdot M_B(\varphi) \cdot T = T^{-1} \cdot A \cdot T = D$, also ist jeder Basisvektor v_j' ein Eigenvektor von φ. □

33. *Anwendung 1: Singulärwertzerlegung*

Als eine erste Folgerung aus dem Spektralsatz betrachten wir die ***Singulärwertzerlegung*** einer Matrix (englisch: ***singular value decomposition***, kurz ***SVD***). Diese wird in zahlreichen Anwendungen aus den unterschiedlichsten Bereichen eingesetzt; siehe

C. D. Martin and M. A. Porter, The extraordinary SVD, Amer. Math. Monthly **119** (2012), 838–851; http://dx.doi.org/10.4169/amer.math.monthly.119.10.838

für einen Überblick mit vielen weiterführenden Referenzen. (Sie werden auch sicherlich in Vorlesungen zur angewandten Mathematik mehr dazu erfahren.)

Lemma 33.1. *Sei* $A \in \mathbb{R}^{m \times n}$ *beliebig und* $r := \mathrm{Rang}(A) \leqslant \min(n, m)$*. Dann gilt:*

(a) *Es ist auch* $\mathrm{Rang}(A^{tr} \cdot A) = r$*.*

(b) *Die Eigenwerte von* $A^{tr} \cdot A \in M_n(\mathbb{R})$ *sind reell und* $\geqslant 0$*; es gibt* r *Eigenwerte* $\neq 0$*.*

Beweis. (a) Sei $x \in N(A)$, also $A \cdot x = 0_m$. Dann ist auch $(A^{tr} \cdot A) \cdot x = 0_n$, also $x \in N(A^{tr} \cdot A)$. Umgekehrt: Sei $x \in N(A^{tr} \cdot A)$, also $(A^{tr} \cdot A) \cdot x = 0_n$. Dann ist auch $\|A \cdot x\|^2 = \langle A \cdot x, A \cdot x\rangle = (A \cdot x)^{tr} \cdot (A \cdot x) = x^{tr} \cdot (A^{tr} \cdot A) \cdot x = 0$, also $A \cdot x = 0_m$ und damit $x \in N(A)$. Damit ist $N(A) = N(A^{tr} \cdot A)$ gezeigt, also folgt mit Lemma 20.6 (Kapitel IV):

$$\mathrm{Rang}(A) = n - \dim N(A) = n - \dim N(A^{tr} \cdot A) = \mathrm{Rang}(A^{tr} \cdot A).$$

(b) Sei $B := A^{tr} \cdot A \in M_n(\mathbb{R})$; dann ist B symmetrisch. Nach dem Spekralsatz gibt es eine orthogonale Matrix $V \in M_n(\mathbb{R})$ so dass $D := V^{-1} \cdot B \cdot V = V^{tr} \cdot B \cdot V$ eine Diagonalmatrix ist. Nach Kapitel IV, Bemerkung 19.14, gilt $\mathrm{Rang}(D) = \mathrm{Rang}(B) = r$, d.h., genau r der Diagonaleinträge von D sind ungleich 0. Sei schließlich $\lambda \in \mathbb{R}$ ein Eigenwert von $B = A^{tr} \cdot A$ (also ein Diagonaleintrag von D) und $v \in \mathbb{R}^n$ ein zugehöriger Eigenvektor, also $v \neq 0_n$ und $B \cdot v = \lambda v$. Dann gilt $\|A \cdot v\|^2 = \langle A \cdot v, A \cdot v\rangle = (A \cdot v)^{tr} \cdot (A \cdot v) = v^{tr} \cdot (B \cdot v) = v^{tr} \cdot (\lambda v) = \lambda \|v\|^2$. Wegen $\|v\| > 0$ und $\|A \cdot v\| \geqslant 0$ folgt dann auch $\lambda \geqslant 0$. $\square$

Satz 33.2 (***SVD***)**.** *Sei $A \in \mathbb{R}^{m \times n}$ mit $m, n \geqslant 1$ beliebig; sei $r = \mathrm{Rang}(A) \geqslant 0$. Dann gibt es orthogonale Matrizen $U \in M_m(\mathbb{R})$ und $V \in M_n(\mathbb{R})$ mit $A = U \cdot S \cdot V^{tr}$, wobei $S = [s_{ij}] \in \mathbb{R}^{m \times n}$ Diagonalgestalt hat, mit $s_{ij} = 0$ für alle $i \neq j$ sowie $s_{11} \geqslant \ldots \geqslant s_{rr} > 0$ und $s_{ii} = 0$ für $i > r$. Die Diagonaleinträge $s_{11}, s_{22}, \ldots$ von S heißen **Singulärwerte** von A.*

Beweis. Sei $B := A^{tr} \cdot A \in M_n(\mathbb{R})$. Wie oben im Beweis gibt es eine orthogonale Matrix $V \in M_n(\mathbb{R})$ so dass $V^{-1} \cdot B \cdot V = V^{tr} \cdot B \cdot V$ eine Diagonalmatrix ist; seien $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ die Einträge auf der Diagonalen. Wir könnnen dies so einrichten, dass $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ gilt. Nach Lemma 33.1 gilt $\lambda_i \geqslant 0$ für alle i; außerdem ist $\lambda_i = 0$ für $i > r$. Setze $\alpha_i := \sqrt{\lambda_i} \in \mathbb{R}$ für $1 \leqslant i \leqslant n$; dann ist auch $\alpha_1 \geqslant \ldots \geqslant \alpha_r > \alpha_{r+1} = \ldots = \alpha_n = 0$. Seien $v_1, \ldots, v_n \in \mathbb{R}^n$ die Spalten von V; dann gilt $B \cdot v_i = \lambda_i v_i$ für $1 \leqslant i \leqslant n$. Wir setzen $u_j := \alpha_j^{-1} A \cdot v_j \in \mathbb{R}^m$ für $1 \leqslant j \leqslant r$. Weil V eine orthogonale Matrix ist, ist $\{v_1, \ldots, v_n\}$ eine Orthonormalbasis von $\mathbb{R}^n$. Damit folgt für $1 \leqslant i, j \leqslant r$:

$$\begin{aligned}\langle u_i, u_j\rangle &= \alpha_i^{-1}\alpha_j^{-1}\langle A \cdot v_i, A \cdot v_j\rangle = \alpha_i^{-1}\alpha_j^{-1}\big(v_i^{tr} \cdot A^{tr} \cdot A \cdot v_j\big) = \alpha_i^{-1}\alpha_j^{-1}\big(v_i^{tr} \cdot (B \cdot v_j)\big)\\ &= \alpha_i^{-1}\alpha_j^{-1}\big(v_i^{tr} \cdot (\lambda_j v_j)\big) = \alpha_i^{-1}\alpha_j^{-1}\lambda_j\langle v_i, v_j\rangle.\end{aligned}$$

Ist $i \neq j$, so ist $\langle v_i, v_j\rangle = 0$, also auch $\langle u_i, u_j\rangle = 0$. Ist $i = j$, so ist $\lambda_j = \alpha_i\alpha_j$ und $\langle v_i, v_i\rangle = 1$, also auch $\langle u_i, u_i\rangle = 1$. Aus diesen Relationen folgt sofort, dass $(u_1, \ldots, u_r)$ l.u. ist (siehe Übungen). Sei $W := \langle u_1, \ldots, u_r\rangle_{\mathbb{R}} \subseteq \mathbb{R}^m$. Nach Lemma 30.6 gilt $\dim W^{\perp} = m - \dim W = m - r$. Ist $w \in W \cap W^{\perp}$, so folgt $\langle w, w\rangle = 0$ also $w = 0_m$. Damit ist $W \cap W^{\perp} = \{0_m\}$ und es folgt $\dim(W + W^{\perp}) = \dim W + \dim W^{\perp} - \dim(W \cap W^{\perp}) = \dim W + \dim W^{\perp} = m$, also $\mathbb{R}^m = W + W^{\perp}$. Mit dem Gram–Schmidt–Verfahren können wir eine Orthonormalbasis $\{u_{r+1}, \ldots, u_m\}$ von $W^{\perp}$ finden. Wegen $\mathbb{R}^m = W + W^{\perp}$ und $\dim W + \dim W^{\perp} = m$ ist $\{u_1, \ldots, u_m\}$ insgesamt eine Basis, also eine Orthonormalbasis von $\mathbb{R}^m$. Sei $U \in M_m(\mathbb{R})$ die orthogonale Matrix mit Spalten $u_1, \ldots, u_m$. Jetzt betrachte $S = [s_{ij}] := U^{-1} \cdot A \cdot V =$

$U^{tr} \cdot A \cdot V$. Für $1 \leqslant i \leqslant m$ und $1 \leqslant j \leqslant n$ gilt also $s_{ij} = u_i^{tr} \cdot A \cdot v_j$. Für i beliebig und $1 \leqslant j \leqslant r$ gilt $A \cdot v_j = \alpha_j u_j$ und damit $s_{ij} = u_i^{tr} \cdot A \cdot v_j = \alpha_j \langle u_i, u_j \rangle = \alpha_j \delta_{ij}$. Für $j > r$ ist $\|A \cdot v_j\|^2 = \langle A \cdot v_j, A \cdot v_j \rangle = v_j^{tr} \cdot (A^{tr} \cdot A) \cdot v_j = v_j^{tr} \cdot (B \cdot v_j) = v_j^{tr} \cdot 0_n = 0$, also $A \cdot v_j = 0_m$ und damit auch $s_{ij} = u_i^{tr} \cdot A \cdot v_j = 0$ für alle i. Folglich ist $A = U \cdot S \cdot V^{-1} = U \cdot S \cdot V^{tr}$ und S hat die gewünschte Diagonalgestalt. $\square$

Beachte: Ist $m < n$, so ist es einfacher, erst die SVD für A^{tr} zu bestimmen und am Ende das Ergebnis wieder zu transponieren, um die SVD für A zu erhalten.

Beispiel 33.3. Sei $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3 \times 2}$. Dann ist $B := A^{tr} \cdot A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \in M_2(\mathbb{R})$ mit $\chi_B = (X-3) * (X-1)$. Mit dem Verfahren in Beispiel 32.6 finden wir eine orthogonale Matrix $V \in M_2(\mathbb{R})$, so dass $V^{tr} \cdot B \cdot V$ eine Diagonalmatrix ist; in diesem Fall:

$$V^{tr} \cdot B \cdot V = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{mit} \qquad V = \sqrt{2}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Die Singulärwerte von A sind also $\alpha_1 = \sqrt{3}$ und $\alpha_2 = 1$. Seien $v_1, v_2 \in \mathbb{R}^2$ die beiden Spalten von V. Wie oben im Beweis setzen wir anschließend

$$u_1 := \alpha_1^{-1} A \cdot v_1 = \sqrt{6}^{-1} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \qquad \text{und} \qquad u_2 := \alpha_2^{-1} A \cdot v_2 = \sqrt{2}^{-1} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

Zum Schluss müssen wir noch einen Spaltenvektor $u_3 \in \mathbb{R}^3$ finden, so dass die Matrix $U \in M_3(\mathbb{R})$ mit Spalten u_1, u_2, u_3 eine orthogonale Matrix ist; in diesem Fall geht dies mit

$$u_3 = \sqrt{3}^{-1} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad \text{und damit} \qquad U := \sqrt{6}^{-1} \begin{bmatrix} 1 & \sqrt{3} & \sqrt{2} \\ 2 & 0 & -\sqrt{2} \\ 1 & -\sqrt{3} & \sqrt{2} \end{bmatrix}.$$

Mit $S := \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{3 \times 2}$ erhalten wir also die Singulärwertzerlegung $A = U \cdot S \cdot V^{tr}$.

Und dann ist $A^{tr} = U' \cdot S^{tr} \cdot (V')^{tr}$ mit $U' := V$, $V' := U$ die Singulärwertzerlegung für A^{tr}.

Bemerkung 33.4. Sei $A = U \cdot S \cdot V^{tr}$ und $r = \text{Rang}(A)$, wie in Satz 33.2. Für $1 \leqslant k \leqslant r$ sei $S_{(k)} \in \mathbb{R}^{m \times n}$ die Matrix in Diagonalgestalt mit Diagonaleinträgen $s_{11}, \ldots, s_{kk}$, d.h., $S_{(k)}$ entsteht aus S, indem wir die ersten k Diagonaleinträge in S beibehalten und die restlichen Diagonaleinträge an den Positionen $k+1, \ldots, r$ alle gleich 0 setzen. Dann heißt

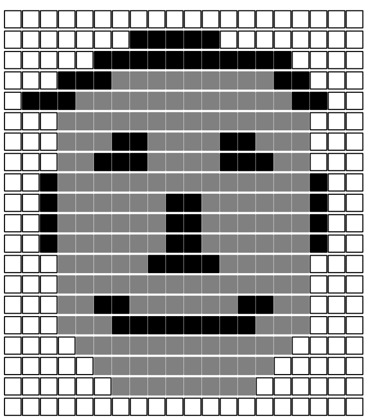
$$A_{(k)} := U \cdot S_{(k)} \cdot V^{tr} \in \mathbb{R}^{m \times n} \qquad \text{“}\textbf{\textit{Rang-}}k\textbf{\textit{-Approximation}}\text{” von } A.$$

Beachte, dass tatsächlich $\text{Rang}(A_{(k)}) = k$ gilt; man kann zeigen, dass $A_{(k)}$ die “beste” Approximation von A durch eine Matrix mit Rang gleich k ist (Satz von Eckart–Young; siehe §1 im oben zitierten Artikel von Martin und Porter). Für $1 \leqslant i \leqslant k$ sei $u_i \in \mathbb{R}^m$ die i-te Spalte von U und $v_i \in \mathbb{R}^n$ die i-te Spalte von V. Dann ist $u_i \cdot v_i^{tr} \in \mathbb{R}^{m \times n}$ und die Definition der Matrixmultiplikation zeigt sofort folgende Formel: $A_{(k)} := \sum_{i=1}^{k} s_{ii} \big(u_i \cdot v_i^{tr}\big)$.

Die mn Einträge von $A_{(k)}$ sind also bestimmt durch die insgesamt $k(m+n+1)$ Einträge in den Spaltenvektoren $u_1, \dots, u_k \in \mathbb{R}^m$, $v_1, \dots, v_k \in \mathbb{R}^n$ plus den k Singulärwerten $s_{11}, \dots, s_{kk}$.

Dies findet Anwendungen zum Beispiel im Bereich der ***Datenkompression***. Denken Sie etwa an die Millionen von Farb- und Helligkeitswerten in einem digitalen Photo; wenn man dies auf einem kleinen Bildschirm betrachten möchte, so braucht man nicht alle diese Werte genau zu kennen. Es genügt also, eine komprimierte Version der Matrix aller Werte zu speichern.

Beispiel 33.5. Zur Veranschaulichung betrachten wir ein extrem vereinfachtes Beispiel, mit digitalen Schwarz-Weiss-Photos in einem 20×20 Raster; jeder Punkt im Raster ist entweder weiss, grau oder schwarz. Wir stellen dies mit einer Matrix $A = [a_{ij}] \in M_{20}(\mathbb{R})$ dar, wo a_{ij} gleich 0, 1 oder 2 ist, je nach dem, ob der entsprechende Punkt im Raster weiss, grau oder schwarz ist. Für ein solches Photo müssen wir also die 400 Einträge von A speichern. (Die Photos mit heutigen Smartphones brauchen natürlich viel mehr Speicherplatz.) Beispiel:



$$A = \begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&2&2&2&2&2&0&0&0&0&0&0&0&0\\
0&0&0&0&0&2&2&2&2&2&2&2&2&2&2&2&0&0&0&0\\
0&0&0&2&2&2&1&1&1&1&1&1&1&1&1&2&2&0&0&0\\
0&2&2&2&1&1&1&1&1&1&1&1&1&1&1&1&2&2&0&0\\
0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&0&0&0\\
0&0&0&1&1&1&2&2&1&1&1&1&2&2&1&1&1&0&0&0\\
0&0&0&1&1&2&2&2&1&1&1&1&2&2&2&1&1&0&0&0\\
0&0&2&1&1&1&1&1&1&1&1&1&1&1&1&1&1&2&0&0\\
0&0&2&1&1&1&1&1&1&2&2&1&1&1&1&1&1&2&0&0\\
0&0&2&1&1&1&1&1&1&2&2&1&1&1&1&1&1&2&0&0\\
0&0&2&1&1&1&1&1&1&2&2&1&1&1&1&1&1&2&0&0\\
0&0&0&1&1&1&1&1&2&2&2&2&1&1&1&1&1&0&0&0\\
0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&0&0&0\\
0&0&0&1&1&2&2&1&1&1&1&1&1&2&2&1&1&0&0&0\\
0&0&0&1&1&1&2&2&2&2&2&2&2&2&1&1&1&0&0&0\\
0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&0&0&0&0\\
0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&0&0&0&0&0\\
0&0&0&0&0&0&1&1&1&1&1&1&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}.$$

Sei $A = U \cdot S \cdot V^{\mathrm{tr}}$ die SVD von A. Zum Beispiel mit Sage erhalten wir diese wie folgt:

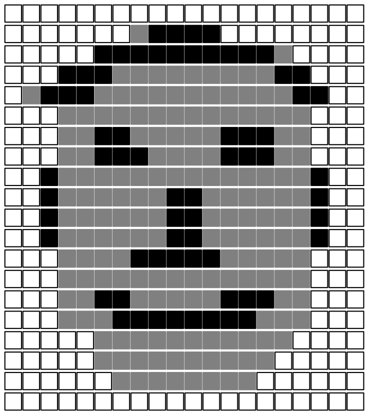
```
SageMath version 9.3, Release Date: 2021-05-09
Using Python 3.9.2. Type "help()" for help.
sage: A=matrix(RDF,[                              # obige Matrix A
  [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0],
  [0,0,0,0,0,0,0,2,2,2,2,2,0,0,0,0,0,0,0,0],
  [0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,0,0,0,0],
  ...
  [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]])
sage: U,S,V=A.SVD()
```

Der Rang von A ist gleich 12; die Singulärwerte ungleich 0 (also die ersten 12 Diagonaleinträge von S) sind wie folgt in Dezimaldarstellung gegeben, gerundet auf 2 Nachkommastellen:

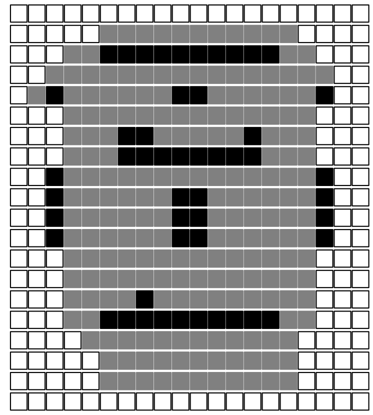
$$18{,}88 \quad 6{,}02 \quad 4{,}71 \quad 3{,}06 \quad 2{,}27 \quad 1{,}79 \quad 1{,}26 \quad 1{,}15 \quad 0{,}79 \quad 0{,}69 \quad 0{,}56 \quad 0{,}40.$$

Betrachten wir die Rang-4-Approximation von A. Runden wir die Einträge von $A_{(4)}$ jeweils zu 0, 1, 2 auf oder ab, so erhalten wir eine Annäherung an das ursprüngliche Bild; siehe Tabelle 1. Nun beachte, dass $A_{(4)}$ gegeben ist durch zwei Matrizen der Größe 20×4, plus die

TABELLE 1. Approximationen an A



Rang-4-Approximation $A_{(4)}$



Rang-2-Approximation $A_{(2)}$

4 Singulärwerte, also brauchen wir für $A_{(4)}$ nur noch $164 = 2 \cdot 80 + 4$ Zahlen zu speichern — deutlich weniger als die Hälfte der ursprünglichen Anzahl! Für die Rang-2-Approximation bräuchten wir noch viel weniger Speicherplatz, aber die Qualität des Bildes ist dagegen ziemlich schlecht; hier bleibt im Wesentlichen nur die ungefähre Form erhalten.
Experimentieren Sie selbst mit der `SVD`–Funktion, um zu sehen, wie sich die Qualität des Bildes für unterschiedliche Werte von $k \in \{1, 2, \ldots, 12\}$ ändert.

34. *Anwendung 2: Quadriken und Hauptachsentransformation*

Als eine klassische Anwendung des Spektralsatzes behandeln wir in diesem Abschnitt einen kleinen Höhepunkt der elementaren analytischen Geometrie. Dabei geht es um Lösungsmengen nicht nur von linearen Gleichungen, sondern von quadratischen Gleichungen in n Variablen. Für $n = 2$ hat eine solche Gleichung die allgemeine Form

$$f(x_1, x_2) := \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_{12} x_1 x_2 + \beta_1 x_1 + \beta_2 x_2 + \gamma = 0,$$

wobei die Koeffizienten $\alpha_1, \alpha_2, \alpha_{12}, \beta_1, \beta_2, \gamma \in \mathbb{R}$ fest vorgegeben und alle $x_1, x_2 \in \mathbb{R}$ gesucht sind, so dass $f(x_1, x_2) = 0$ gilt. Zum Beispiel ist für $f(x_1, x_2) = x_1^2 + x_2^2 - 1$ die Lösungsmenge ein Kreis in der reellen Ebene mit Radius 1 und Mittelpunkt im Ursprung $(0, 0)$. Gibt es mehr Terme in der Gleichung mit Koeffizienten ungleich 0, so wird es schon schwieriger, gleich zu sehen, um welche geometrische Figur es sich handelt; Beispiele:

$$f(x_1, x_2) = 5x_1^2 + 2\sqrt{3}x_1 x_2 + 7x_2^2 - 16 = 0,$$
$$f(x_1, x_2) = x_1^2 + 2x_1 x_2 + x_2^2 + 3x_1 - 4x_2 + 1 = 0.$$

Die Idee ist nun, eine Koordinatentransformation vorzunehmen, so dass in den “neuen” Koordinaten die Gleichung “möglichst einfach” wird. Behandeln wir dies gleich für beliebiges $n \geqslant 1$. Die allgemeine Form einer quadratischen Gleichung ist

$$f(x_1,\dots,x_n) := \underbrace{\alpha_1 x_1^2 + \dots + \alpha_n x_n^2}_{\text{rein quadratisch}} + \underbrace{\sum_{1\leqslant i<j\leqslant n} \alpha_{ij} x_i x_j}_{\text{gemischt quadratisch}} + \underbrace{\beta_1 x_1 + \dots + \beta_n x_n}_{\text{linear}} + \gamma = 0,$$

mit vorgegebenen Koeffizienten $\alpha_i, \alpha_{ij}, \beta_j, \gamma \in \mathbb{R}$. Um triviale Sonderfälle zu vermeiden, nehmen wir an, dass mindestens ein α_i oder ein α_{ij} ungleich 0 ist, also quadratische Terme (rein oder gemischt) tatsächlich vorkommen. Die Lösungsmenge

$$Q := \left\{(x_1,\dots,x_n) \in \mathbb{R}^{1\times n} \,\middle|\, f(x_1,\dots,x_n) = 0\right\}$$

heißt ***Quadrik*** oder ***Hyperfläche zweiten Grades***; für $n = 2$ spricht man auch von ***Kegelschnitten***. Siehe `https://de.wikipedia.org/wiki/Quadrik` für Hintergrund dazu.

Bemerkung 34.1. Sei $T \in M_n(\mathbb{R})$ eine orthogonale Matrix. Ist $x \in \mathbb{R}^n$, so gilt

$$\|T\cdot x\|^2 = \langle T\cdot x, T\cdot x\rangle = (T\cdot x)^{\mathrm{tr}}\cdot(T\cdot x) = x^{\mathrm{tr}}\cdot(T^{\mathrm{tr}}\cdot T)\cdot x = x^{\mathrm{tr}}\cdot x = \|x\|^2,$$

und damit $\|T\cdot x\| = \|x\|$, d.h., T erhält die Norm von Vektoren. Etwas allgemeiner heißt eine Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine ***Bewegung***, wenn es eine orthogonale Matrix $T \in M_n(\mathbb{R})$ und einen Spaltenvektor $u \in \mathbb{R}^n$ gibt mit $\varphi(x) = T\cdot x + u$ für alle $x \in \mathbb{R}^n$. In diesem Fall gilt

$$\|\varphi(x) - \varphi(y)\| = \|T\cdot x - T\cdot y\| = \|T\cdot(x-y)\| = \|x-y\|,$$

für alle $x, y \in \mathbb{R}^n$, d.h., φ erhält die Abstände zwischen Vektoren.

Als geeignete Koordinatentransformation betrachten wir nun eine Bewegung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ wie oben; es gibt also eine orthogonale Matrix $T = [t_{ij}] \in M_n(\mathbb{R})$ und ein $u \in \mathbb{R}^n$ mit $\varphi(y) = T\cdot y + u$ für alle $y \in \mathbb{R}^n$. Seien $u_1,\dots,u_n \in \mathbb{R}$ die Komponenten von u. Jedes $x \in \mathbb{R}^n$ können wir schreiben als $x = \varphi(y) = T\cdot y + u$ mit $y \in \mathbb{R}^n$. Sind $x_1,\dots,x_n \in \mathbb{R}$ die Komponenten von x und $y_1,\dots,y_n \in \mathbb{R}$ die Komponenten von y, so gilt also

$$x_i = \Big(\sum_{j=1}^{n} t_{ij} y_j\Big) + u_i \qquad \text{für } 1 \leqslant i \leqslant n.$$

Setzen wir diese Ausdrücke in die Gleichung $f(x_1,\dots,x_n) = 0$ ein, so erhalten wir eine neue quadratische Gleichung $f'(y_1,\dots,y_n) = 0$ in den Variablen $y_1,\dots,y_n$.

Satz 34.2 (***Normalformen von Quadriken***). *Es gibt eine Bewegung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$, so dass die ursprüngliche quadratische Gleichung $f(x_1,\dots,x_n) = 0$ mittels der Transformation $x = \varphi(y)$ in eine der folgenden Gleichungen in $y_1,\dots,y_n$ überführt werden kann:*

(Typ 1) $\quad \lambda_1 y_1^2 + \dots + \lambda_r y_r^2 + \gamma = 0 \quad$ *mit* $1 \leqslant r \leqslant n$, *oder*

(Typ 2) $\quad \lambda_1 y_1^2 + \dots + \lambda_r y_r^2 + \beta y_n = 0 \quad$ *mit* $1 \leqslant r < n$,

wobei jeweils $\lambda_i, \beta, \gamma \in \mathbb{R}$ sowie $\beta \neq 0$ und $\lambda_i \neq 0$ für $1 \leqslant i \leqslant r$ gilt.

Eine Bewegung, die die ursprüngliche Gleichung $f(x_1,\dots,x_n) = 0$ in eine der Gleichungen 1) oder 2) überführt, heißt ***Hauptachsentransformation***.

Beweis. Wir definieren eine Matrix $A = [a_{ij}] \in M_n(\mathbb{R})$ durch $a_{ij} := \begin{cases} \alpha_i & \text{für } i = j, \\ \alpha_{ij}/2 & \text{für } i < j, \\ \alpha_{ji}/2 & \text{für } i > j; \end{cases}$
dann ist $A \neq 0_{n\times n}$ (weil mindestens ein α_i oder ein α_{ij} ungleich 0 ist). Sei $b \in \mathbb{R}^{1\times n}$ der Zeilenvektor mit Komponenten $\beta_1, \dots, \beta_n$. Dann ist A symmetrisch und es gilt

$$f(x_1, \dots, x_n) = x^{tr} \cdot A \cdot x + b \cdot x + \gamma \qquad \text{für alle } x \in \mathbb{R}^n.$$

Wir zeigen nun, dass man in maximal 4 Schritten diese Gleichung auf Typ (1) oder (2) transformieren kann.

1. Schritt. Nach dem Spektralsatz gibt es eine orthogonale Matrix $T \in M_n(\mathbb{R})$, so dass $D := T^{-1} \cdot A \cdot T = T^{tr} \cdot A \cdot T$ eine Diagonalmatrix ist; seien $\lambda_1 \dots, \lambda_n \in \mathbb{R}$ die Diagonaleinträge von D. Wir können diese so anordnen, dass $\lambda_i \neq 0$ für $1 \leqslant i \leqslant r$ und $\lambda_{r+1} = \dots = \lambda_n = 0$ gilt (wobei $r \in \{1, \dots, n\}$, weil $A \neq 0_{n\times n}$). Betrachte die Bewegung $\varphi_1 \colon \mathbb{R}^n \to \mathbb{R}^n$, $y \mapsto T \cdot y$; schreibe $x = \varphi_1(y)$ und setze $b' := b \cdot T \in \mathbb{R}^{1\times n}$; dann hat die neue Gleichung die Form

$$\begin{aligned} f'(y_1, \dots, y_n) &= (T \cdot y)^{tr} \cdot A \cdot (T \cdot y) + b \cdot (T \cdot y) + \gamma \\ &= y^{tr} \cdot D \cdot y + (b \cdot T) \cdot y + \gamma = \lambda_1 y_1^2 + \dots + \lambda_r y_r^2 + b' \cdot y + \gamma; \end{aligned}$$

d.h., es gibt keine gemischt quadratischen Terme mehr. — Je nach Fall werden wir jetzt noch weitere Bewegungen anwenden müssen, um Typ (1) oder (2) zu erreichen.

2. Schritt. Damit wir nicht ständig neue Namen für die Variablen einführen müssen, nehmen wir nun an, dass die Ursprungsgleichung die gerade erreichte Form hat, also

$$f(x_1, \dots, x_n) = \lambda_1 x_1^2 + \dots + \lambda_r x_r^2 + \beta_1 x_1 + \cdots + \beta_n x_n + \gamma,$$

wobei $\lambda_i, \beta_i, \gamma \in \mathbb{R}$ und $\lambda_i \neq 0$ für $1 \leqslant i \leqslant r$. Betrachte nun die Bewegung $\varphi_2 \colon \mathbb{R}^n \to \mathbb{R}^n$, $y \mapsto y + c$, wobei $c \in \mathbb{R}^n$ ein Spaltenvektor mit Komponenten $c_1, \dots, c_r, 0, \dots, 0$ sei (und $c_1, \dots, c_r \in \mathbb{R}$ noch zu bestimmen sind). Schreibe $x_i = y_i + c_i$ für $1 \leqslant i \leqslant r$ und $x_i = y_i$ für $r + 1 \leqslant i \leqslant n$. Dann ergibt sich als neue Gleichung

$$\begin{aligned} f'(y_1, \dots, y_n) &= \sum_{i=1}^{r} \lambda_i (y_i + c_i)^2 + \sum_{i=1}^{r} \beta_i (y_i + c_i) + \sum_{i=r+1}^{n} \beta_i y_i + \gamma \\ &= \sum_{i=1}^{r} \lambda_i y_i^2 + \sum_{i=1}^{r} (2\lambda_i c_i + \beta_i) y_i + \sum_{i=r+1}^{n} \beta_i y_i + \underbrace{\sum_{i=1}^{r} (\lambda_i c_i^2 + \beta_i c_i) + \gamma}_{\gamma' :=} \end{aligned}$$

Mit $c_i := -\beta_i/(2\lambda_i)$ für $1 \leqslant i \leqslant r$ erhalten wir:

$$f'(y_1, \dots, y_n) = \sum_{i=1}^{r} \lambda_i y_i^2 + \sum_{i=r+1}^{n} \beta_i y_i + \gamma',$$

d.h., die Variablen in den quadratischen und den linearen Termen wurden getrennt.

3. Schritt. Nehmen wir nun wieder an, dass die Ursprungsgleichung die gerade erreichte Form hat, also $f(x_1, \dots, x_n) = \lambda_1 x_1^2 + \dots + \lambda_r x_r^2 + \beta_{r+1} x_{r+1} + \dots + \beta_n x_n + \gamma$, wobei $\lambda_i, \beta_i, \gamma \in \mathbb{R}$ und $\lambda_i \neq 0$ für $1 \leqslant i \leqslant r$. Ist $r = n$ oder $\beta_{r+1} = \dots = \beta_n = 0$, so haben wir Typ (1) erreicht. Nehmen wir schließlich an, es sei $r < n$ und es gibt ein $k \in \{r + 1, \dots, n\}$ mit $\beta_k \neq 0$.

Betrachte nun die Bewegung $\varphi_3\colon \mathbb{R}^n \to \mathbb{R}^n$, $y \mapsto y + c e_k$, wobei $c \in \mathbb{R}$ noch zu bestimmen ist. Schreibe $x_k = y_k + c$ und $x_i = y_i$ für $i \neq k$. Dann ergibt sich als neue Gleichung

$$f'(y_1, \dots, y_n) = \sum_{i=1}^{r} \lambda_i y_i^2 + \sum_{i=r+1}^{n} \beta_i y_i + \underbrace{\beta_k c + \gamma}_{\gamma' :=}.$$

Mit $c := -\gamma/\beta_k$ erhalten wir $\gamma' = 0$, d.h., es gibt keinen konstanten Term mehr.

4. Schritt. Nehmen wir wiederum an, dass die Ursprungsgleichung die gerade erreichte Form hat, also $f(x_1, \dots, x_n) = \lambda_1 x_1^2 + \dots + \lambda_r x_r^2 + \beta_{r+1} x_{r+1} + \dots + \beta_n x_n$, wobei $\lambda_i, \beta_i \in \mathbb{R}$ und $\lambda_i \neq 0$ für $1 \leqslant i \leqslant r$; außerdem ist mindestens ein β_i ungleich 0. Um Typ (2) zu erreichen, müssen wir noch eine Bewegung $\varphi_4\colon \mathbb{R}^n \to \mathbb{R}^n$ finden, so dass nach der Transformation $x = \varphi_4(y)$ nur der lineare Term in y_n übrig bleibt.

Dazu: Sei $v \in \mathbb{R}^n$ der Spaltenvektor mit Komponenten $0, \dots, 0, \beta_{r+1}, \dots, \beta_n$. Dann ist $v \neq 0_n$; sei $\beta := \|v\| > 0$. Sei $w := \beta e_n \in \mathbb{R}^n$; dann ist $\|w\| = \beta = \|v\|$. Ist $v = w$, so sind wir fertig. Sei nun $v \neq w$. Nach Folgerung 31.5 gibt es ein $\varphi_4 \in O(\mathbb{R}^n, \beta_0)$ mit $\varphi_4(v) = w$ (nämlich die Spiegelung mit Wurzel $v - w$). Sei $T \in M_n(\mathbb{R})$ die Matrix mit $\varphi_4(x) = T \cdot x$ für alle $x \in \mathbb{R}^n$; nach Lemma 32.1 ist $T \in O_n(\mathbb{R})$. Schreibe wieder $x = \varphi_4(y) = T \cdot y$; da die ersten r Komponenten von v und w gleich 0 sind, gilt $e_i \perp (v - w)$ und damit $T \cdot e_i = \varphi_4(e_i) = e_i$, also auch $x_i = y_i$ für $1 \leqslant i \leqslant r$. Als neue Gleichung ergibt sich:

$$f'(y_1, \dots, y_n) = \sum_{i=1}^{r} \lambda_i y_i^2 + \sum_{i=r+1}^{n} \beta_i \Big(\sum_{j=1}^{n} t_{ij} y_j \Big) = \sum_{i=1}^{r} \lambda_i y_i^2 + \underbrace{\sum_{j=1}^{n} \Big(\sum_{i=r+1}^{n} \beta_i t_{ij} \Big) y_j}_{=(v^{\mathrm{tr}} \cdot T) \cdot y}.$$

Nun ist $T \cdot w = v$, also $v^{\mathrm{tr}} = w^{\mathrm{tr}} \cdot T^{\mathrm{tr}} = w^{\mathrm{tr}} \cdot T^{-1}$ und damit $v^{\mathrm{tr}} \cdot T = w^{\mathrm{tr}} = \beta e_n^{\mathrm{tr}}$. Schließlich folgt $(v^{\mathrm{tr}} \cdot T) \cdot y = \beta e_n^{\mathrm{tr}} \cdot y = \beta y_n$, wie gewünscht. □

Beispiel 34.3. Sei $f(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2^2 + 3x_1 - 4x_2 + 1$. Wir gehen die obigen Schritte durch. Die Matrix A ist gegeben durch $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, mit $\chi_A = \det(A - XI_2) = X^2 - 2X$; es gibt also die beiden Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = 0$. Wie in Beispiel 32.6 finden wir

$$T^{-1}AT = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{mit der orthogonalen Matrix } T = \sqrt{2}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Die Transformation $x_1 = \sqrt{2}^{-1}(y_1 + y_2)$, $x_2 = \sqrt{2}^{-1}(y_1 - y_2)$ ergibt die neue Gleichung

$$f'(y_1, y_2) = 2y_1^2 - \sqrt{2}^{-1} y_1 + 7\sqrt{2}^{-1} y_2 + 1 = 0.$$

Schreibe nun $y_1 = z_1 + u$, $y_2 = z_2$ wobei $u \in \mathbb{R}$ noch zu bestimmen ist. Einsetzen ergibt

$$2(z_1+u)^2 - \sqrt{2}^{-1}(z_1+u) + 7\sqrt{2}^{-1} z_2 + 1 = 2z_1^2 + 4z_1u + 2u^2 - \sqrt{2}^{-1} z_1 - \sqrt{2}^{1} u + 7\sqrt{2}^{-1} z_2 + 1.$$

Mit $u := (4\sqrt{2})^{-1}$ erhalten wir die Gleichung $f''(z_1, z_2) = 2z_1^2 + 7\sqrt{2}^{-1} z_2 + 15/16 = 0$. Schließlich schreibe $z_1 = w_1$, $z_2 = w_2 + r$ wobei $r \in \mathbb{R}$ so zu bestimmen ist, dass der konstante Term verschwindet. Mit $r = -15\sqrt{2}/112$ erhalten wir die Gleichung $f'''(w_1, w_2) =$

$2w_1^2 + 7\sqrt{2}^{-1}w_2 = 0$ (d.h. Typ 2), was in der reellen Ebene $\mathbb{R}^2$ (mit Koordinatenachsen w_1, w_2) einer Parabel entspricht.

Ab hier Woche 11

35. *Normale Matrizen und die Spektralzerlegung über* $\mathbb{C}$

In diesem kurzen Abschnitt wollen wir untersuchen, unter welchen Bedingungen eine Spektralzerlegung für Matrizen über $K = \mathbb{C}$ möglich ist. (Dieser Abschnitt ist logisch unabhängig von den vorherigen Abschnitten, obwohl ähnliche Argumente verwendet werden; er hätte also auch am Anfang dieses Kapitels stehen können.) Um das Standard-Skalarprodukt von $\mathbb{R}^n$ auf $\mathbb{C}^n$ zu verallgemeinern, ist es sinnvoll, irgendwie die komplexe Konjugation ins Spiel zu bringen. Für $n = 1$ ist der Absolutbetrag von $z \in \mathbb{C}$ durch $|z| := \sqrt{z\bar{z}}$ definiert. Für $n \geqslant 1$ beliebig definieren wir nun eine Abbildung $\langle\ ,\ \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ durch

$$\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle := x_1\bar{y}_1 + \ldots + x_n\bar{y}_n \qquad \text{für alle } x_i, y_j \in \mathbb{C}.$$

Sei $x \in \mathbb{C}^n$ mit Komponenten $x_1, \ldots, x_n \in \mathbb{C}$. Dann definieren wir $\bar{x} \in \mathbb{C}^n$ als den Spaltenvektor mit Komponenten $\bar{x}_1, \ldots, \bar{x}_n \in \mathbb{C}$. Sei $y \in \mathbb{C}^n$ mit Komponenten $y_1, \ldots, y_n \in \mathbb{C}$. Dann gilt die Regel:

$$\overline{\langle x, y\rangle} = \overline{x_1\bar{y}_1 + \ldots + x_n\bar{y}_n} = \bar{x}_1\bar{\bar{y}}_1 + \ldots + \bar{x}_n\bar{\bar{y}}_n = y_1\bar{x}_1 + \ldots + y_n\bar{x}_n = \langle y, x\rangle.$$

Insbesondere folgt $\langle x, x\rangle = \overline{\langle x, x\rangle} = |x_1|^2 + \ldots + |x_n|^2 \in \mathbb{R}$ für alle $x \in \mathbb{C}^n$. Das Ergebnis ist sogar $\geqslant 0$, also können wir $\|x\| := \sqrt{\langle x, x\rangle}$ als die Norm von x definieren; beachte: Auch hier gilt $\|x\| = 0$ nur für $x = 0_n$. Wir bezeichnen $\langle\ ,\ \rangle$ als die standard-***hermitesche Form*** auf $\mathbb{C}^n$. Analog zu Kapitel IV, §18, können wir dann auch schreiben:

$$\langle x, y\rangle = x^{\mathrm{tr}} \cdot \bar{y} \qquad \text{für alle } x, y \in \mathbb{C}^n,$$

wobei das Ergebnis eine 1×1-Matrix der Form $[u]$ mit $u \in \mathbb{C}$ ist, wofür wir nach unseren Konventionen einfach nur $u \in \mathbb{C}$ schreiben.

Satz 35.1 (***Gram–Schmidt–Orthogonalisierung über*** $\mathbb{C}$). *Sei* $U \subseteq \mathbb{C}^n$ *ein Teilraum mit* $m := \dim U \geqslant 1$ *und* $\{v_1, \ldots, v_m\}$ *eine Basis von* U. *Dann gibt es eine Basis* $\{w_1, \ldots, w_m\}$ *von* U *mit folgenden Eigenschaften:*

(a) $d_i := \|w_i\|^2 > 0$ *und* $\langle w_i, w_j\rangle = 0$ *für alle* $1 \leqslant i, j \leqslant m$ *mit* $i \neq j$*;*

(b) $w_1 = v_1$ *und* $w_k = v_k - \sum_{j=1}^{k-1} d_j^{-1}\langle v_k, w_j\rangle . w_j$ *für* $k = 2, 3, \ldots, m$.

Beweis. Völlig analog zu Satz 18.3 in Kapitel IV. □

Definition 35.2. (a) Sei $A = [a_{ij}] \in M_n(\mathbb{C})$ beliebig. Dann setzen wir $\bar{A} := [\bar{a}_{ij}] \in M_n(\mathbb{C})$ und $A^* := \bar{A}^{\mathrm{tr}}$. Man sieht sofort $(A^*)^* = A$ und $(A \cdot B)^* = B^* \cdot A^*$ für $B \in M_n(\mathbb{C})$.

Ist $A = A^*$ so heißt A eine ***hermitesche Matrix***. (Ist also $A \in M_n(\mathbb{R})$, so ist A hermitesch genau dann, wenn A symmetrisch ist.)

(b) Eine Matrix $U \in M_n(\mathbb{C})$ heißt ***unitäre Matrix***, wenn $U^* \cdot U = I_n$ gilt. In diesem Fall ist also U invertierbar, und die inverse Matrix ist wiederum auf besonders einfache Weise gegeben, nämlich durch $U^{-1} = U^*$. (Es folgt dann auch $U \cdot U^* = I_n$.)

Lemma 35.3. *Sei $U \in M_n(\mathbb{C})$ eine beliebige Matrix; seien $v_1, \ldots, v_n \in \mathbb{C}^n$ die Spalten von U. Genau dann ist U unitär, wenn $B = \{v_1, \ldots, v_n\}$ eine Orthonormalbasis von $\mathbb{C}^n$ bezüglich der standard-hermiteschen Form $\langle\ ,\ \rangle\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$, $(x, y) \mapsto x^{\mathrm{tr}} \cdot \bar{y}$, ist.*

Beweis. Für $1 \leqslant i, j \leqslant n$ gilt $\langle v_i, v_j\rangle = \langle U \cdot e_i, U \cdot e_j\rangle = (U \cdot e_i)^{\mathrm{tr}} \cdot \overline{U \cdot e_j} = e_i^{\mathrm{tr}} \cdot (U^{\mathrm{tr}} \cdot \bar{U}) \cdot \bar{e}_j =$ Eintrag an der Stelle (i, j) von $U^{\mathrm{tr}} \cdot \bar{U} = (U^* \cdot U)^{\mathrm{tr}}$. Also ist $U^* \cdot U = I_n$ genau dann, wenn $\langle v_i, v_i\rangle = 1$ und $\langle v_i, v_j\rangle = 0$ für $1 \leqslant i, j \leqslant n$ mit $i \neq j$ gilt. □

Beispiel 35.4. (a) Für $\theta \in \mathbb{R}$, $0 \leqslant \theta < 2\pi$, und $z, w \in \mathbb{C}$ mit $|z|^2 + |w|^2 = 1$ setzen wir

$$U := \begin{bmatrix} w & z \\ -u(\theta)\bar{z} & u(\theta)\bar{w} \end{bmatrix} \in M_2(\mathbb{C}) \qquad \text{wobei} \qquad u(\theta) := \cos(\theta) + \sin(\theta)i \in \mathbb{C}.$$

Dann prüft man leicht nach, dass U unitär ist. Umgekehrt kann man zeigen, dass jede unitäre Matrix in $M_2(\mathbb{C})$ obige Form hat. (Dies ist eine sehr gute Übungsaufgabe.)

(b) Sei $0_n \neq v_1 \in \mathbb{C}^n$ mit $\|v_1\| = 1$. Völlig analog zu Beispiel 32.2 gibt es dann eine unitäre Matrix $U \in M_n(\mathbb{R})$ deren erste Spalte durch v_1 gegeben ist.

Eine Matrix $A \in M_n(\mathbb{C})$ heißt ***unitär diagonalisierbar***, wenn es eine unitäre Matrix $U \in M_n(\mathbb{C})$ gibt, so dass $U^{-1} \cdot A \cdot U = U^* \cdot A \cdot U$ eine Diagonalmatrix ist. Wir wollen herausfinden, unter welchen Bedingungen dies möglich ist.

Bemerkung 35.5. Sei $A \in M_n(\mathbb{C})$ unitär diagonalisierbar. Dann gilt $A \cdot A^* = A^* \cdot A$.

Denn sei $U \in M_n(\mathbb{C})$ unitär, so dass $D := U^* \cdot A \cdot U$ eine Diagonalmatrix ist. Dann ist $A = U \cdot D \cdot U^*$ und $A^* = (U \cdot D \cdot U^*)^* = (U^*)^* \cdot D^* \cdot U^* = U \cdot D^* \cdot U^*$. Es folgt $A \cdot A^* = (U \cdot D \cdot U^*) \cdot (U \cdot D^* \cdot U^*) = U \cdot (D \cdot D^*) \cdot U^*$ und analog $A^* \cdot A = U \cdot (D^* \cdot D) \cdot U^*$. Da D und D^* beide Diagonalmatrizen sind, gilt $D \cdot D^* = D^* \cdot D$ und damit auch $A \cdot A^* = A^* \cdot A$.

Ist $A \in M_n(\mathbb{C})$ und gilt $A \cdot A^* = A^* \cdot A$, so heißt A eine ***normale Matrix***. Zum Beispiel sind hermitesche oder unitäre Matrizen automatisch auch normale Matrizen.

Lemma 35.6. *Ist $A \in M_n(\mathbb{C})$ eine normale Matrix, so gilt $\|A \cdot x\| = \|A^* \cdot x\|$ für $x \in \mathbb{C}^n$.*

Beweis. Sei $x \in \mathbb{C}^n$. Dann ist $(A^* \cdot x)^{\mathrm{tr}} = (\bar{A}^{\mathrm{tr}} \cdot x)^{\mathrm{tr}} = x^{\mathrm{tr}} \cdot \bar{A}$ und damit

$$\|A^* \cdot x\|^2 = \langle A^* \cdot x, A^* \cdot x\rangle = (A^* \cdot x)^{\mathrm{tr}} \cdot \overline{A^* \cdot x} = x^{\mathrm{tr}} \cdot (\bar{A} \cdot \overline{A^*}) \cdot \bar{x} = x^{\mathrm{tr}} \cdot (\overline{A \cdot A^*}) \cdot \bar{x};$$

andererseits gilt auch $\|A\cdot x\|^2 = \langle A\cdot x, A\cdot x\rangle = (A\cdot x)^{\mathrm{tr}} \cdot \overline{A\cdot x} = x^{\mathrm{tr}}\cdot(A^{\mathrm{tr}} \cdot \bar{A})\cdot\bar{x} = x^{\mathrm{tr}}\cdot(\overline{A^* \cdot A})\cdot\bar{x}$. Nach Voraussetzung sind die beiden rechten Seiten gleich. $\square$

Satz 35.7 (***Spektralzerlegung über*** $\mathbb{C}$)**.** *Sei* $A \in M_n(\mathbb{C})$ *eine normale Matrix, also* $A\cdot A^* = A^* \cdot A$*. Dann ist* A *unitär diagonalisierbar; es gibt also eine unitäre Matrix* $U \in M_n(\mathbb{C})$ *so dass* $D := U^{-1} \cdot A \cdot U = U^* \cdot A \cdot U \in M_n(\mathbb{C})$ *eine Diagonalmatrix ist.*

Beweis. (Vollständige Induktion nach n.) Ist $n = 1$, so ist $A = [a]$ mit $a \in \mathbb{C}$ und die Aussage gilt mit $U = [1]$, $D = [a]$. Sei nun $n \geqslant 2$. Da $\mathbb{C}$ algebraisch abgeschlossen ist, hat das Minimalpolynom $\mu_A \in \mathbb{C}[X]$ eine Nullstelle, also gibt es einen Eigenwert $\lambda_1 \in \mathbb{C}$; sei $w_1 \in \mathbb{C}^n$ ein zugehöriger Eigenvektor, d.h., $w_1 \neq 0_n$ und $A \cdot w_1 = \lambda_1 w_1$. Setzen wir $v_1 := \|w_1\|^{-1} w_1 \in \mathbb{C}^n$, so ist auch v_1 ein Eigenvektor (mit Eigenwert λ_1) und $\|v_1\| = 1$. Nach Beispiel 35.4(b) gibt es eine unitäre Matrix $U_1 \in M_n(\mathbb{C})$ deren erste Spalte durch v_1 gegeben ist. Wir setzen $A_1 := U_1^* \cdot A \cdot U_1 \in M_n(\mathbb{C})$. Wir zeigen nun, dass auch A_1 normal ist. Dazu: Es gilt $A_1^* = (U_1^* \cdot A \cdot U_1)^* = U_1^* \cdot A^* \cdot (U_1^*)^* = U_1^* \cdot A^* \cdot U_1$. Mit $A \cdot A^* = A^* \cdot A$ erhalten wir $A_1 \cdot A_1^* = (U_1^* \cdot A) \cdot (U_1 \cdot U_1^*) \cdot (A^* \cdot U_1) = U_1^* \cdot (A \cdot A^*) \cdot U_1 = U_1^* \cdot (A^* \cdot A) \cdot U_1 = (U_1^* \cdot A^*) \cdot (U_1 \cdot U_1^*) \cdot (A \cdot U_1) = A_1^* \cdot A_1$, wie behauptet.

Nun ist $U_1 \cdot e_1 = v_1$, also ist die erste Spalte von A_1 gegeben durch $A_1 \cdot e_1 = (U_1^* \cdot A)\cdot(U_1 \cdot e_1) = U_1^* \cdot (A \cdot v_1) = \lambda_1(U_1^* \cdot v_1) = \lambda_1(U_1^{-1} \cdot v_1) = \lambda_1 e_1$. Die Einträge in der ersten Spalte von A_1 sind also $\lambda_1, 0, \ldots, 0$; insbesondere ist $\|A_1 \cdot e_1\| = |\lambda_1|$. Seien $b_1, \ldots, b_n \in \mathbb{C}$ die Einträge in der ersten Zeile von A_1; insbesondere $b_1 = \lambda_1$. Wegen $A_1^* = \overline{A_1}^{\mathrm{tr}}$ sind dann $\bar{b}_1, \ldots, \bar{b}_n$ die Einträge in der ersten Spalte von A_1^*. Mit Lemma 35.6 folgt

$$|\lambda_1|^2 + |b_2|^2 + \ldots + |b_n|^2 = |b_1|^2 + |b_2|^2 + \ldots + |b_n|^2 = \|A_1^* \cdot e_1\|^2 = \|A_1 \cdot e_1\|^2 = |\lambda_1|^2,$$

und damit $b_2 = \ldots = b_n = 0$. Also hat A_1 eine Blockdiagonalgestalt wie folgt:

$$U_1^* \cdot A \cdot U_1 = A_1 = \left[\begin{array}{c|c} \lambda_1 & 0_{n-1}^{\mathrm{tr}} \\ \hline 0_{n-1} & A_2 \end{array}\right] \qquad \text{mit} \qquad A_2 \in M_{n-1}(\mathbb{C}).$$

Da A_1 normal ist, folgt mit einer leichten Rechnung, dass auch A_2 normal ist. Nach Induktion gibt es also eine unitäre Matrix $U_2 \in M_{n-1}(\mathbb{C})$, so dass $D_2 := U_2^* \cdot A_2 \cdot U_2 \in M_{n-1}(\mathbb{C})$ eine Diagonalmatrix ist. Setzen wir

$$U := U_1 \cdot U_2' \qquad \text{mit} \qquad U_2' := \left[\begin{array}{c|c} 1 & 0_{n-1}^{\mathrm{tr}} \\ \hline 0_{n-1} & U_2 \end{array}\right] \in M_n(\mathbb{C})$$

so sieht man sofort, dass U_2' und dann U unitär sind; außerdem gilt

$$U^* \cdot A \cdot U = (U_2')^* \cdot \big(U_1^* \cdot A \cdot U_1\big) \cdot U_2' = (U_2')^* \cdot \left[\begin{array}{c|c} \lambda_1 & 0_{n-1}^{\mathrm{tr}} \\ \hline 0_{n-1} & A_2 \end{array}\right] \cdot U_2' = \left[\begin{array}{c|c} \lambda_1 & 0_{n-1}^{\mathrm{tr}} \\ \hline 0_{n-1} & U_2^* \cdot A_2 \cdot U_2 \end{array}\right],$$

und die rechte Seite ist eine Diagonalmatrix, deren Diagonaleinträge durch λ_1 und die Diagonaleinträge von D_2 gegeben sind. $\square$

Ist $A \in M_n(\mathbb{C})$ normal und $D = U^* \cdot A \cdot U$ die Spektralzerlegung wie oben, so sind die Diagonaleinträge von D genau die Eigenwerte von A. Für hermitesche Matrizen ergibt sich:

Lemma 35.8. *Sei $A = [a_{ij}] \in M_n(\mathbb{C})$ beliebig.*
(a) *Ist $\lambda \in \mathbb{C}$ ein Eigenwert von A, so ist $\bar{\lambda} \in \mathbb{C}$ ein Eigenwert von $\bar{A}$.*
(b) *Ist A hermitesch (also $A = A^*$), so sind alle Eigenwerte von A reell.*

Beweis. (a) Sei $v \in \mathbb{C}^n$ ein zu λ gehöriger Eigenvektor, also $v \neq 0_n$ und $A \cdot v = \lambda v$. Seien $z_1, \ldots, z_n \in \mathbb{C}$ die Komponenten von v. Für $i \in \{1, \ldots, n\}$ ist dann die i-te Komponente von $A \cdot v$ gleich λz_i. Sei $\bar{v} \in \mathbb{C}^n$ der Spaltenvektor mit Komponenten $\bar{z}_1, \ldots, \bar{z}_n$. Die i-te Komponente von $\bar{A} \cdot \bar{v}$ ist dann gegeben durch

$$\sum_{j=1}^{n} \bar{a}_{ij}\bar{z}_j = \sum_{j=1}^{n} \overline{a_{ij}z_j} = \overline{\sum_{1 \leqslant j \leqslant n} a_{ij}z_j} = \overline{\lambda z_i} = \bar{\lambda}\bar{z}_i;$$

also gilt $\bar{A} \cdot \bar{v} = \bar{\lambda}\bar{v}$ und damit ist $\bar{\lambda}$ ein Eigenwert von $\bar{A}$, mit Eigenvektor $\bar{v}$.

(b) Sei λ ein Eigenwert von A, mit Eigenvektor $0_n \neq v \in \mathbb{C}^n$. Nach (a) ist $\bar{\lambda}$ ein Eigenwert von $\bar{A}$, mit Eigenvektor $\bar{v} \in \mathbb{C}^n$. Wegen $\bar{A} = A^{\mathrm{tr}}$ folgt einerseits

$$v^{\mathrm{tr}} \cdot \bar{A} \cdot \bar{v} = (v^{\mathrm{tr}} \cdot A^{\mathrm{tr}}) \cdot \bar{v} = (A \cdot v)^{\mathrm{tr}} \cdot \bar{v} = (\lambda v^{\mathrm{tr}}) \cdot \bar{v} = \lambda(v^{\mathrm{tr}} \cdot \bar{v}) = \lambda\|v\|^2.$$

Andererseits ist die linke Seite auch gleich $v^{\mathrm{tr}} \cdot (\bar{A} \cdot \bar{v}) = v^{\mathrm{tr}} \cdot (\bar{\lambda}\bar{v}) = \bar{\lambda}(v^{\mathrm{tr}} \cdot \bar{v}) = \bar{\lambda}\|v\|^2$. Also folgt $\bar{\lambda}\|v\|^2 = \lambda\|v\|^2$. Wegen $v \neq 0_n$ ist $\|v\| \neq 0$ und damit $\lambda = \bar{\lambda}$. □

Bemerkung 35.9. Sei $A \in M_n(\mathbb{C})$ eine normale Matrix. Um die Spektralzerlegung in Satz 35.7 zu berechnen, kann man entweder induktiv wie im obigen Beweis vorgehen, oder mit einem Verfahren analog zu Beispiel 32.6: also zuerst Eigenwerte bestimmen, dann Basen für die Eigenräume, sowie schließlich Orthonormalbasen dieser Eigenräume (zum Beispiel mit Gram–Schmidt).

Sei zum Beispiel $A = \begin{bmatrix} 1 & i & i & 1 \\ -i & 1 & 1 & -i \\ -i & 1 & 3 & i \\ 1 & i & -i & 3 \end{bmatrix} \in M_4(\mathbb{C})$. Dann ist $A = A^*$, also insbesondere A eine normale Matrix. Wir erhalten $\chi_A = \det(A - XI_4) = \ldots = X^2 * (X-4)^2$; es gibt also die beiden Eigenwerte $\lambda_1 = 0$ und $\lambda_2 = 4$. Mit dem obigen Verfahren ergibt sich mit

$$U := \frac{1}{2\sqrt{3}} \begin{bmatrix} -i\sqrt{6} & -\sqrt{3} & i\sqrt{2} & 1 \\ \sqrt{6} & i\sqrt{3} & \sqrt{2} & -i \\ 0 & -i\sqrt{3} & 2\sqrt{2} & i \\ 0 & \sqrt{3} & 0 & 3 \end{bmatrix} \in M_4(\mathbb{C}),$$

dass U unitär ist und $D := U^* \cdot A \cdot U$ eine Diagonalmatrix mit Einträgen $0, 0, 4, 4$ auf der Diagonalen; Details in den Übungen (oder selbst).

Kapitel VIII: Allgemeine Theorie der Vektorräume

In diesem Kapitel ist K wieder ein beliebiger Körper und V ein Vektorraum über K. Es geht hier darum, erstens einige der Begriffe aus früheren Kapiteln von Matrizen auf Endomorphismen von V zu verallgemeinern, und zweitens neue Methoden und Konstruktionen einzuführen, die ohne irgendwelche Voraussetzungen an die Dimension von V funktionieren.

36. ***Das Lemma von Zorn und die Existenz von Basen***

Sei A eine nicht-leere Menge und $\leqslant$ eine Ordnungsrelation auf A, also $\leqslant$ reflexiv, antisymmetrisch und transitiv (siehe Kapitel I, §4). Die Relation $\leqslant$ heißt eine ***totale Ordnung***, wenn $a \leqslant b$ oder $b \leqslant a$ für alle $a, b \in A$ gilt. Zum Beispiel ist die übliche Relation $\leqslant$ auf $A = \mathbb{N}$ eine totale Ordnung. Ist andererseits zum Beispiel $A = \mathcal{P}(X)$ die Potenzmenge einer nicht-leeren Menge X, so ist $\subseteq$ eine Ordnungsrelation, die aber keine totale Ordnung ist.
Das Lemma von Zorn ist eine rein mengentheoretische Aussage, die unter bestimmten Bedingungen die Existenz von maximalen Elementen in A garantiert. Ein Element $a \in A$ heißt ***maximales Element***, wenn es kein $b \in A$ gibt mit $a < b$. (Allgemein schreiben wir $a < b$, wenn $a \leqslant b$ gilt und $a \neq b$.) Für endliche Mengen ist dies kein Problem:

Lemma 36.1. *Sei A eine endliche Menge. Dann besitzt A ein maximales Element.*

Beweis. Sei $d := |A| \geqslant 1$ und nehmen wir an, es gibt kein maximales Element. Sei $a_1 \in A$ beliebig. Dann ist a_1 nicht maximal, also gibt es ein $a_2 \in A$ mit $a_1 < a_2$. Wiederum ist a_2 nicht maximal, also gibt es ein $a_3 \in A$ mit $a_2 < a_3$. Nach d Schritten haben wir Elemente $a_1, \ldots, a_d \in A$ gefunden mit $a_1 < a_2 < \ldots < a_d$. Dann ist aber $A = \{a_1, a_2, \ldots, a_d\}$ und a_d doch ein maximales Element, Widerspruch. $\square$

Definition 36.2. Eine Teilmenge $\varnothing \neq X \subseteq A$ heißt ***Kette in*** A, wenn $x \leqslant y$ oder $y \leqslant x$ für alle $x, y \in X$ gilt (d.h., $\leqslant$ ist eine totale Ordnung auf X). Wir sagen, dass A ***induktiv geordnet*** ist, wenn die folgende Bedingung erfüllt ist:

(Z) Ist $\varnothing \neq X \subseteq A$ eine Kette in A, so gibt es ein $a \in A$ mit $x \leqslant a$ für alle $x \in X$.

Oder kurz: Jede Kette in A besitzt eine ***obere Schranke*** in A.

Satz 36.3 (***Lemma von Zorn***, 1933)**.** *Ist A induktiv geordnet (und nicht-leer), so gibt es ein maximales Element in A.*

Beweis. Dies folgt aus dem Auswahlaxiom, das wir in Kapitel I, §6, formuliert haben. Die Herleitung ist rein mengentheoretischer Natur, und nicht ganz einfach; für die Details siehe §16 im Buch von Halmos. (Dies wäre auch geeignet für 1-2 Proseminarvorträge.) $\square$

Eine der klassischen Anwendungen des Lemmas von Zorn ist die allgemeine Existenz von Basen in Vektorräumen. Sei also V ein beliebiger K-Vektorraum, $V \neq \{0_V\}$. Eine Teilmenge $\varnothing \neq X \subseteq V$ heißt ***linear unabhängig***, wenn jedes Tupel $(v_1, \ldots, v_d)$ mit $d \geqslant 1$ und paarweise verschiedenen $v_1, \ldots, v_d \in X$ linear unabhängig ist (so wie in Kapitel IV, §17, definiert). Die leere Menge $\varnothing$ wird ebenfalls als linear unabhängig deklariert. Eine Teilmenge $B \subseteq V$ heißt ***Basis*** von V, wenn B linear unabhängig ist und $V = \langle B \rangle_K$ gilt. Wie in Kapitel IV, §17, sieht man dann sofort, dass sich jedes $v \in V$ auf eindeutige Weise schreiben lässt als $v = \sum_{i=1}^{d} s_i.v_i$ wobei $v_1, \ldots, v_d \in B$ paarweise verschieden sind und $s_i \in K$ für alle i gilt.

Man sieht sehr leicht, dass diese Definition mit der "vorläufigen" Definition einer Basis in Kapitel IV, Bemerkung 17.8, übereinstimmt. D.h., eine Teilmenge $B \subseteq V$ ist genau dann eine Basis von V, wenn $V = \langle B \rangle_K$ gilt und $V \neq \langle B' \rangle_K$ für jede echte Teilmenge $B' \subsetneqq B$.

Satz 36.4 (Existenz von Basen). *Sei V ein beliebiger Vektorraum, $V \neq \{0_V\}$. Dann gibt es eine Basis B von V. Genauer: Ist $S \subseteq V$ eine linear unabhängige Teilmenge, so gibt es eine Basis B von V mit $S \subseteq B$.*

Beweis. Sei $S \subseteq V$ linear unabhängig (hier ist $S = \varnothing$ erlaubt). Sei $\mathcal{A}$ die Menge aller linear unabhängigen Teilmengen $B \subseteq V$ mit $S \subseteq B$. Wegen $S \in \mathcal{A}$ ist $\mathcal{A} \neq \varnothing$. Durch $\subseteq$ wird eine Ordnungsrelation auf $\mathcal{A}$ definiert. Wir zeigen, dass damit $\mathcal{A}$ induktiv geordnet ist.
Sei $\varnothing \neq \mathcal{X} \subseteq \mathcal{A}$ eine Kette in $\mathcal{A}$, und bilde $B' := \bigcup_{B \in \mathcal{X}} B$. Dann gilt sicherlich $B \subseteq B'$ für alle $B \in \mathcal{X}$; also ist B' eine obere Schranke für $\mathcal{X}$. Wir müssen noch zeigen, dass B' linear unabhängig ist. Sei dazu $d \geqslant 1$ und seien $v_1, \ldots, v_d \in B'$ paarweise verschieden. Zu jedem i gibt es ein $B_i \in \mathcal{X}$ mit $v_i \in B_i$. Nach Lemma 36.1 hat die geordnete Menge $\{B_1, \ldots, B_d\} \subseteq \mathfrak{X}$ ein maximales Element. Wir wählen die Bezeichnungen so, dass dies B_d ist; es gilt also $v_i \in B_d$ für alle i. Wegen $B_d \in \mathcal{X}$ ist das Tupel $(v_1, \ldots, v_d)$ linear unabhängig, wie verlangt.
Nach dem Lemma von Zorn gibt es ein maximales Element in $\mathcal{A}$; sei $B \in \mathcal{A}$ ein solches maximales Element. Insbesondere gilt $S \subseteq B$ und B ist linear unabhängig; wir müssen noch zeigen, dass $V = \langle B \rangle_K$ gilt. Annahme, es wäre $\langle B \rangle_K \subsetneqq V$; sei dann $v \in V \setminus \langle B \rangle_K$. Wir bilden damit die Menge $B' := B \cup \{v\}$. Analog zu Kapitel IV, Lemma 17.10, zeigen wir nun, dass B' linear unabhängig ist. Seien also $d \geqslant 1$ und $v_1, \ldots, v_d \in B'$ paarweise verschieden mit $0 = \sum_{i=1}^{d} s_i.v_i$, wobei $s_i \in K$. Gilt $\{v_1, \ldots, v_d\} \subseteq B$, so ist $(v_1, \ldots, v_d)$ linear unabhängig, d.h., $s_i = 0$ für alle i. Sei nun $v_i = v$ für ein i; wir wählen die Bezeichnungen so, dass $i = d$ ist (und damit $v_1, \ldots, v_{d-1} \in B$). Ist $s_d = 0$, so erhalten wir auch $\sum_{i=1}^{d-1} s_i.v_i = 0$; weil $(v_1, \ldots, v_{d-1})$ linear unabhängig ist, folgt dann auch $s_i = 0$ für $i = 1, \ldots, d-1$. Ist andererseits $s_d \neq 0$, so folgt $s_d.v_d \in \langle v_1, \ldots, v_{d-1} \rangle_K$, also auch $v = v_d \in \langle v_1, \ldots, v_{d-1} \rangle_K \subseteq \langle B \rangle_K$, Widerspruch.

Also ist B' linear unabhängig, d.h., $B' \in \mathcal{A}$. Wegen $B \subsetneqq B'$ erhalten wir also einen Widerspruch dazu, dass B ein maximales Element von $\mathcal{T}$ ist. $\square$

Bemerkung 36.5. Sei B eine Basis von V; wir schreiben $B = \{b_i \mid i \in I\}$ mit einer Indexmenge I. (Die Abbildung $I \to B$, $i \mapsto b_i$, ist also eine Bijektion.) Ist $0_V \neq v \in V$ beliebig, so können wir v auf eindeutige Weise schreiben als $v = s_1.b_{i_1} + \ldots + s_d.b_{i_d}$, wobei $d \geqslant 1$, $s_1, \ldots, s_d \in K \setminus \{0\}$ und $i_1, \ldots, i_d \in I$ paarweise verschieden sind.
(a) Sei nun W ein weiterer K-Vektorraum. Für jedes $i \in I$ sei ein $w_i \in W$ gegeben. (Hier wird nicht verlangt, dass $w_i \neq w_j$ für $i \neq j$ gilt.) Analog zu Kapitel IV, Lemma 19.6(c), gibt es dann genau eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(b_i) = w_i$ für alle $i \in I$. (Ist $0_V \neq v \in V$ beliebig, so schreiben wir $v = s_1.b_{i_1} + \ldots + s_d.b_{i_d}$ wie oben; dann ist $\varphi(v) = s_1.w_{i_1} + \ldots + s_d.w_{i_d}$.)
(b) Für $i \in I$ definiere die lineare Abbildung $\gamma_i\colon V \to K$ durch $\gamma_i(b_i) := 1$ und $\gamma_i(b_{i'}) := 0$ für alle $i \neq i' \in I$. Sei nun $v \in V$ und $I_v := \{i \in I \mid \gamma_i(v) \neq 0\}$. Dann ist $|I_v| < \infty$ und es gilt $v = \sum_{i \in I_v} \gamma_i(v).b_i$. (Denn: Ist $v = 0_V$, so $I_v = \varnothing$; andernfalls schreiben wir wie oben $v = s_1.b_{i_1} + \ldots + s_d.b_{i_d}$; dann ist $I_v = \{i_1, \ldots, i_d\}$ und $s_1 = \gamma_{i_1}(v)$, ..., $s_d = \gamma_{i_d}(v)$.)

Folgerung 36.6. *Sei V ein beliebiger Vektorraum, $V \neq \{0_V\}$. Ist $0_V \neq v_0 \in V$ gegeben, so gibt es eine lineare Abbildung $\lambda\colon V \to K$ mit $\lambda(v_0) = 1$.*

Beweis. Nach Satz 36.4 gibt es eine Basis B von V mit $v_0 \in B$. Nach obiger Bemerkung gibt es eine lineare Abbildung $\lambda\colon V \to K$ mit $\lambda(v_0) = 1$ und $\lambda(v) = 0$ für alle $v_0 \neq v \in B$. □

Man kann zeigen, dass das Auswahlaxiom und das Lemma von Zorn äquivalent sind. Das Auswahlaxiom erscheint allerdings als unmittelbar einsichtiger ...

Ab hier Woche 12

37. *Faktorräume und direkte Summen*

In Kapitel II, §8, haben wir die Ringe $\mathbb{Z}/m\mathbb{Z}$ (für $m \in \mathbb{N}$) mit Hilfe einer Äquivalenzrelation auf $\mathbb{Z}$ eingeführt. Eine analoge Konstruktion ist auch für Vektorräume möglich.

Sei V ein K-Vektorraum und $U \subseteq V$ ein Teilraum. Für $v, w \in V$ definiere $v \sim w$ falls $v - w \in U$ gilt. Dies ist eine Äquivalenzrelation:
Reflexivität: $v - v = 0_V \in U$, also $v \sim v$. Symmetrie: Aus $v \sim w$ folgt $v - w \in U$, dann auch $w - v = -(v - w) \in U$, also $w \sim v$. Transitivität: Aus $u \sim v$ und $v \sim w$ folgt $u - v \in U$ und $v - w \in U$, also auch $u - w = (u - v) + (v - w) \in U$, d.h., $u \sim w$.
Die Äquivalenzklasse von $v \in V$ ist wie folgt gegeben:
$$\begin{aligned} K_v &= \{w \in V \mid v \sim w\} = \{w \in V \mid v - w \in U\} \\ &= \{w \in V \mid w = v + u \text{ für ein } u \in U\} =: v + U. \end{aligned}$$
Für $v, v' \in V$ folgt dann sofort: $\quad v + U = v' + U \;\Leftrightarrow\; v - v' \in U \;\Leftrightarrow v \sim v'$.

Für $v \in V$ heißt $\bar{v} := v + U$ die Nebenklasse von v nach U oder auch ***affiner Unterraum***.

Beispiel 37.1. Sei $A \in K^{m\times n}$ und $b \in K^m$. Sei $L := \{x \in K^n \mid A \cdot x = b\} \subseteq K^n$ die Lösungsmenge des inhomogenenen LGS mit erweiterter Matrix $[A|b]$. Sei $L \neq \varnothing$ und $x_0 \in L$ eine feste Lösung. Dann gilt $L = x_0 + U$ wobei $U = N(A) = \{x \in K^n \mid A \cdot x = 0_m\}$ der Lösungsraum des homogenen LGS mit Matrix A ist; siehe Kapitel IV, Beispiel 16.5.
Also ist die Lösungsmenge eines LGS (falls sie nicht-leer ist) stets ein affiner Unterraum.

Satz 37.2. *Sei $U \subseteq V$ ein Teilraum und $V/U = \{\overline{v} = v + U \mid v \in V\}$ die Menge der Äquivalenzklassen von $\sim$ (wie oben). Dann gilt:*
(a) V/U *ist ein* K*-Vektorraum mit* $\overline{v}+\overline{w} := \overline{v+w}$ *und* $s.\overline{v} := \overline{s.v}$ *für alle* $v, w \in V$ *und* $s \in K$. *Das neutrale Element ist* $\overline{0}_V = 0_V + U = U$*; das Inverse von* $\overline{v} \in V/U$ *ist* $-\overline{v} = \overline{-v} \in V/U$.
(b) *Die Abbildung* $\pi_U\colon V \to V/U$, $v \mapsto \overline{v}$, *ist linear und surjektiv, mit* $\mathrm{Kern}(\pi_U) = U$.

Der Vektorraum V/U heißt ***Faktorraum*** oder ***Quotientenraum*** von V nach U, und die Abbildung $\pi_U\colon V \to V/U$ der ***kanonische Homomorphismus***.

Beweis. (a) Wie im Fall der Ringe $\mathbb{Z}/m\mathbb{Z}$ müssen wir zeigen, dass die obigen Definitionen der Verknüpfungen auch tatsächlich "wohl-definiert" sind, d.h., konkret: Sind $v, v', w, w' \in V$ mit $\overline{v} = \overline{v'}$ und $\overline{w} = \overline{w'}$ gegeben, so gilt auch $\overline{v+w} = \overline{v'+w'}$ und $\overline{s.v} = \overline{s.v'}$ für alle $s \in K$. Dazu: Wegen $\overline{v} = \overline{v'}$ und $\overline{w} = \overline{w'}$ folgt $v - v' \in U$ und $w - w' \in U$, also $(v+w)-(v'+w') = (v-v') + (w-w') \in U$ und damit $\overline{v+w} = \overline{v'+w'}$. Völlig analog sieht man $\overline{s.v} = \overline{s.v'}$.
Die Vektorraumaxiome für V/U ergeben sich dann unmittelbar aus den Vektorraumaxiomen für V (analog zum Beweis von Sazt 8.1 in Kapitel II).

(b) Für $v, w \in V$ gilt $\pi_U(v + w) = \overline{v+w} = \overline{v} + \overline{w} = \pi_U(v) + \pi_U(w)$, sowie $\pi_U(s.v) = \overline{s.v} = s.\overline{v} = s.\pi_U(v)$ für $s \in S$. Also ist π_U linear; Surjektivität ist klar. Schließlich gilt $\mathrm{Kern}(\pi_U) = \{v \in V \mid \overline{v} = \overline{0}_V\} = \{v \in V \mid v \sim 0_V\} = \{v \in V \mid v \in U\} = U$. □

Satz 37.3. *Sei* $\dim V < \infty$. *Dann gilt* $\dim(V/U) = \dim V - \dim U$.

Beweis. Mit der Kern-Bild-Dimensionsformel folgt $\dim V = \dim \mathrm{Kern}(\pi_U) + \dim \mathrm{Bild}(\pi_U)$. Nach obigem Satz gilt $\mathrm{Kern}(\pi_U) = U$ und $\mathrm{Bild}(\pi_U) = V/U$. Aus dem Beweis dieses Satzes erhalten wir auch die folgende Aussage zu Basen. Sei $1 \leqslant n := \dim V < \infty$ und $1 \leqslant d := \dim U \leqslant n$. Sei $\{v_1, \ldots, v_d\}$ eine Basis von U; wir ergänzen diese zu einer Basis $\{v_1, \ldots, v_d, v_{d+1}, \ldots, v_n\}$ zu einer Basis von V. Dann ist $\{\overline{v}_{d+1}, \ldots, \overline{v}_n\}$ eine Basis von V/U. □

Satz 37.4 (***Homomorphiesatz***)**.** *Seien* V, W *Vektorräume über* K *und* $\varphi\colon V \to W$ *eine lineare Abbildung. Sei* $U \subseteq V$ *ein Teilraum mit* $U \subseteq \mathrm{Kern}(\varphi)$. *Dann gibt es eine eindeutige lineare Abbildung* $\overline{\varphi}\colon V/U \to W$ *mit* $\varphi = \overline{\varphi} \circ \pi_U$. *Es gilt* $\mathrm{Bild}(\overline{\varphi}) = \mathrm{Bild}(\varphi)$.
Ist $U = \mathrm{Kern}(\varphi)$, *so ist* $\overline{\varphi}$ *injektiv und* $\overline{\varphi}\colon V/\mathrm{Kern}(\varphi) \to \mathrm{Bild}(\varphi)$ *ist ein Isomorphismus; in Zeichen* $V/\mathrm{Kern}(\varphi) \cong \mathrm{Bild}(\varphi)$.

Beweis. Für $v \in V$ setze $\overline{\varphi}(\overline{v}) := \varphi(v)$. Wiederum müssen wir zuerst zeigen, dass dies "wohldefiniert" ist, d.h., sind $v, v' \in V$ gegeben mit $\overline{v} = \overline{v}'$, so muss $\varphi(v) = \varphi(v')$ gelten. Dazu: Aus $\overline{v} = \overline{v}'$ folgt $v - v' \in U \subseteq \text{Kern}(\varphi)$, also $\varphi(v) - \varphi(v') = \varphi(v - v') = 0_V$ und damit $\varphi(v) = \varphi(v')$. Dann folgt auch sofort, dass $\overline{\varphi}$ linear ist. Es gilt $\text{Bild}(\overline{\varphi}) = \{\overline{\varphi}(\overline{v}) \mid v \in V\} = \{\varphi(v) \mid v \in V\} = \text{Bild}(\varphi)$. Nach Konstruktion ist $(\overline{\varphi} \circ \pi_U)(v) = \overline{\varphi}(\overline{v}) = \varphi(v)$ für alle $v \in V$, also gilt $\overline{\varphi} \circ \pi_U = \varphi$. Sei auch auch $\psi\colon V/U \to W$ linear mit $\psi \circ \pi_U = \varphi$. Dann gilt $\psi(\overline{v}) = \psi(\pi_U(w)) = (\psi \circ \pi_U)(v) = \varphi(v) = \overline{\varphi}(\overline{v})$ für alle $v \in V$, also $\psi = \overline{\varphi}$.
Sei schließlich $U = \text{Kern}(\varphi)$. Sei $\overline{v} \in \text{Kern}(\overline{\varphi})$, also $\varphi(v) = \overline{\varphi}(\overline{v}) = 0_W$, d.h, $v \in \text{Kern}(\varphi) = U$ und damit $\overline{v} = \overline{0}_V$. Also ist $\overline{\varphi}$ injektiv. □

Beispiel 37.5. Wie in Kapitel II, §9, sei $\mathcal{F} := \text{Abb}(\mathbb{N}_0, \mathbb{Q})$ die Menge aller Folgen $f = (a_n)_{n \geqslant 0}$ mit $a_n \in \mathbb{Q}$ für alle $n \in \mathbb{N}_0$. Mit den dort definierten Verknüpfungen ist $\mathcal{F}$ ein $\mathbb{Q}$-Vektorraum, mit $\dim \mathcal{F} = \infty$. Sei $\mathcal{F}_0$ die Menge aller $f = (a_n)_{n \geqslant 0} \in \mathcal{F}$, die eine Cauchy–Folge sind (siehe Analysis); dann sieht man leicht, dass $\mathcal{F}_0$ ein Teilraum von $\mathcal{F}$ ist. Jede Cauchy-Folge hat einen Grenzwert in $\mathbb{R}$; wir erhalten damit eine lineare Abbilding $\varphi\colon \mathcal{F}_0 \to \mathbb{R}$, $(a_n)_{n \geqslant 0} \mapsto \lim\limits_{n \to \infty}(a_n)$ (wobei wir $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum auffassen). Dann ist φ surjektiv (denn jede reelle Zahl ist Grenzwert einer Cauchy-Folge) und $\text{Kern}(\varphi)$ besteht genau aus den Null-Folgen in $\mathcal{F}_0$. Mit dem Homomorphiesatz folgt $\mathcal{F}_0/\text{Kern}(\varphi) \cong \mathbb{R}$ (Isomorphismus als $\mathbb{Q}$-Vektorräume).

Beispiel 37.6. Sei $\varphi\colon V \to V$ linear und $U \subseteq V$ ein ***invarianter Teilraum***, d.h., es gilt $\varphi(U) \subseteq U$. Betrachte die lineare Abbildung $\varphi' := \pi_U \circ \varphi\colon V \to V/U$. Sei $u \in U$. Dann ist nach Voraussetzung $\varphi(u) \in U$, also $\varphi'(u) = \pi_U(\varphi(u)) = 0_{V/U}$, d.h., es gilt $U \subseteq \text{Kern}(\varphi')$. Nach Satz 37.4 gibt es genau eine lineare Abbildung $\tilde{\varphi}\colon V/U \to V/U$ mit $\pi_U \circ \varphi = \varphi' = \tilde{\varphi} \circ \pi_U$, d.h., $\tilde{\varphi}(\overline{v}) = \tilde{\varphi}(\pi_U(v)) = \pi_U(\varphi(v)) = \overline{\varphi(v)}$ für alle $v \in V$. Sei nun $1 \leqslant n := \dim V < \infty$ und $1 \leqslant d := \dim U < n$. Sei $C = \{v_1, \ldots, v_d\}$ eine Basis von U; wir ergänzen diese zu einer Basis $B = \{v_1, \ldots, v_d, v_{d+1}, \ldots, v_n\}$ von V. Wegen $\varphi(U) \subseteq U$ hat dann $A = M_B(\varphi) \in M_n(K)$ eine Blockdiagonalgestalt der Form

$$A = M_B(\varphi) = \left[\begin{array}{c|c} A' & * \\ \hline 0 & A'' \end{array}\right] \quad \text{mit} \quad A' \in M_d(K) \text{ und } A'' \in M_{n-d}(K),$$

wobei A' die Matrix von $\varphi|_U \in \text{End}(U)$ bezüglich C ist und A'' die Matrix von $\tilde{\varphi} \in \text{End}(V/U)$ bezüglich der Basis $\{\overline{v}_{d+1}, \ldots, \overline{v}_n\}$ von V/U. (Wir haben dies implizit bereits im Beweis des Lemmas von Jacob in §28 benutzt.) Dazu beachte: Für $j > d$ gilt

$$\tilde{\varphi}(\overline{v}_j) = \big(\tilde{\varphi} \circ \pi_U\big)(v_j) = \varphi'(v_j) = \big(\pi_U \circ \varphi\big)(v_j) = \pi_U\big(\textstyle\sum_{i=1}^n a_{ij} v_i\big) = \sum_{i=1}^n a_{ij} \overline{v}_i$$

und in der letzten Summe ist $\overline{v}_i = \overline{0}_V$ für $1 \leqslant i \leqslant d$; die Summation braucht also nur über $i = d+1, \ldots, n$ zu laufen.

Wir kommen nun zu direkten Summen von Vektorräumen. Sei $d \geqslant 2$ und seien $V_1, \ldots, V_d$ Vektorräume über K. Wir betrachten das kartesische Produkt

$$V_1 \times \ldots \times V_d = \{(v_1, \ldots, v_d) \mid v_i \in V_i \text{ für alle } i\};$$

dieses ist wieder ein K-Vektorraum mit den Verknüpfungen

$$(v_1, \ldots, v_d) + (v_1', \ldots, v_d') := (v_1 + v_1', \ldots, v_d + v_d') \text{ und } s.(v_1, \ldots, v_d) := (s.v_1, \ldots, s.v_d),$$

wobei $v_i, v_i' \in V_i$ und $s \in K$. (Den Fall $d = 2$ haben wir bereits in Kapitel IV, Bemerkung 19.9, betrachtet.) Man bezeichnet $V_1 \times \ldots \times V_d$ auch als ***äußere direkte Summe*** von $V_1, \ldots, V_d$. Gilt $\dim V_i < \infty$ für alle i, so ist (siehe wieder Kapitel IV, Bemerkung 19.9, für $d = 2$):

$$\dim(V_1 \times \ldots \times V_d) = \dim V_1 + \ldots + \dim V_d.$$

Definition 37.7. Sei V ein K-Vektorraum. Gegeben seien Unterräume $U_1, \ldots, U_d \subseteq V$ (wobei $d \geqslant 2$). Wir bilden $U_1 \times \ldots \times U_d$ wie oben und betrachten die Abbildung

$$\varphi\colon U_1 \times \ldots \times U_d \to V, \qquad (u_1, \ldots, u_d) \mapsto u_1 + \ldots + u_d.$$

Man sieht sofort, dass φ linear ist, mit $\mathrm{Bild}(\varphi) = U_1 + \ldots + U_d$. Wir bezeichnen V als die ***direkte Summe*** (oder genauer auch ***innere direkte Summe***) von $U_1, \ldots, U_d$, wenn φ bijektiv ist, d.h., wenn $V = U_1 + \ldots + U_d$ gilt und sich jedes $v \in V$ auf eindeutige Weise schreiben lässt als $v = u_1 + \ldots + u_d$ mit $u_i \in U_i$ für $= 1, \ldots, d$.
In diesem Fall schreiben wir $V = U_1 \oplus \ldots \oplus U_d$.

Lemma 37.8. *Seien* $U_1, \ldots, U_d \subseteq V$ *Teilräume (mit* $d \geqslant 2$*). Sei* $\dim V < \infty$. *Genau dann ist* $V = U_1 \oplus \ldots \oplus U_d$*, wenn* $V = U_1 + \ldots + U_d$ *und* $\dim V = \dim U_1 + \ldots + \dim U_d$ *gilt.*

Beweis. Sei zuerst $V = U_1 \oplus \ldots \oplus U_d$, d.h., φ ist bijektiv und damit $\dim(U_1 \times \ldots \times U_d) = \dim V$ (siehe Kapitel IV, Beispiel 19.8). Die linke Seite ist auch gleich $\dim U_1 + \ldots + \dim U_d$ (siehe oben). Nun gelte umgekehrt $V = U_1 + \ldots + U_d$ und $\dim V = \dim U_1 + \ldots + \dim U_d$. Dann ist φ surjektiv, also $\mathrm{Bild}(\varphi) = V$. Mit der Kern-Bild-Dimensionsformel folgt $\dim \mathrm{Kern}(\varphi) = \dim(U_1 \times \ldots \times U_d) - \dim \mathrm{Bild}(\varphi) = \dim U_1 + \ldots + \dim U_d - \dim V = 0$, also ist φ injektiv. $\square$

Beispiel 37.9. Sei $A \in M_n(K)$ diagonalisierbar. Seien $\lambda_1, \ldots, \lambda_r \in K$ die verschiedenen Eigenwerte von A. Für jedes i sei $E_A(\lambda_i) \subseteq K^n$ der zugehörige Eigenraum. Nach Bemerkung 26.10 ist dann $K^n = E_A(\lambda_1) + \ldots + E_A(\lambda_r)$ und $n = \dim E_A(\lambda_i) + \ldots + \dim E_A(\lambda_i)$. Also folgt mit Lemma 37.8: $\quad K^n = E_A(\lambda_1) \oplus \ldots \oplus E_A(\lambda_r) \quad$ (direkte Summe der Eigenräume).

Beispiel 37.10. Seien $U_1, U_2 \subseteq V$ Teilräume mit $V = U_1 \oplus U_2$. Wir definieren wie folgt eine Abbildung $\pi\colon V \to U_2$. Sei $v \in V$; dann schreibe $v = u_1 + u_2$ mit eindeutigen $u_1 \in U_1$, $u_2 \in U_2$, und setze $\pi(v) := u_2$. Man sieht sofort, dass π linear und surjektiv ist, mit $\mathrm{Kern}(\pi) = U_1$. Nach dem Homomorphiesatz erhalten wir also einen Isomorphismus $\overline{\pi}\colon V/U_1 \to U_2$.

Bemerkung 37.11. Für $d = 2$ gibt es noch eine weitere Charakterisierung von direkten Summen. Seien also $U_1, U_2 \subseteq V$ Teilräume und betrachte wieder $\varphi\colon U_1 \times U_2 \to V$. Nun ist

$$\mathrm{Kern}(\varphi) = \{(u_1, u_2) \in U_1 \times U_2 \mid u_1 = -u_2\} = \{(u, -u) \mid u \in U_1 \cap U_2\}.$$

Also ist φ injektiv genau dann, wenn $U_1 \cap U_2 = \{0_V\}$ gilt. Weiterhin gilt natürlich wie oben, dass φ surjektiv ist genau dann, wenn $V = U_1 + U_2$. Damit folgt:

$$V = U_1 \oplus U_2 \quad \Leftrightarrow \quad V = U_1 + U_2 \text{ und } U_1 \cap U_2 = \{0_V\}.$$

Beispiel 37.12. Sei V der $\mathbb{R}$-Vektorraum aller Funktionen $f\colon \mathbb{R} \to \mathbb{R}$. Wir sagen, dass $f \in V$ gerade ist, wenn $f(-x) = f(x)$ für alle $x \in \mathbb{R}$ gilt; wir sagen, dass $f \in V$ ungerade ist, wenn $f(-x) = -f(x)$ für alle $x \in \mathbb{R}$ gilt. Man sieht leicht, dass die Teilmenge $V^+ \subseteq V$ aller geraden Funktionen ein Teilraum ist; analog ist die Teilmenge $V^- \subseteq V$ aller ungeraden Funktionen ein Teilraum. Wir behaupten, dass $V = V^+ \oplus V^-$ gilt.
Dazu: Sei $f \in V^+ \cap V^-$. Für $x \in \mathbb{R}$ gilt dann $f(x) = f(-x) = -f(x)$, also $2f(x) = 0$ und damit $f(x) = 0$. Also ist $V^+ \cap V^- = \{\underline{0}\}$. Sei nun $f \in V$ beliebig. Wir definieren $f^\pm \in V$ durch $f^+(x) := \frac{1}{2}(f(x) + f(-x))$ und $f^-(x) := \frac{1}{2}(f(x) - f(-x))$ für alle $x \in \mathbb{R}$. Man sieht sofort, dass $f^+ \in V^+$ und $f^- \in V^-$ gilt. Außerdem ist $f(x) = f^+(x) + f^-(x)$ für alle $x \in \mathbb{R}$ und damit $f \in V^+ + V^-$, d.h., es ist $V = V^+ + V^-$. Also folgt die Behauptung mit Bemerkung 37.11.

Bemerkung 37.13. Seien $U_1, U_2 \subseteq K^n$ beliebige Teilräume sowie $A_1 \in K^{n\times n_1}$ und $A_2 \in K^{n\times n_2}$ mit $n_i \geqslant 1$ und $U_i = \mathrm{SR}(A_i)$ für $i = 1, 2$. Mit dem folgenden ***Zassenhaus–Algorithmus*** kann man Basen für $U_1 \cap U_2$ und $U_1 + U_2$ berechnen.
Dazu bilde die folgenden Blockmatrizen der Größe $(n_1 + n_2) \times 2n$:

$$\left[\begin{array}{c|c} A_1^{\mathrm{tr}} & A_1^{\mathrm{tr}} \\ \hline A_2^{\mathrm{tr}} & 0 \end{array}\right] \quad \xrightarrow{\text{Gauß-Verfahren}} \quad \left[\begin{array}{c|c} C^{\mathrm{tr}} & * \\ \hline 0 & D^{\mathrm{tr}} \\ \hline 0 & 0 \end{array}\right],$$

wobei die rechte Seite in Stufenform ist und die Aufteilung in Blöcke so erfolgt, dass C^{tr} und D^{tr} jeweils n Spalten haben sowie alle Zeilen von C^{tr} und D^{tr} ungleich der Null-Zeile sind. Dann ist $U_1 + U_2 = \mathrm{SR}(C)$ und $U_1 \cap U_2 = \mathrm{SR}(D)$.
[Begründung: Sei $U_i' := \mathrm{ZR}(A_i^{\mathrm{tr}}) \subseteq K^{1\times n}$ für $i = 1, 2$. Wie im Beweis von Kapitel IV, Satz 19.10, sei $W' := \{u_1 + u_2, u_1) \mid u_1 \in U_1', u_2 \in U_2'\} \subseteq K^{1\times n} \times K^{1\times n}$; dort haben wir dann gesehen, dass $\pi_1(W') = U_1' + U_2'$ und $\mathrm{Kern}(\pi_1) \cap W' = \{(0_n, u) \mid u \in U_1' \cap U_2'\}$ gilt, wobei $\pi_1\colon U_1' \times U_2' \to U_1'$ die Projektionsabbildung auf die erste Komponente ist. Wir fassen nun die Paare von Zeilenvektoren in $K^{1\times n} \times K^{1\times n}$ als Zeilenvektoren in $K^{1\times 2n}$ auf; damit ist also $W' \subseteq K^{1\times 2n}$. Da W' von den Zeilenvektoren (u_1, u_1) (mit $u_1 \in U_1'$) und $(u_2, 0_V)$ (mit $u_2 \in U_2'$) erzeugt wird, ist W' der Zeilenraum der linken oberen Matrix. Nach Kapitel IV, Beispiel 17.9, ändert sich der Zeilenraum einer Matrix unter elementaren Zeilenumformungen nicht. Also ist W' auch der Zeilenraum der rechten oberen Matrix. Schauen wir jeweils nur auf die ersten n Spalten der obigen Matrizen, so ist klar, dass $\mathrm{ZR}(C^{\mathrm{tr}})$ der von den Zeilen von A_1^{tr} und A_2^{tr} erzeugte Teilraum ist, also $\mathrm{ZR}(C^{\mathrm{tr}}) = U_1 + U_2$. Andererseits besteht $\mathrm{Kern}(\pi_1) \cap W'$ genau aus denjenigen Zeilenvektoren in W', bei denen alle Komponenten in den ersten n Spalten gleich 0 sind, also folgt $\mathrm{Kern}(\pi_1') \cap W' = \{(0_V, u) \mid u \in \mathrm{ZR}(D^{\mathrm{tr}}\}$.]

Gegegen seien zum Beispiel $U_1 = \langle u_1, u_2\rangle_{\mathbb{R}} \subseteq \mathbb{R}^4$ und $U_2 = \langle u_1', u_2'\rangle_{\mathbb{R}} \subseteq \mathbb{R}^4$ mit

$$u_1 := \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \qquad u_2 := \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \qquad u_1' := \begin{bmatrix} 5 \\ 0 \\ -3 \\ 3 \end{bmatrix}, \qquad u_2' := \begin{bmatrix} 0 \\ 5 \\ -3 \\ -2 \end{bmatrix}.$$

Dann bilden wir

$$\left[\begin{array}{rrrr|rrrr} 1 & -1 & 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ \hline 5 & 0 & -3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & -3 & -2 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{Gauß-Verfahren}} \left[\begin{array}{rrrr|rrrr} 1 & 0 & 0 & 0 & * & * & * & * \\ 0 & 1 & 0 & -1 & * & * & * & * \\ 0 & 0 & 1 & -1 & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right].$$

Also folgt: $$U_1 + U_2 = \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle \quad \text{und} \quad U_1 \cap U_2 = \left\langle \begin{bmatrix} 1 \\ -1 \\ 0 \\ -1 \end{bmatrix} \right\rangle.$$

38. *Dualraum und transponierte Abbildungen*

Sind V, W Vektorräume über K, so ist $\mathrm{Hom}(V, W) = \{\varphi\colon V \to W \mid \varphi \text{ linear}\}$ wieder ein K-Vektorraum; siehe Kapitel IV, §20. Wir betrachten nun den Spezialfall $W = K$.

Definition 38.1. Der K-Vektorraum $V^* := \mathrm{Hom}(V, K)$ heißt der ***Dualraum*** von V. Die Elemente von V^* sind also lineare Abbildungen $\lambda\colon V \to K$; auch genannt ***Linearformen***.
Beachte: Ist $\dim V < \infty$, so gilt $\dim V^* = \dim \mathrm{Hom}(V, K) = (\dim V)(\dim K) = \dim V$ (siehe Kapitel IV, Satz 20.4).

Beispiel 38.2. (a) Sei $V = K^n$ und seien $a_1, \dots, a_n \in K$ fest. Definiere $\lambda\colon V \to K$ durch

$$\lambda\Big(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\Big) := a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad \text{für alle } x_i \in K;$$

dann ist $\lambda \in V^*$. Nach Kapitel IV, Lemma 19.4, ist umgekehrt jede Linearform in V^* von dieser Form (also gegeben durch einen Zeilenvektor in $K^{1\times n}$).

(b) Sei $V := \{f\colon [a, b] \to \mathbb{R} \mid f \text{ stetig}\}$. Dann ist V ein $\mathbb{R}$-Vektorraum und die Abbildung $\lambda\colon V \to \mathbb{R}$, $f \mapsto \int_a^b f(x)dx$, eine Linearform (nach Sätzen der Analysis).

(c) Sei $X \neq \varnothing$ eine Menge und $V := \mathrm{Abb}(X, K)$. Sei $x_0 \in X$ fest. Dann ist $\lambda_{x_0}\colon V \to K$, $f \mapsto f(x_0)$, eine Linearform ("Auswertung an x_0"); wie bereits in Kapitel IV, Beispiel 19.2, bemerkt. Seien nun $x_1, \dots, x_n \in X$ paarweise verschieden und $\lambda_i := \lambda_{x_i} \in V^*$ für $i = 1, \dots, n$.
Behauptung: Das Tupel $(\lambda_1, \dots, \lambda_n)$ ist linear unabhängig in V^*.
Dazu: Seien $s_1, \dots, s_n \in K$ so, dass $\sum_{i=1}^n s_i.\lambda_i = \underline{0}$ gilt. Für $1 \leqslant j \leqslant n$ definiere $f_j \in V$ durch $f_j(x_j) := 1$ und $f_j(x) := 0$ für alle $x \neq x_j$. Dann ist $f_j(x_j) = 1$ und $f_j(x_i) = 0$ für alle $i \neq j$.
Damit folgt $0 = \big(\sum_{i=1}^n s_i.\lambda_i\big)(f_j) = \sum_{i=1}^n s_i f_j(x_i) = s_j$ für alle j.

Beispiel 38.3. Sei $\beta\colon V \times V \to K$ eine reflexive Bilinearform. Für festes $w \in V$ erhalten wir dann eine Abbildung $\beta_w\colon V \to K$, $v \mapsto \beta(v, w)$. Wegen der Bilinearität von β ist $\beta_w \in V^*$ und die Abbildung $\widehat{\beta}\colon V \to V^*$, $w \mapsto \beta_w$, linear. Es gilt

$$\mathrm{Kern}(\widehat{\beta}) = \{w \in V \mid \beta_w = \underline{0}\} = \{w \in V \mid \beta(v, w) = \beta_w(v) = 0 \text{ für alle } v \in V\} = V^{\perp}.$$

Also ist $\widehat{\beta}$ genau dann injektiv, wenn β nicht-ausgeartet ist.

Satz 38.4. *Sei $n = \dim V < \infty$ und $B = \{v_1, \ldots, v_n\}$ eine Basis von V. Zu jedem v_i gibt es genau ein $\lambda_i \in V^*$ mit $\lambda_i(v_i) = 1$ und $\lambda_i(v_j) = 0$ für alle $j \neq i$. Dann ist $B^* := \{\lambda_1, \ldots, \lambda_n\}$ eine Basis von V^* und heißt die zu B* ***duale Basis*** *von V^*; wir schreiben dann auch $\lambda_i = v_i^*$ für alle i. Für $v \in V$ beliebig gilt dann $v = \sum_{i=1}^n \lambda_i(v)v_i$, und für $\lambda \in V^*$ gilt $\lambda = \sum_{i=1}^n \lambda(v_i)\lambda_i$.*

Beweis. Existenz und Eindeutigkeit der λ_i ist klar nach allgemeinen Sätzen zu linearen Abbildungen; siehe Kapitel IV, Lemma 19.6. Wir wissen bereits, dass $\dim V^* = \dim V$ gilt. Wir zeigen jetzt, dass $V^* = \langle \lambda_1, \ldots, \lambda_n \rangle_K$ gilt. Dazu sei $\lambda \in V^*$ beliebig. Dann ist

$$\left(\sum_{i=1}^n \lambda(v_i)\lambda_i\right)(v_j) = \sum_{i=1}^n \lambda(v_i)\lambda_i(v_j) = \lambda(v_j) \qquad \text{für } j = 1, \ldots, n.$$

Also gilt $\lambda = \sum_{i=1}^n \lambda(v_i)\lambda_i \in \langle \lambda_1, \ldots, \lambda_n \rangle_K$; außerdem ist damit die Formel für die Zerlegung von λ gezeigt. Schließlich sei $v \in V$ beliebig und $v = \sum_{i=1}^n s_i.v_i$ mit $s_i \in K$. Dann gilt $\lambda_j(v) = \sum_{i=1}^n s_i\lambda_j(v_i) = s_j$ für alle j, also gilt auch die Formel für die Zerlegung von v. □

Bemerkung 38.5. Die Aussage des obigen Satzes gilt nicht wenn $\dim V = \infty$. Sei zum Beispiel $V = K[X]$ der Polynomring über K in der Unbestimmten X; dann ist V ein K-Vektorraum mit Basis $B = \{X^i \mid i \in \mathbb{N}_0\}$. Für $i \in \mathbb{N}_0$ definiere wie oben $\lambda_i \in V^*$ durch $\lambda_i(X^i) = 1$ und $\lambda_i(X_j) = 0$ für alle $j \neq i$. Angenommen, es wäre $V^* = \langle \lambda_i \mid i \in \mathbb{N}_0 \rangle_K$. Sei dann $\lambda \in V^*$ die Auswertung an 1, also $\lambda(f) = f(1)$ für $f \in K[X]$. Wegen $V^* = \langle \lambda_i \mid i \in \mathbb{N}_0 \rangle_K$ gibt es dann ein $n \geqslant 0$ und $s_0, \ldots, s_n \in K$ mit $\lambda = \sum_{i=0}^n s_i\lambda_i$. Aber andererseits gilt $1 = \lambda(X^{n+1}) = \sum_{i=0}^n s_i\lambda_i(X^{n+1}) = 0$, Widerspruch.

Ab hier Woche 13

Satz 38.6. *Sei $V^{**} := (V^*)^*$ der* ***Bidualraum*** *von V.*

(a) *Für jedes $v \in V$ ist $\varepsilon_v\colon V^* \to K$, $\lambda \mapsto \lambda(v)$, eine lineare Abbildung, also $\varepsilon_v \in V^{**}$.*

(b) *Die Abbildung $\varepsilon\colon V \to V^{**}$, $v \mapsto \varepsilon_v$, ist linear und injektiv.*

(c) *Ist $\dim V < \infty$, so ist $\varepsilon\colon V \to V^{**}$ ein Isomorphismus.*

Beweis. (a) Dies ist klar weil ε_v die Auswertung an v ist (siehe Beispiel 38.2(c)).

(b) Für $v, w \in V$ und $\lambda \in V^*$ gilt $\varepsilon(v + w)(\lambda) = \varepsilon_{v+w}(\lambda) = \lambda(v + w) = \lambda(v) + \lambda(w) = \varepsilon_v(\lambda) + \varepsilon_w(\lambda) = (\varepsilon_v + \varepsilon_w)(\lambda) = (\varepsilon(v) + \varepsilon(w))(\lambda)$; also ist $\varepsilon(v + w) = \varepsilon(v) + \varepsilon(w)$. Genauso sieht man $\varepsilon(s.v)(\lambda) = (s.\varepsilon(v))(\lambda)$, d.h., $\varepsilon(s.v) = s.\varepsilon(v)$. Also ist ε linear. Nehmen wir an, es sei $\mathrm{Kern}(\varepsilon) \neq \{0_V\}$; sei dann $0_V \neq v_0 \in \mathrm{Kern}(\varepsilon)$ und $\lambda \in V^*$ wie in Folgerung 36.6(b), also $\lambda(v_0) = 1$. Wegen $v_0 \in \mathrm{Kern}(\varepsilon)$ gilt $\varepsilon_{v_0} = \varepsilon(v_0) = \underline{0}$, d.h., $\lambda(v_0) = \varepsilon_{v_0}(\lambda) = 0$, Widerspruch.

(c) Wegen $\dim V < \infty$ ist $\dim V^{**} = \dim V^* = \dim V < \infty$. Da $\varepsilon\colon V \to V^{**}$ injektiv, muss φ bijektiv sein (siehe Kapitel IV, Beispiel 19.8). □

Definition 38.7. Sei $M \subseteq V$ eine Teilmenge. Der ***Annulator*** von M ist definiert als

$$M^\circ := \{\lambda \in V^* \mid \lambda(m) = 0 \text{ für alle } m \in M\}.$$

Man sieht sofort, dass M° stets ein Teilraum von V^* ist.

Satz 38.8. (a) *Sei* $M \subseteq V$ *eine Teilmenge und* $U := \langle M \rangle_K$. *Dann gilt* $M^\circ = U^\circ$.
(b) *Ist* $\dim V < \infty$ *und* $U \subseteq V$ *ein Teilraum, so gilt* $\dim U^\circ = \dim V - \dim U$.

Beweis. (a) Wegen $M \subseteq U$ ist es klar, dass $U^\circ \subseteq M^\circ$ gilt. Sei umgekehrt $\lambda \in M^\circ$, also $\lambda(m) = 0$ für alle $m \in M$. Weil λ linear ist, gilt dann auch $\lambda(m') = 0$ für jede Linearkombination $m' \in U$ von Elementen von M. Also ist auch $m \in U^\circ$.
(b) Betrachte die Abbildung $f\colon V^* \to U^*$, $\lambda \mapsto \lambda|_U$ (Einschränkung von λ auf U). Dann ist f linear und $\mathrm{Kern}(f) = \{\lambda \in V^* \mid \lambda(u) = 0 \text{ für alle } u \in U\} = U^\circ$. Behauptung: f ist surjektiv. Dazu sei $\mu \in U^*$ beliebig. Sei $n := \dim V < \infty$ und $d := \dim U < \infty$. Sei $\{v_1, \dots, v_d\}$ eine Basis von U; diese können wir zu einer Basis $\{v_1, \dots, v_d, v_{d+1}, \dots, v_n\}$ von V ergänzen. Nun definiere $\hat{\mu} \in V^*$ durch $\hat{\mu}(v_i) := \mu(v_i)$ für $1 \leqslant i \leqslant d$, und $\hat{\mu}(v_i) := 0$ für $i > d$. Dann ist $\mu = \hat{\mu}|_U = f(\hat{\mu}) \in \mathrm{Bild}(f)$, wie gewünscht. Mit der Kern-Bild-Dimensionsformel folgt schließlich $\dim V - \dim U^\circ = \dim V^* - \dim U^\circ = \mathrm{Bild}(f) = \dim U^* = \dim U$. □

Sei $U \subseteq K^n$ ein beliebiger Teilraum, $d := \dim U$. In Kapitel IV, Beispiel 19.12, haben wir gesehen, dass es eine Matrix $A \in K^{(n-d)\times n}$ gibt mit $U = N(A) = \{x \in K^n \mid A \cdot x = 0_{n-d}\}$. Der folgende Satz erklärt nun etwas besser, wie A zustande kommt.

Folgerung 38.9. *Sei* $\dim V < \infty$ *und* $U \subseteq V$ *ein Teilraum. Sei* $r = \dim U^\circ$ *und* $\{\lambda_1, \dots, \lambda_r\}$ *eine Basis von* U°. *Dann gilt* $\dim U = \dim V - r$ *und* $U = \{v \in V \mid \lambda_i(v) = 0 \text{ für } i = 1, \dots, r\}$.

Beweis. Für $u \in U$ ist $\lambda_i(u) = 0$ für $i = 1, \dots, r$. Also ist $U \subseteq U' := \{v \in V \mid \lambda_i(v) = 0 \text{ für } i = 1, \dots, r\}$. Mit Satz 38.8(b) folgt $\dim V - r = \dim U \leqslant \dim U'$. Also genügt es, $\dim U' \leqslant \dim V - r$ zu zeigen. Dazu setze $Y := U^\circ = \langle \lambda_1, \dots, \lambda_r \rangle_K \subseteq V^*$ und betrachte die Abbildung $\varepsilon\colon V \to V^{**}$, $v \mapsto \varepsilon_v$, in Satz 38.6. Dann gilt $\varepsilon(U') \subseteq Y^\circ \subseteq V^{**}$.
Denn sei $v \in U'$ beliebig. Für $1 \leqslant i \leqslant r$ gilt dann $\varepsilon(v)(\lambda_i) = \varepsilon_v(\lambda_i) = \lambda_i(v) = 0$. Weil $\{\lambda_1, \dots, \lambda_r\}$ eine Basis von Y ist, folgt $\varepsilon(v)(y) = 0$ für alle $y \in Y$, also $\varepsilon(v) \in Y^\circ$.

Weil ε bijektiv ist, folgt schließlich $\dim U' = \dim \varepsilon(U') \leqslant \dim Y^\circ$. Und mit Satz 38.8(b) dann auch $\dim U' \leqslant \dim Y^\circ = \dim V^* - \dim Y = \dim V^* - \dim U^\circ = \dim V - r$. □

Definition 38.10. Seien V, W Vektorräume über K und $\varphi\colon V \to W$ eine lineare Abbildung. Für $\mu \in W^*$ ist dann $\mu \circ \varphi\colon V \to K$ auch linear, also erhalten wir eine Abbildung

$$\varphi^*\colon W^* \to V^*, \qquad \mu \mapsto \mu \circ \varphi.$$

Dann heißt φ^* die zu φ ***transponierte Abbildung*** (oder auch ***duale Abbildung***).
Man sieht leicht, dass φ^* linear ist. Denn für $\mu, \mu' \in W^*$ und $v \in V$ folgt:

$$\begin{aligned}\big(\varphi^*(\mu+\mu')\big)(v) &= \big((\mu+\mu')\circ\varphi\big)(v) = (\mu+\mu')(\varphi(v)) = \mu(\varphi(v)) + \mu'(\varphi(v))\\ &= (\mu\circ\varphi)(v) + (\mu'\circ\varphi)(v) = \varphi^*(\mu)(v) + \varphi^*(\mu')(v) = \big(\varphi^*(\mu)+\varphi^*(\mu')\big)(v);\end{aligned}$$

also $\varphi^*(\mu+\mu') = \varphi^*(\mu)+\varphi^*(\mu')$; analog zeigt man $\varphi^*(s.\mu) = s.\varphi^*(\mu)$ für $s \in K$.

Es folgt eine Interpretation der transponierten Matrix mittels Homomorphismen:

Satz 38.11. *Sei* $\varphi\colon V \to W$ *linear, wobei* $n = \dim V < \infty$ *und* $m = \dim W < \infty$. *Sei* B *eine Basis von* V *und* C *eine Basis* C. *Sei* $A = M_C^B(\varphi) \in K^{m\times n}$ *die Matrix von* φ *bezüglich* B, C. *Dann gilt* $M_{B^*}^{C^*}(\varphi^*) = A^{\mathrm{tr}} \in K^{n\times m}$, *wobei* B^* *und* C^* *die zu* B *und* C *dualen Basen sind.*

Beweis. Sei $B = \{v_1, \ldots, v_n\}$ und $C = \{w_1, \ldots, w_m\}$. Nach Definition von $A = [a_{ij}] \in K^{m\times n}$ gilt $\varphi(v_j) = \sum_{i=1}^m a_{ij}v_i$ für $j = 1, \ldots, n$. Nun folgt einerseits für $1 \leqslant j \leqslant m$ und $1 \leqslant l \leqslant n$:

$$\big(\varphi^*(w_j^*)\big)(v_l) = (w_j^*\circ\varphi)(v_l) = w_j^*(\varphi(v_l)) = \sum_{k=1}^m a_{kl}w_j^*(w_k) = \sum_{k=1}^m a_{kl}\delta_{jk} = a_{jl};$$

und andererseits: $$\Big(\sum_{i=1}^n a_{ji}v_i^*\Big)(v_l) = \sum_{i=1}^n a_{ji}v_i^*(v_l) = \sum_{i=1}^n a_{ji}\delta_{il} = a_{jl}.$$

Also gilt $\varphi^*(w_j^*) = \sum_{i=1}^n a_{ji}v_i^*$ für $j = 1, \ldots, m$, d.h., $M_{B^*}^{C^*}(\varphi^*) = A^{\mathrm{tr}}$. □

39. *Multilineare Abbildungen und Tensorprodukte*

Genauso wie Determinanten sind Tensorprodukte in einer einführenden Vorlesung zur Linearen Algebra meist ein etwas kniffliges Thema. Das liegt einerseits daran, dass die Definition ziemlich abstrakt ist; andererseits kommen substanzielle Anwendungen von Tensorprodukten meist erst in weiterführenden Vorlesungen vor. Nichtsdestotrotz sind Tensorprodukte ein vielseitiges und grundlegendes Instrument mit diversen Anwendungen nicht nur in der Mathematik selbst, sondern zum Beispiel auch in der Physik.
Nachdem wir bereits Bilinearformen und Dualräume behandelt haben, bietet sich hier eine erste, kurze Einführung zu Tensorprodukten an, wobei wir einem Online-Artikel von Timothy Gowers (Fields-Medaille 1998) folgen mit dem Titel "How to loose your fear of tensor products"; siehe `https://www.dpmms.cam.ac.uk/~wtg10/tensors3.html`.

Sei K ein Körper. Sei $d \geqslant 2$ und seien $V_1, \ldots, V_d$ Vektorräume über K. Wir betrachten das kartesische Produkt $V_1 \times \ldots \times V_d = \{(v_1, \ldots, v_d) \mid v_i \in V_i \text{ für alle } i\}$. Sei auch W ein K-Vektorraum. Eine Abbildung $f\colon V_1 \times \ldots \times V_d \to W$ heißt ***multilinear***, wenn f linear in jedem Argument ist, d.h., für jedes $i \in \{1, \ldots, d\}$ gilt

$$f(v_1, \ldots, v_{i-1}, s.v + t.w, v_{i+1}, \ldots, d_d)$$

$$= s.f(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_d) + t.f(v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, d_d)$$

wobei $v, w \in V_i$, $s, t \in K$ und $v_j \in V_j$ für alle $j \neq i$. Sei nun $W = K$. Dann bezeichnen wir mit $\mathcal{L}(V_1, \dots, V_d; K)$ die Menge aller multilinearen Abbildungen $f\colon V_1 \times \dots \times V_d \to K$; man sieht sofort, dass dies ein Teilraum von $\mathrm{Abb}(V_1 \times \dots \times V_d, K)$ ist. Wir betrachten im Folgenden den Dualraum $\mathcal{L}(V_1, \dots, V_d; K)^*$.

Definition 39.1. Sind $v_i \in V_i$ für $i = 1, \dots, d$ gegeben, so erhalten wir eine Linearform $\mu_{v_1,\dots,v_d} \in \mathcal{L}(V_1, \dots, V_d; K)^*$ mit $\mu_{v_1,\dots,v_d}(f) := f(v_1, \dots, v_d)$ für alle $f \in \mathcal{L}(V_1, \dots, V_d; K)$.
Das ***Tensorprodukt*** von $V_1, \dots, V_d$ ist dann definiert als der Teilraum

$$V_1 \otimes \dots \otimes V_d := \left\langle \mu_{v_1,\dots,v_d} \mid v_i \in V_i \text{ für } i = 1, \dots, d \right\rangle_K \subseteq \mathcal{L}(V_1, \dots, V_d; K)^*.$$

Wir bezeichnen $\mu_{v_1,\dots,v_d}$ auch mit $v_1 \otimes \dots \otimes v_d$. Mit dieser Bezeichnung ist also

$$V_1 \otimes \dots \otimes V_d = \left\langle v_1 \otimes \dots \otimes v_d \mid v_i \in V_i \text{ für } i = 1, \dots, d \right\rangle_K.$$

Lemma 39.2. *Sei $W := V_1 \otimes \dots \otimes V_d$. Dann ist die Abbildung $\tau\colon V_1 \times \dots \times V_d \to W$, $(v_1, \dots, v_d) \mapsto v_1 \otimes \dots \otimes v_d$, multilinear.*

Beweis. Sei $i \in \{1, \dots, d\}$. Seien $v, w \in V_i$, $s, t \in K$ und $v_j \in V_j$ für alle $j \neq i$. Für $f \in \mathcal{L}(V_1, \dots, V_d; K)$ gilt dann

$$\tau(v_1, \dots, v_{i-1}, s.v + t.w, v_{i+1}, \dots, v_d)(f) = (v_1 \otimes \dots \otimes v_{i-1} \otimes (s.v + t.w) \otimes v_{i+1} \otimes \dots \otimes v_d)(f)$$
$$= \mu_{v_1,\dots,v_{i-1},s.v+t.w,v_{i+1},\dots,v_d}(f) = f(v_1, \dots, v_{i-1}, s.v + t.w, v_{i+1}, \dots, v_d).$$

Weil f multilinear ist, ist dies gleich

$$sf(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_d) + tf(v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_d) = \dots$$
$$= (s.\tau(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_d) + t.\tau(v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_d))(f).$$

Da f beliebig war, folgt also

$$\tau(v_1, \dots, v_{i-1}, s.v + t.w, v_{i+1}, \dots, v_d)$$
$$= s.\tau(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_d) + t.\tau(v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_d).$$ □

Satz 39.3. *Seien $V_1, \dots, V_d$ Vektorräume über K mit $\dim V_i < \infty$ für $1 \leqslant i \leqslant d$. Dann gilt*

$$\dim(V_1 \otimes \dots \otimes V_d) = (\dim V_1) \cdot \ldots \cdot (\dim V_d) < \infty.$$

Beweis. Sei $B_i = \{b_j^{(i)} \mid j \in I_i\}$ Basis von V_i, wobei I_i eine Indexmenge ist mit $|I_i| = \dim V_i$.
Behauptung: $B := \{b_{j_1}^{(1)} \otimes \dots \otimes b_{j_d}^{(d)} \mid j_1 \in I_1, \dots, j_d \in I_d\}$ ist eine Basis von $V_1 \otimes \dots \otimes V_d$.

Dazu: $V_1 \otimes \dots \otimes V_d$ wird erzeugt von allen $v_1 \otimes \dots \otimes v_d$ mit $v_i \in V_i$. Jedes v_i ist hier eine Linearkombination von B_i. Weil τ wie in Lemma 39.2 multilinear ist, folgt sofort, dass $v_1 \otimes \dots \otimes v_d$ eine Linearkombination von B ist. Also gilt $V_1 \otimes \dots \otimes V_d = \langle B \rangle_K$. Es ist dann noch zu zeigen, dass B linear unabhängig ist. Sei

$$\sum_{j_1 \in I_1,\dots,j_d \in I_d} s_{j_1,\dots,j_d}.b_{j_1}^{(1)} \otimes \dots \otimes b_{j_d}^{(d)} = 0_{V_1 \otimes \dots \otimes V_d} \qquad \text{mit} \qquad s_{j_1,\dots,j_d} \in K.$$

Sei $B_i^* = \{\hat{b}_j^{(i)} \mid j \in I_i\}$ die zu B_i duale Basis von V_i^*. Seien $k_i \in I_i$ für $i = 1, \dots, d$. Dann definiere $\lambda_{k_1,\dots,k_d}\colon V_1 \times \ldots \times V_d \to K$ durch

$$\lambda_{k_1,\dots,k_d}(v_1,\dots,v_d) := \hat{b}^{(1)}_{k_1}(v_1)\cdot\ldots\cdot\hat{b}^{(d)}_{k_d}(v_d) \qquad \text{für alle } v_i \in V_i.$$

Man sieht sofort, dass $\lambda_{k_1,\dots,k_d}$ multilinear ist, also $\lambda_{k_1,\dots,k_d} \in \mathcal{L}(V_1,\dots,V_d;K)$. Nach obiger Definition ist $V_1 \otimes \ldots \otimes V_d \subseteq \mathcal{L}(V_1,\dots,V_d;K)^*$. Es folgt nun:

$$\begin{aligned} 0 &= \sum_{j_1\in I_1,\dots,j_d\in I_d} s_{j_1,\dots,j_d}\big(b^{(1)}_{j_1}\otimes\ldots\otimes b^{(d)}_{j_d}\big)(\lambda_{k_1,\dots,k_d}) = \sum_{j_1\in I_1,\dots,j_d\in I_d} s_{j_1,\dots,j_d}\lambda_{k_1,\dots,k_d}(b^{(1)}_{j_1},\dots,b^{(d)}_{j_d}) \\ &= \sum_{j_1\in I_1,\dots,j_d\in I_d} s_{j_1,\dots,j_d}\underbrace{\hat{b}^{(1)}_{k_1}(b^{(1)}_{j_1})}_{=\delta_{k_1j_1}}\cdots\underbrace{\hat{b}^{(d)}_{k_d}(b^{(d)}_{j_d})}_{=\delta_{k_dj_d}} = s_{k_1,\dots,k_d}. \end{aligned}$$

Also sind alle Koeffizienten in obiger Linearkombination gleich 0. □

Der folgende Satz ist das zentrale Hilfsmittel, um mit Tensorprodukten zu arbeiten.

Satz 39.4 (***Universelle Eigenschaft des Tensorprodukts***). *Gegeben seien ein Vektorraum W über K und eine multilineare Abbildung $\varphi\colon V_1 \times \ldots \times V_d \to W$.*
Dann gibt es genau eine lineare Abbildung $\tilde{\varphi}\colon V_1 \otimes \ldots \otimes V_d \to W$ mit $\varphi = \tilde{\varphi}\circ\tau$, d.h., es gilt $\tilde{\varphi}(v_1\otimes\ldots\otimes v_d) = (\tilde{\varphi}\circ\tau)(v_1,\dots,v_d) = \varphi(v_1,\dots,v_d)$ für alle $v_i \in V_i$ $(1 \leqslant i \leqslant d)$.

Beweis. Wir beginnen mit der Existenz von $\tilde{\varphi}$. Dazu betrachten wir zuerst den Spezialfall $W = K$. In diesem Fall ist $\varphi \in \mathcal{L}(V_1,\dots,V_d;K)$ und wir erhalten eine lineare Abbildung $\mathcal{L}(V_1,\dots,V_d;K)^* \to K$, $\mu \mapsto \mu(\varphi)$. Wegen $V_1 \otimes \ldots \otimes V_d \subseteq \mathcal{L}(V_1,\dots,V_d;K)$ erhalten wir durch Einschränkung eine lineare Abbildung $\tilde{\varphi}\colon V_1\otimes\ldots\otimes V_d \to K$ mit

$$\tilde{\varphi}(v_1\otimes\ldots\otimes v_d) = (v_1\otimes\ldots\otimes v_d)(\varphi) = \mu_{v_1,\dots,v_d}(\varphi) = \varphi(v_1,\dots,v_d)$$

für alle $v_i \in V_i$. Die linke Seite ist gleich $(\tilde{\varphi}\circ\tau)(v_1,\dots,v_d)$; also folgt $\tilde{\varphi}\circ\tau = \varphi$.

Nun zum allgemeinen Fall, wo W ein beliebiger K-Vektorraum ist. Sei B eine Basis von W. Wir schreiben $B = \{b_i \mid i \in I\}$ wie in Bemerkung 36.5; für $i \in I$ sei auch die lineare Abbildung $\gamma_i\colon W \to K$ wie dort definiert. Nun ist $\gamma_i\circ\varphi \in \mathcal{L}(V_1,\dots,V_d;K)$. Nach dem gerade behandelten Spezialfall gibt es also eine lineare Abbildung

$$\tilde{\varphi}_i\colon V_1\otimes\ldots\otimes V_d \to K \qquad \text{mit} \qquad \gamma_i\circ\varphi = \tilde{\varphi}_i\circ\tau.$$

Damit können wir jetzt eine Abbildung $\tilde{\varphi}\colon V_1\otimes\ldots\otimes V_d \to W$ definieren durch:

$$\tilde{\varphi}(x) := \sum_{i\in I}\tilde{\varphi}_i(x).b_i \qquad \text{für alle } x \in V_1\otimes\ldots\otimes V_d.$$

(Beachte: Falls $\dim W = \infty$, so muss man sich hier zuerst überlegen, dass es zu jedem x nur endlich viele $i \in I$ gibt mit $\tilde{\varphi}_i(x) \neq 0$.—Übung oder selbst.) Weil jedes $\tilde{\varphi}_i$ linear ist, folgt sofort, dass $\tilde{\varphi}$ linear ist. Außerdem ist

$$\begin{aligned} (\tilde{\varphi}\circ\tau)(v_1,\dots,v_d) &= \tilde{\varphi}(v_1\otimes\ldots\otimes v_d) = \textstyle\sum_{i\in I}\tilde{\varphi}_i(v_1\otimes\ldots\otimes v_d).b_i \\ &= \textstyle\sum_{i\in I}\gamma_i(\varphi(v_1,\dots,v_d)).b_i = \varphi(v_1,\dots,v_d) \end{aligned}$$

für alle $v_i \in V_i$; also $\tilde{\varphi}\circ\tau = \varphi$.

Schließlich zur Eindeutigkeit von $\tilde{\varphi}$: Sei auch $\psi\colon V_1\otimes\ldots\otimes V_d \to W$ linear mit $\varphi = \psi\circ\tau$.

Für $v_i \in V_i$ gilt $\psi(v_1 \otimes \ldots \otimes v_d) = \psi(\tau(v_1, \ldots, v_d)) = (\psi \circ \tau)(v_1, \ldots, v_d) = \varphi(v_1, \ldots, v_d)$, und analog $\tilde{\varphi}(v_1 \otimes \ldots \otimes v_d) = \varphi(v_1, \ldots, v_d)$. Also gilt $\psi(v_1 \otimes \ldots \otimes v_d) = \tilde{\varphi}(v_1 \otimes \ldots \otimes v_d)$ für alle $v_i \in V_i$. Wegen $V_1 \otimes \ldots \otimes V_d = \langle v_1 \otimes \ldots \otimes v_d \mid v_i \in V_i \text{ für } i = 1, \ldots, d \rangle_K$ folgt $\psi = \tilde{\varphi}$. $\square$

Im Folgenden werden wir praktisch nie wieder die ursprüngliche Definition des Tensorprodukts benutzen, sondern immer nur die obige "universelle Eigenschaft".

Beispiel 39.5. Seien V, W Vektorräume über K. Für $\lambda \in V^*$ und $w \in W$ erhalten wir eine Abbildung $\sigma_{\lambda,w} \colon V \to W$ mit $\sigma_{\lambda,w}(v) := \lambda(v).w$ für alle $v \in V$. Man prüft sofort nach, dass $\sigma_{\lambda,w}$ linear ist und dann die Abbildung $\sigma \colon V^* \times W \to \mathrm{Hom}(V, W)$, $(\lambda, w) \mapsto \sigma_{\lambda,w}$, bilinear ist. Nach der universellen Eigenschaft des Tensorprodukts gibt es also eine lineare Abbildung

$$\tilde{\sigma} \colon V^* \otimes W \to \mathrm{Hom}(V, W) \qquad \text{mit} \qquad \tilde{\sigma}(\lambda \otimes w) = \sigma_{\lambda,w} \text{ für alle } \lambda \in V^* \text{ und } w \in W.$$

Behauptung: Sind $\dim V < \infty$ und $\dim W < \infty$, so ist $\tilde{\sigma}$ ein Isomorphismus.

Dazu: Sei $n = \dim V$ und $m = \dim W$. Sei $B = \{v_1, \ldots, v_n\}$ eine Basis von V und $C := \{w_1, \ldots, w_m\}$ eine Basis von W. Sei $B^* = \{v_1^*, \ldots, v_n^*\}$ die duale Basis von V^*. Für $1 \leqslant k \leqslant m$ und $1 \leqslant l \leqslant n$ definiere $\psi_{kl} \in \mathrm{Hom}(V, W)$ durch $\psi_{kl}(v_l) := w_k$ und $\psi_{kl}(v_j) := 0_W$ falls $j \neq l$. Dann ist $M_C^B(\psi_{kl})$ die Standardmatrix $E_{kl}^{(m,n)} \in K^{m \times n}$ wie in Kapitel III, Beispiel 11.7 (mit 1 an der Position (k, l) und 0 überall sonst). Diese Standardmatrizen bilden eine Basis von $K^{m \times n}$; also ist $\{\psi_{kl} \mid 1 \leqslant k \leqslant m, 1 \leqslant l \leqslant n\}$ eine Basis von $\mathrm{Hom}(V, W)$ (nach Kapitel IV, Satz 19.19). Man prüft sofort nach, dass $\psi_{kl}(v_j) = \sigma_{v_l^*, w_k}(v_j)$ für $1 \leqslant j \leqslant n$ gilt, also ist $\psi_{kl} = \sigma_{v_l^*, w_k} = \tilde{\sigma}(v_l^* \otimes w_k) \in \mathrm{Bild}(\tilde{\sigma})$. Weil die ψ_{kl} eine Basis von $\mathrm{Hom}(V, W)$ bilden, ist also $\tilde{\sigma}$ surjektiv. Andererseits gilt $\dim(V^* \otimes W) = (\dim V^*)(\dim W) = (\dim V)(\dim W) = \dim \mathrm{Hom}(V, W)$, also ist $\tilde{\sigma}$ auch injektiv.

Beispiel 39.6. Sei V ein K-Vektorraum und $\varphi \colon V \to V$ linear. Sei $n = \dim V < \infty$ und $B = \{v_1, \ldots, v_n\}$ eine Basis von V. Bilden wir die darstellende Matrix $A = [a_{ij}] = M_B(\varphi) \in M_n(K)$, so haben wir $\mathrm{Spur}(A) := \sum_{i=1}^n a_{ii} \in K$ definiert, und in Lemma 24.3 gesehen, dass $\mathrm{Spur}(A)$ nicht von der Wahl der Basis B anhängt. Man kann also auch $\mathrm{Spur}(\varphi) := \mathrm{Spur}(M_B(\varphi))$ definieren. Gibt es eine direkte Definition von $\mathrm{Spur}(\varphi)$, ohne den Umweg über darstellende Matrizen? Die Antwort is JA, und zwar mit Hilfe des Tensorprodukts! Betrachte dazu die Abbildung $\psi \colon V^* \times V \to K$, $(\lambda, v) \mapsto \lambda(v)$. Man sieht sofort, dass ψ bilinear ist, also gibt es eine lineare Abbildung $\tilde{\psi} \colon V^* \otimes V \to K$ mit $\tilde{\psi}(\lambda \otimes v) = \lambda(v)$ für alle $\lambda \in V^*$ und $v \in V$. Nach Beispiel 39.5 haben wir einen Isomorphismus $\tilde{\sigma} \colon V^* \otimes V \to \mathrm{Hom}(V, V)$. Wir überlassen es dann als Übung zu zeigen, dass $\mathrm{Spur}(\varphi) = (\tilde{\psi} \circ \tilde{\sigma}^{-1})(\varphi)$ für alle $\varphi \in \mathrm{Hom}(V, V)$ gilt.

Ab hier Woche 14

Beispiel 39.7. Seien V, W Vektorräume über K. Seien $\varphi \colon V \to V$ und $\psi \colon W \to W$ linear. Dann ist $V \times W \to V \otimes W$, $(v, w) \mapsto \varphi(v) \otimes \psi(w)$, bilinear; also gibt es nach der universellen Eigenschaft eine lineare Abbildung $V \otimes W \to V \otimes W$ mit $v \otimes w \mapsto \varphi(v) \otimes \psi(w)$ für alle

$v \in V$ und $w \in W$. Diese wird mit $\varphi \otimes \psi$ bezeichnet; es gilt also

$$(\varphi \otimes \psi)(v \otimes w) = \varphi(v) \otimes \psi(w) \qquad \text{für alle } v \in V \text{ und } w \in W.$$

Sei nun $n = \dim V < \infty$ und $m = \dim W < \infty$. Sei $B_1 = \{v_1, \dots, v_n\}$ eine Basis von V und $B_2 = \{w_1, \dots, w_m\}$ von W. Sei $A = M_{B_1}(\varphi) \in M_n(K)$ und $B := M_{B_2}(\psi) \in M_m(K)$. Was ist die darstellende Matrix von $\varphi \otimes \psi$ bezüglich der Basis von $V \otimes W$ in Satz 39.3? Die Antwort ist die Blockmatrix

$$A \otimes B := \left[\begin{array}{c|c|c} a_{11}B & \dots & a_{1n}B \\ \hline \vdots & & \vdots \\ \hline a_{n1}B & \dots & a_{nn}B \end{array}\right] \in M_{nm}(K),$$

die auch als ***Kronecker–Produkt*** von A und B bezeichnet wird. (Details siehe Übungen.)

Beispiel 39.8. Sei K in einem größeren Körper L enthalten, so dass die Operationen in K durch die Einschränkungen der entsprechenden Operationen in L gegeben sind. Typische Beispiele sind $K = \mathbb{Q} \subseteq L = \mathbb{R}$ oder $K = \mathbb{R} \subseteq L = \mathbb{C}$. Sei V ein beliebiger K-Vektorraum. Wie in Kapitel IV, Beispiel 16.1(b), bemerkt, ist in obiger Situation auch L ein K-Vektorraum. Also können wir das Tensorprodukt $V_L := L \otimes V$ bilden, das als ***Skalarerweiterung*** von K nach L bezeichnet wird. Wir behaupten, dass V_L auch ein L-Vektorraum ist. Dazu müssen wir eine skalare Multiplikation $L \times V_L \to V_L$ definieren. Dies geht wie folgt: Für ein festes $s \in L$ betrachte die Abbildung $L \times V \to V_L$, $(t, v) \mapsto (st) \otimes v$. Diese ist K-bilinear, also gibt es eine K-lineare Abbildung $\gamma_s \colon V_L \to V_L$ mit $\gamma_s(t \otimes v) = (st) \otimes v$ für alle $t \in L$ und $v \in V$. Dies liefert die gewünschte Abbildung $L \times V_L \to V_L$, $(s, x) \mapsto \gamma_s(x)$. (Man muss natürlich noch zeigen, dass damit die Vektorraumaxiome erfüllt sind.) Es gilt dann auch: Ist B eine Basis von V, so ist $B_L := \{1 \otimes b \mid b \in B\}$ eine Basis von V_L (als L-Vektorraum); insbesondere ist die Dimension von V über K gleich der Dimension von V_L über L.

40. *Affine und projektive Räume*

In jedem Vektorraum V gibt es einen ausgezeichneten Punkt, nämlich 0_V; jede lineare Abbildung von V in sich lässt diesen Punkt fest. Motiviert durch Anwendungen ist es allerdings wünschenswert, einen Formalismus zu entwickeln, in dem alle Punkte gleichberechtigt sind und auch allgemeinere Abbildungen benutzt werden können (etwa die Bewegungen bei Hauptachsentransformationen). Zum Beispiel in physikalischen Anwendungen wirken Kräfte an Punkten des 3-dimensionalen Anschauungsraums; diese haben eine Richtung und einen Wert, und können durch einen Pfeil veranschaulicht werden, der von einem beliebigen Punkt ausgeht. Dies führt zur abstrakten Definition eines affinen Raumes.

Definition 40.1. Sei K ein Körper und V ein K-Vektorraum. Eine Menge $\mathcal{A} \neq \varnothing$ heißt ***affiner Raum*** über V, wenn es eine Abbildung $\mathcal{A} \times \mathcal{A} \to V$, $(P, Q) \mapsto \overrightarrow{PQ}$, gibt mit folgenden Eigenschaften:

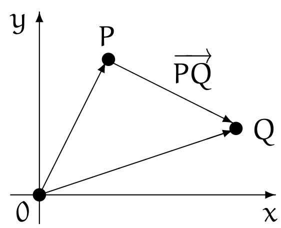
(A1) Zu jedem $P \in \mathcal{A}$ und $v \in V$ gibt es genau ein $Q \in \mathcal{A}$ mit $\overrightarrow{PQ} = v$.
(A2) Für $P, Q, R \in \mathcal{A}$ gilt $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$.

Zu festem $v \in V$ können wir damit auch eine Abbildung $\tau_v \colon \mathcal{A} \to \mathcal{A}$ wie folgt definieren: Zu $P \in \mathcal{A}$ gibt es nach (A1) genau ein $Q \in \mathcal{A}$ mit $\overrightarrow{PQ} = v$; setze dann $\tau_v(P) := Q$. Die Abbildung τ_v wird als ***Translation*** (oder auch "Verschiebung") bezeichnet.

Die Elemente von $\mathcal{A}$ bezeichnen wir meist als "Punkte", und $\overrightarrow{PQ}$ als den "Pfeil" von P nach Q. Wir setzen auch $\dim \mathcal{A} := \dim V$.

Bemerkung 40.2. (a) Für alle $P, Q \in \mathcal{A}$ gilt $\overrightarrow{PQ} = -\overrightarrow{QP}$ und $\overrightarrow{PP} = 0_V$. Denn: Nach (A2) gilt $\overrightarrow{PP} + \overrightarrow{PP} = \overrightarrow{PP}$, also $\overrightarrow{PP} = 0_V$. Wiederum nach (A2) folgt dann auch $\overrightarrow{PQ} + \overrightarrow{QP} = \overrightarrow{PP} = 0_V$.
(b) Zu festem $P_0 \in \mathcal{A}$ ist $\mathcal{A} \to V$, $Q \mapsto \overrightarrow{P_0Q}$, bijektiv. (Dies folgt sofort aus (A1).)

Beispiel 40.3. (a) Sei V ein K-Vektorraum und $\mathcal{A} := V$. Zu $P, Q \in V$ definiere $\overrightarrow{PQ} := Q - P \in V$. Dann gelten (A1), (A2). Dies ist der ***affine Standardraum*** zu V.
Für $V = K^2$ heißt $\mathcal{A}$ ***affine Ebene*** über K. Also zum Beispiel für $\mathcal{A} = V = \mathbb{R}^2$:



$\overrightarrow{0P} + \overrightarrow{PQ} = \overrightarrow{0Q}$
(Vektoraddition)

(b) Sei W ein K-Vektorraum und $V \subseteq W$ ein Teilraum. Sei $w_0 \in W$ fest und setze $\mathcal{A} := w_0 + V = \{w_0 + v \mid v \in V\}$; dies ist also ein affiner Unterraum wie in §37. Für $P, Q \in \mathcal{A}$ setze wieder $\overrightarrow{PQ} := Q - P \in V$. Dann ist $\mathcal{A}$ affiner Raum über V. Konkretes Beispiel:

$$V := \left\{ \begin{bmatrix} 0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_i \in K \right\} \subseteq W := K^{n+1} \qquad \text{und} \qquad \mathcal{A} := \left\{ \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} \;\middle|\; x_i \in K \right\}.$$

Beispiel 40.4. Sei $\mathcal{A} = V = \mathbb{R}^2$ die affine Ebene über $\mathbb{R}^2$ wie in Beispiel 40.3(a). (Wir schreiben der Einfachheit halber die Vektoren in $\mathbb{R}^2$ als Zeilenvektoren.)
Gegeben seien zwei Punkte $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ mit $(x_1, y_1) \neq (x_2, y_2)$. Dann gibt es genau eine Gerade durch diese beiden Punkte; diese ist gegeben durch

$$G := \{(u, v) \in \mathbb{R}^2 \mid au + bv = c\}, \quad \text{wobei } a := y_1 - y_2,\ b := x_2 - x_1,\ c := x_2y_1 - x_1y_2;$$

beachte $a \neq 0$ oder $b \neq 0$ gilt. Seien nun zwei Geraden G_1, G_2 gegeben, also

$$G_i = \{(u, v) \in \mathbb{R}^2 \mid a_iu + b_iv = c_i\} \quad \text{für } i = 1, 2 \text{ (wobei } a_i, b_i, c_i \in \mathbb{R}).$$

Dann erhält man die Schnittmenge $G_1 \cap G_2$ als Lösungsmenge des LGS $a_1x + b_1y = c_1$, $a_2x + b_2y = c_2$. Es gilt also 3 Fälle: Entweder ist $G_1 \cap G_2 = \varnothing$ (die Geraden sind parallel, verschieden), oder $|G_1 \cap G_2| = 1$ (die Geraden sind nicht parallel) oder $G_1 = G_2$.

Definition 40.5. Sei $\mathcal{A}$ ein affiner Raum über V und $n = \dim V < \infty$. Sei $P_0 \in \mathcal{A}$ fest. Eine Menge $S = \{P_0, P_1, \ldots, P_n\} \subseteq \mathcal{A}$ heißt ***affines Koordinatensystem***, wenn $B := \{\overrightarrow{P_0P_1}, \ldots, \overrightarrow{P_0P_n}\}$ eine Basis von V ist; der Punkt P_0 heißt dann hier "Ursprung".
Ist $P \in \mathcal{A}$ gegeben, so setze $v := \overrightarrow{P_0P} \in V$. Dann ist $P = \tau_v(P_0)$ und v lässt sich auf eindeutige Weise schreiben als $v = \sum_{i=1}^n s_i.\overrightarrow{P_0P_i}$ mit $s_i \in K$.

Ist zum Beispiel $\mathcal{A} = V = K^n$ und $B = \{e_1, \ldots, e_n\}$ die Standardbasis von K^n, so ist $S = \{0_n, e_1, \ldots, e_n\}$ das affine "Standard"-Koordinatensystem (mit "Ursprung" 0_n); aber man könnte eben auch jeden anderen Punkt $P_0 \in \mathcal{A} = K^n$ als "Ursprung" nehmen.

Lemma 40.6. *Sei $\mathcal{A}$ ein affiner Raum über V. Gegeben seien $P_1, \ldots, P_n \in \mathcal{A}$ ($n \geqslant 2$) und $s_1, \ldots, s_n \in K$ mit $s_1 + \ldots + s_n = 1$. Dann gibt es genau ein $Q \in \mathcal{A}$ mit*

$$\overrightarrow{PQ} = s_1.\overrightarrow{PP_1} + \ldots + s_n.\overrightarrow{PP_n} \qquad \text{für alle } P \in \mathcal{A}.$$

Dann heißt Q das ***Baryzentrum*** (oder der ***Schwerpunkt***) der mit den s_i gewichteten Punkte P_i. (Zur Veranschaulichung stelle man sich s_i als die "Masse" von P_i vor.)

Beweis. Sei $P \in \mathcal{A}$ und $v := s_1.\overrightarrow{PP_1} + \ldots + s_n.\overrightarrow{PP_n} \in V$. Nach (A1) gibt es genau ein $Q \in \mathcal{A}$ mit $\overrightarrow{PQ} = v$. Sei nun auch $P' \in \mathcal{A}$ und $v' := s_1.\overrightarrow{P'P_1} + \ldots + s_n.\overrightarrow{P'P_n} \in V$; sei $Q' \in \mathcal{A}$ mit $\overrightarrow{P'Q'} = v'$. Mit (A2) ist $\overrightarrow{PP'} = \overrightarrow{PP_i} + \overrightarrow{P_iP'} = \overrightarrow{PP_i} - \overrightarrow{P'P_i}$ für alle i und damit

$$\overrightarrow{PP'} = \Big(\sum_{i=1}^n s_i\Big).\overrightarrow{PP'} = \sum_{i=1}^n s_i.\overrightarrow{PP'} = \sum_{i=1}^n s_i.\overrightarrow{PP_i} - \sum_{i=1}^n s_i.\overrightarrow{P_iP'} = v - v'.$$

Es folgt $\overrightarrow{PQ'} = \overrightarrow{PP'} + \overrightarrow{P'Q'} = (v - v') + v' = v = \overrightarrow{PQ}$, also $Q = Q'$ mit (A1). □

Definition 40.7. Seien V, V' Vektorräume über K. Sei $\mathcal{A}$ ein affiner Raum über V und $\mathcal{A}'$ ein affiner Raum über V'. Eine Abbildung $\alpha\colon \mathcal{A} \to \mathcal{A}'$ heißt ***affine Abbildung***, wenn es eine lineare Abbildung $\varphi\colon V \to V'$ gibt mit $\varphi(\overrightarrow{PQ}) = \overrightarrow{\alpha(P)\alpha(Q)}$ für alle $P, Q \in \mathcal{A}$.

Beispiel 40.8. Sei $\mathcal{A}$ ein affiner Raum über V. Für jedes $v \in V$ ist die Translation $\tau_v\colon \mathcal{A} \to \mathcal{A}$ eine bijektive affine Abbildung.

Dazu: Die Bijektivität folgt sofort aus (A1). Seien nun $P, Q \in \mathcal{A}$; setze $P' := \tau_v(P)$ und $Q' := \tau_v(Q)$; also $\overrightarrow{PP'} = v$ und $\overrightarrow{QQ'} = v$. Mit (A2) folgt $\overrightarrow{\tau_v(P)\tau_v(Q)} = \overrightarrow{P'Q'} = \overrightarrow{P'Q} + \overrightarrow{QQ'} = (\overrightarrow{P'P} + \overrightarrow{PQ}) + \overrightarrow{QQ'}$. Nach Bemerkung 40.2(a) gilt $\overrightarrow{P'P} = -\overrightarrow{PP'} = -v = -\overrightarrow{QQ'}$, also ist $(\overrightarrow{P'P} + \overrightarrow{PQ}) + \overrightarrow{QQ'} = (-v + \overrightarrow{PQ}) + v = \overrightarrow{PQ}$.

Beispiel 40.9. Sei $\mathcal{A} = V = K^n$ und $\mathcal{A}' = V' = K^m$ (jeweils wie in Beispiel 40.3(a)). Eine Abbildung $\alpha\colon \mathcal{A} \to \mathcal{A}'$ ist affin genau dann, wenn es eine Matrix $A \in K^{m \times n}$ und einen Vektor $b \in K^m$ gibt mit $\alpha(v) = A \cdot v + b$ für alle $v \in K^n$.
Dazu: Sei zuerst α affin und $\varphi\colon K^n \to K^m$ die zugehörige lineare Abbildung. Dann gibt es

eine Matrix $A \in K^{m\times n}$ mit $\varphi(v) = A \cdot v$ für alle $v \in K^n$; setze $b := \alpha(0_n) \in K^m$. Sei nun $P \in \mathcal{A} = K^n$ und $v := \overrightarrow{0_n P} = P - 0_n = P$; dann folgt $A \cdot v = \varphi(v) = \varphi(\overrightarrow{0_n P}) = \overrightarrow{\alpha(0_n)\alpha(P)} = \overrightarrow{b\alpha(P)}$, also $A \cdot v = \alpha(P) - b$. Seien umgekehrt $A \in K^{m\times n}$ und $b \in K^m$ gegeben; definiere $\varphi\colon K^n \to K^m$ durch $\varphi(v) := A \cdot v$ für alle $v \in K^n$. Für $P, Q \in \mathcal{A} = K^n$ folgt dann $\varphi(\overrightarrow{PQ}) = \varphi(Q - P) = \varphi(Q) - \varphi(P) = \varphi(Q) + b - (\varphi(P) + b) = \alpha(P) - \alpha(Q) = \overrightarrow{\alpha(P)\alpha(Q)}$.

Definition 40.10. Sei V ein K-Vektorraum. Dann heißt $\mathbb{P}(V) := \{\langle v\rangle_K \mid 0_V \neq v \in V\}$ der ***projektive Raum*** über V; als Menge besteht $\mathbb{P}(V)$ also genau aus den 1-dimensionalen Teilräumen von V. (Ist $V = \{0_V\}$, so gilt $\mathbb{P}(V) = \varnothing$.) Ist $1 \leqslant n = \dim V < \infty$, so heißt $\dim \mathbb{P}(V) := n - 1$ die Dimension von $\mathbb{P}(V)$.
Beachte: Für $0 \neq s \in K$ und $0_V \neq v \in V$ gilt $\langle v\rangle_K = \langle s.v\rangle_K$; der durch v definierte 1-dimensionale Teilraum ändert sich also nicht bei Multiplikation mit einem Skalar $\neq 0$.

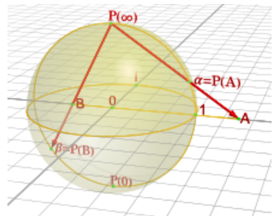
Ist $V = K^{n+1}$, so bezeichne kurz $\mathbb{P}^n(K) := \mathbb{P}(K^{n+1})$. Ist $0_{n+1} \neq v \in K^{n+1}$ gegeben, mit Komponenten $x_0, x_1, \ldots, x_n \in K$, so schreiben wir auch $\langle v\rangle_K = (x_0 : x_1 : \ldots : x_n) \in \mathbb{P}^n(K)$.
Es gilt dann $(x_0 : x_1 : \ldots : x_n) = (sx_0 : sx_1 : \ldots : sx_n)$ für $0 \neq s \in K$.

Beispiel 40.11. Für $n = 0$ besteht $\mathbb{P}^0(K)$ aus einem Element, nämlich $\langle [1]\rangle_K$.
Sei nun $n = 1$. Wir beschreiben alle Elemente $(x : y) \in \mathbb{P}^1(K)$, wobei $x, y \in K$ mit $x \neq 0$ oder $y \neq 0$. Ist $x \neq 0$, so gilt $(x : y) = (1 : z)$ mit $z := x^{-1}y \in K$. Ist $x = 0$, so ist $y \neq 0$ und $(0 : y) = (0 : 1)$. Damit erhalten wir $\mathbb{P}^1(K) = \{(1 : z) \mid z \in K\} \cup \{(0 : 1)\}$. — Identifizieren wir $(1 : z)$ mit $z \in K$, und schreiben $P(\infty) := (0 : 1)$, so gilt also $\mathbb{P}^1(K) = K \cup \{P(\infty)\}$, d.h., $\mathbb{P}^1(K)$ entsteht aus K durch Hinzunahme eines weiteren Punktes.

Beispiel 40.12. Für $K = \mathbb{R}$ ist noch eine andere Konstruktion möglich. Betrachten wir nur den Fall $n = 2$. Sei $S^2 := \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ die Oberfläche der Kugel mit Radius 1 und Mittelpunkt im Ursprung von $\mathbb{R}^3$. (Wir schreiben hier wieder die Elemente von $\mathbb{R}^3$ als Zeilenvektoren.) Dann ist die Abbildung $S^2 \to \mathbb{P}^2(\mathbb{R})$, $(x, y, z) \mapsto (x : y : z)$, surjektiv. (Zu jedem $x = \langle v\rangle_{\mathbb{R}} \in \mathbb{P}^3(\mathbb{R})$ können wir v so wählen, dass $\|v\| = 1$ gilt; dann gibt es genau zwei Urbilder von x unter f, nämlich $\pm v$. Es ist also S^2 “fast” gleich zu $\mathbb{P}^2(\mathbb{R})$.)
Sei $N := (0, 0, 1) \in S^2$ der “Nordpol” der Kugel und definiere

$$f\colon S^2 \setminus \{N\} \to \mathbb{R}^2, \qquad (x, y, z) \mapsto \Big(\frac{x}{1 - z}, \frac{y}{1 - z}\Big).$$

Diese Abbildung ist bijektiv und heißt ***stereographische Projektion***. Anschaulich betrachtet man die Gerade durch N und $(x, y, z) \in S^2 \setminus \{N\}$; dann ist $f(x, y, z) \in \mathbb{R}^2$ der Schnittpunkt dieser Geraden mit der “Äquatorebene” $\{(x, y, 0) \mid x, y \in \mathbb{R}\}$ (wie im Bild unten, wo $N = P(\infty)$, $\alpha, \beta \in S^2$ und A, B die entsprechenden Punkte in der Äquatorebene sind; für mehr dazu siehe https://de.wikipedia.org/wiki/Projektiver_Raum).



Es folgt dann also $\mathbb{P}^2(\mathbb{R}) = \{\langle N\rangle_{\mathbb{R}}\} \cup \{\langle f^{-1}(u,v)\rangle_{\mathbb{R}} \mid u,v \in \mathbb{R}\}$.

Beispiel 40.13. Sei $n = 2$; dann heißt $\mathbb{P}^2(K)$ die projektive Ebene über K.
Eine Teilmenge $X \subseteq \mathbb{P}^2(K)$ heißt ***projektive Gerade***, wenn es ein Tripel $(a,b,c) \in K^{1\times 3}$ gibt mit $(a,b,c) \neq (0,0,0)$ und $X = \{(x:y:z) \in \mathbb{P}^2(K) \mid ax+by+cz = 0\}$.
Wir wollen zeigen, dass durch je zwei Punkte von $\mathbb{P}^2(K)$ genau eine projektive Gerade geht (völlig analog zur Situation in der affinen Ebene $\mathcal{A} = K^2$).
Seien also $\langle v_1\rangle_K \neq \langle v_2\rangle_K$ in $\mathbb{P}^2(K)$ gegeben. Dann ist das Tupel (v_1,v_2) linear unabhängig und $\dim\langle v_1,v_2\rangle^\perp = 1$ (siehe Lemma 30.6, wobei $\perp$ bezüglich des Standard-Skalarproduktes gebildet wird). Sei $(0,0,0) \neq (a,b,c) \in \langle v_1,v_2\rangle^\perp$. Wir betrachten die projektive Gerade $X = \{(x:y:z) \in \mathbb{P}^2(K) \mid ax+by+cz = 0\}$. Wegen $(a,b,c) \in \langle v_1,v_2\rangle_K^\perp$ folgt $\langle v_1\rangle_K, \langle v_2\rangle_K \in X$. Sei nun auch $(a',b',c') \neq (0,0,0)$ ein Tripel mit $\langle v_1\rangle_K, \langle v_2\rangle_K \in X' := \{(x:y:z) \in \mathbb{P}^2(K) \mid a'x+b'y+c'z = 0\}$. Dann ist wiederum $(a',b',c') \in \langle v_1,v_2\rangle_K^\perp$; wegen $\dim\langle v_1,v_2\rangle^\perp = 1$ muss also (a',b',c') ein skalares Vielfaches von (a,b,c) sein und damit $X' = X$.

Die folgende Aussage zeigt allerdings einen bemerkenswerten Unterschied zwischen der affinen und der projektiven Ebene. — Mehr dazu in einer Geometrie–Vorlesung.

Lemma 40.14. *Seien $X_1, X_2 \subseteq \mathbb{P}^2(K)$ beliebige projektive Geraden. Dann gilt $X_1 \cap X_2 \neq \varnothing$.*

Beweis. Für $i = 1,2$ schreiben wir $X_i = \{(x:y:z) \in \mathbb{P}^2(K) \mid a_ix+b_iy+c_iz = 0\}$ mit $(0,0,0) \neq (a_i,b_i,c_i) \in K^{1\times 3}$. Das homogene lineare Gleichungssystem mit Matrix
$$A := \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix} \in K^{2\times 3}$$
besitzt dann eine nicht-triviale Lösung $0_3 \neq v \in K^3$ (einfach deshalb, weil das homogene LGS mehr Variablen als Gleichungen hat). Es folgt $\langle v\rangle_K \in X_1 \cap X_2$. □

Ausblick

Wie geht's weiter in den nächsten Semestern im Bereich Algebra ?

Im 3. Semester gibt es die Algebra-Vorlesung. Sie ist nicht Pflicht, aber Grundlage für vieles Weitere im Bereich Algebra. Insbesondere ist diese empfohlen, wenn Sie an einen Master in Mathematik denken. (Für das Lehramt gibt es auch eine Algebra-Vorlesung im Sommersemester.) Es werden unter Anderem folgende Themen behandelt:

- Das weitere Studium von Gruppen, Ringen und Körpern; vor allem nicht-abelsche Gruppen bieten reichhaltiges und interessantes Material; siehe zum Beispiel
  https://de.wikipedia.org/wiki/Endliche_einfache_Gruppe.
- Die Lösbarkeit von Polynomgleichungen der Form $x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = 0$. Für $n = 2$ kann man die Lösungen bestimmen mit Hilfe von "quadratischer Ergänzung" (jedenfalls über $K = \mathbb{C}$). Gibt es solche Lösungsformeln auch für $n = 3, 4, \ldots$?

Im anschließenden Wahlbereich gibt es (aktuell in Stuttgart) folgende Schwerpunkte:

- Darstellungstheorie: Dort wird die hier behandelte Matrix-Theorie verallgemeinert, indem zum Beispiel versucht wird, simultan gemeinsame Normalformen von mehreren Matrizen zu finden (und nicht nur einer einzelnen Matrix wie hier).
- Lie-Algebren und Lie-Gruppen: Diese haben Verbindungen und Anwendungen in der Mathematik selbst, der Geometrie, der Physik ... In der Algebra werden wichtige Klassen von (endlichen) Gruppen mittels Lie-Theorie konstruiert.
- Algebraische Geometrie: Hier werden noch einmal allgemeinere Gleichungssysteme betrachtet, nämlich nicht nur lineare Gleichungen oder Polynomgleichungen in einer Variablen, sondern Systeme von Polynomgleichungen in mehreren Variablen. Zum Beispiel sind die orthogonalen Matrizen der Gruppe $O_n(K)$ (Beispiel 30.12) durch ein solches Gleichungssystem definiert: Ist $A = [a_{ij}]_{1 \leqslant i,j \leqslant n} \in M_n(K)$, so gilt
  $$A \in O_n(K) \qquad \Leftrightarrow \qquad \sum_{k=1}^{n} a_{ki}a_{kj} = \delta_{ij} \quad \text{für } i, j = 1, \ldots, n.$$
  D.h., wir haben ein System von n^2 Polynomgleichungen vom Grad 2 in den n^2 Einträgen a_{ij}. Kann man zum Beispiel wieder so etwas wie die "Dimension" der Lösungsmenge eines solches Systems definieren? Gibt es eine Verallgemeinerung des Gauß–Verfahrens, um die Lösungen algorithmisch zu bestimmen?

Historisch bedeutsam ist hier "Fermat's Letztes Problem", nämlich die Aussage, dass es für $n \geqslant 3$ keine Lösung $(x, y, z) \in \mathbb{Q}^3$ gibt mit $x^n + y^n = z^n$ und $xyz \neq 0$. Dies wurde erst 1993 von Andrew Wiles bewiesen. (Beachte: Für $n = 2$ gibt es unendlich viele Lösungen, nämlich die sogenannten "Pythagoras–Tripel", zum Beispiel $(x, y, z) = (3, 4, 5)$.)
Siehe dazu auch https://en.wikipedia.org/wiki/Fermat%27s_Last_Theorem.

Index