

Lineare Algebra II
Sommersemester 2011

Karl-Hermann Neeb

27.7.2011

Inhalt

5	Eigenvektoren und Eigenwerte	7
5.1	Ähnlichkeit von Matrizen	7
5.2	Eigenvektoren und Eigenwerte	10
5.3	Formale Polynome	15
5.4	Das charakteristische Polynom einer Matrix	19
5.5	Trigonalisierung	23
5.5.1	Trigonalisierung komplexer Matrizen	23
5.5.2	Blocktrigonalisierung reeller Matrizen	25
5.6	Ideale im Polynomring	27
5.7	Der Satz von Cayley–Hamilton	29
6	Die Jordansche Normalform	35
6.1	Jordankästchen	35
6.2	Hauptvektoren und verallgemeinerte Eigenräume	37
6.3	Die Jordan-Zerlegung	41
6.4	Der Beweis des Hauptsatzes	43
6.5	Die reelle Jordansche Normalform	46
6.6	Matrixpotenzen	50
6.7	Die Matrixexponentialfunktion	53
7	Euklidische und unitäre Räume	59
7.1	Unitäre Räume	59
7.2	Orthonormalbasen	65
7.2.1	Das Gram–Schmidt Verfahren	66
7.2.2	Orthogonalprojektionen	70
7.3	Orthogonale und unitäre Abbildungen	77
7.3.1	Die QR-Zerlegung einer Matrix	84
7.4	Normale Matrizen	87
7.4.1	Diagonalisierungssätze für normale Matrizen	89
7.4.2	Normalform orthogonaler Matrizen	92
7.5	Die Adjungierte einer Abbildung unitärer Räume	93
7.6	Singulärwertzerlegung	95
8	Quotientenraum und Dualraum	99
8.1	Der Quotientenraum	99
8.2	Der Dualraum	105
8.2.1	Lineare Funktionale und duale Basen	106
8.2.2	Die duale Abbildung	108
8.2.3	Der Bidualraum	108
8.2.4	Annullatoren	109

9	Bilinearformen und Quadriken	113
9.1	Bilinearformen und Sesquilinearformen	113
9.1.1	Die Matrix einer Sesquilinearform	113
9.1.2	Automorphismen	118
9.1.3	Orthogonalräume	119
9.1.4	Normalformen	120
9.2	Quadratische Formen	123
9.3	Hauptachsentransformation	124
9.4	Typen reeller Quadriken	126
9.4.1	Quadratische Kurven	127
9.4.2	Quadratische Flächen	129
9.4.3	Quadratische Gleichungen mit linearen Termen	133
9.5	Definitheit von Formen und von Matrizen	134
10	Tensorprodukte	141
10.1	Tensorprodukte von Vektorräumen	141
10.1.1	Tensorprodukte von zwei Vektorräumen	141
10.1.2	Tensorprodukte von endlich vielen Vektorräumen	144
10.2	Symmetrische und alternierende Produkte	146
11	Affine und konvexe Geometrie	149
11.1	Affine Geometrie	149
11.2	Konvexe Mengen	157
11.3	Polyeder	160

Vorwort

Dies ist ein partielles Skriptum zur Vorlesung "Lineare Algebra II". Es beruht auf Vorlesungen zur linearen Algebra, die ich früher gehalten habe und die etwas anders strukturiert waren. Wie im vergangenen Semester empfehle ich wieder das Buchmanuskript "Lineare Algebra" von Barth und Knabner als Begleitlektüre zur Vorlesung. Manche Abschnitte der Vorlesung werden sich sehr eng an dieses Buch halten. Als Begleitlektüre zu den anderen ist dieses Skript gedacht.

Teile dieses Skripts wurden schon im Rahmen der Linearen Algebra I behandelt. Ich habe sie trotzdem nicht herausgenommen, um leichter auf die verweisen zu können. Der Abschnitt über formale Polynome ist z.B. in Abschnitt 5.3 integriert.

Der Stoff des zweiten Teils der Linearen Algebra beginnt in Abschnitt 5.5.

Notation und Erinnerung

Matrizen: $M_{p,q}(R) = R^{(p,q)}$ steht für den Vektorraum der $p \times q$ -Matrizen mit Einträgen in dem Ring R . In der Regel wird $R = \mathbb{K}$ ein Körper sein.

Koordinaten: Ist $B = (v_1, \dots, v_n)$ eine Basis von V , so schreiben wir

$$\kappa_B: V \rightarrow \mathbb{K}^n, \quad \kappa_B(v) = [v]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad v = \sum_{j=1}^n x_j v_j$$

für die zugehörige Koordinatenabbildung.

Matrizen linearer Abbildungen: Ist $\varphi: V \rightarrow W$ eine lineare Abbildung, $B = (v_1, \dots, v_n)$ eine Basis von V und $C = (w_1, \dots, w_m)$ eine Basis von W , so schreiben wir $A = [\varphi]_B^C$ für die Matrix von φ bzgl. der geordneten Basen B und C . Sie bestimmt sich durch die Relation

$$\varphi(v_k) = \sum_{j=1}^m a_{jk} w_j.$$

In den Spalten der Matrix A stehen also die Koordinaten der Bilder $\varphi(v_k)$ der Elemente der Basis B bzgl. der Basis C von W . Für $V = W$ und $B = C$ schreiben wir vereinfacht $[\varphi]_B := [\varphi]_B^B$.

Ein wichtiger Spezialfall ist $V = W$ und $\varphi = \text{id}$. Dann stehen in der Matrix $S = [\text{id}_V]_C^B$ die Koordinaten der Elemente $w_1, \dots, w_n$ der Basis C bzgl. der Basis B .

Transformationsformel: Für zwei Basen B, B' von V und C, C' von W haben wir die Transformationsformel

$$[\varphi]_{B'}^{C'} = T^{-1} [\varphi]_B^C S \quad \text{für} \quad T = [\text{id}_W]_{C'}^C, \quad S = [\text{id}_V]_B^{B'}. \quad (4.1)$$

Transposition von Matrizen: Für $A = (a_{ij}) \in M_n(\mathbb{K})$ schreiben wir $A^T = A^t = (a_{ji})$ für die transponierte Matrix.

Für $A = (a_{ij}) \in M_n(\mathbb{C})$ schreiben wir $A^* = A^\dagger = (\overline{a_{ji}}) = \overline{A}^T$ für die hermitesch transponierte (=adjungierte) Matrix.

Kapitel 5

Eigenvektoren und Eigenwerte

In diesem Kapitel sei V ein n -dimensionaler Vektorraum über dem Körper $\mathbb{K}$. Wir betrachten eine lineare Abbildung $\varphi : V \rightarrow V$. Hauptziel der Überlegungen in diesem Kapitel ist es, eine Basis B von V so zu finden, dass die Matrix $[\varphi]_B$ von φ bzgl. B möglichst einfache Gestalt besitzt. Hierbei kann “möglichst einfach” verschiedene Bedeutungen haben, und wir werden mehrere verfolgen. In diesem Kapitel diskutieren wir, in wie weit man φ durch eine Diagonalmatrix darstellen kann. Dies führt uns auf das Konzept des Eigenvektors und des Eigenwerts einer linearen Abbildung bzw. einer Matrix.

Die Theorie, die wir hier präsentieren, ist nicht nur fundamental für zahlreiche mathematische Anwendungen der linearen Algebra, sondern auch von erfahrbarer physikalischer Relevanz, wenn man sie auf gewisse Operatoren auf Räumen stetiger Funktionen anwendet. Ein klassisches Beispiel ist die Theorie der schwingenden Saite. Hier entsprechen Eigenwerte den Resonanzfrequenzen und Eigenvektoren werden als die zugehörigen Schwingungen hörbar. Wir werden dieses physikalische Beispiel hier nicht breiter diskutieren und verweisen für detailliertere Information auf die schöne Darstellung auf den Seiten 260–272 in Egbert Brieskorns Buch “Lineare Algebra und analytische Geometrie II,” (Vieweg, Braunschweig 1985).

Ein wichtiger neuer Punkt in diesem Kapitel ist, dass es uns von der linearen Algebra in die Algebra höherer Ordnung führt. In der Tat werden wir sehen, dass das Auffinden von Eigenwerten auf das Lösen von Polynomgleichungen führt, deren Grad die Dimension des betrachteten Vektorraums ist. Die Theorie der Gleichungen höherer Ordnung ist sehr verschieden von der linearen Theorie, die wir bisher kennengelernt haben. Zum Beispiel hängt die Lösbarkeit einer Polynomgleichung vom Grad ≥ 2 von dem betrachteten Körper ab; die Gleichung $X^2 + 1 = 0$ hat keine reelle Lösung, aber zwei komplexe Lösungen, $\pm i$. Dieses Verhalten steht in starkem Kontrast zu den Systemen linearer Gleichungen, deren Lösbarkeit über *jedem Körper* durch den Gauß–Jordan Algorithmus entschieden werden kann, wenn man ihn über dem kleinsten Körper, der alle Koeffizienten der Matrix enthält, durchführt.

5.1 Ähnlichkeit von Matrizen

In diesem Abschnitt ist V ein n -dimensionaler $\mathbb{K}$ -Vektorraum und $\varphi \in \text{End}(V)$ ist eine lineare Abbildung.

5.1.1. [Motivation] Sei V ein n -dimensionaler Vektorraum über dem Körper $\mathbb{K}$ und $\varphi : V \rightarrow V$ eine lineare Abbildung. Um φ durch eine Matrix A darstellen zu können, wählen wir eine Basis

$$B = (v_1, v_2, \dots, v_n)$$

von V und berechnen die Bilder der Basisvektoren

$$\varphi(v_j) = \sum_{i=1}^n a_{ij} v_i \quad (j = 1, \dots, n).$$

Dann wird φ (bzgl. der Basis B) durch die $(n \times n)$ -Matrix

$$[\varphi]_B = A := (a_{ij})$$

dargestellt. Wir möchten für die Matrix A eine Basis B so finden, dass die Form der Matrix A möglichst einfach wird (dann kann man leichter Berechnungen mit A durchführen).

Bemerkung 5.1.2. Man muss hier beachten, dass das beschriebene Problem ganz wesentlich davon abhängt, dass wir für Endomorphismen φ eines Vektorraums V die Matrix $[\varphi]_B$ bzgl. einer Basis B betrachten und nicht die Matrix $[\varphi]_B^C$ einer linearen Abbildung $\varphi: V \rightarrow W$ bzgl. zweier Basen B von V und C von W .

Würden wir zwei Basen zulassen, so könnten wir Basen B und C finden, so dass die Matrix $A = [\varphi]_B^C$, deren Spalten die Koordinaten der Bilder der Elemente von B bzgl. der Basis C sind, die Gestalt

$$A = [\varphi]_B^C = \begin{pmatrix} \boxed{\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{matrix}} & \begin{matrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \end{matrix} \\ \begin{matrix} 0 & \cdots & \cdots & \cdots \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & \cdots \end{matrix} & \begin{matrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{matrix} \end{pmatrix}$$

hat. In der Tat ist das eine sehr einfache Gestalt, aber bei unserer Problemstellung sind die Spielregeln anders: Wir haben jeweils nur eine Basis $B = C$ zu verwenden. Diese Einschränkung macht das Problem, eine möglichst einfache Matrix $[\varphi]_B$ zu finden, wesentlich schwieriger.

Definition 5.1.3. [Ähnlichkeit von Matrizen] Zwei $n \times n$ -Matrizen A und $A' \in M_n(\mathbb{K})$ heißen *ähnlich*, geschrieben $A \approx A'$, wenn es eine invertierbare $n \times n$ -Matrix $S \in \text{GL}_n(\mathbb{K})$ gibt mit

$$A' = S^{-1}AS. \quad (5.1)$$

Lemma 5.1.4. *Ähnlichkeit von Matrizen ist eine Äquivalenzrelation, d.h. für $A, A', A'' \in M_n(\mathbb{K})$ gilt:*

- (R) $A \approx A$ (Reflexivität).
- (S) $A \approx A' \Rightarrow A' \approx A$ (Symmetrie).
- (T) $A \approx A', A' \approx A'' \Rightarrow A \approx A''$ (Transitivität).

Beweis. (R) Für die Einheitsmatrix $S := \mathbf{1}_n$ gilt $A = \mathbf{1}_n^{-1}A\mathbf{1}_n$, also $A \approx A$.

(S) Gilt $A \approx A'$, so gibt es eine invertierbare Matrix S , so dass $A' = S^{-1}AS$ gilt. Durch Multiplikation mit S von der linken Seite und durch S^{-1} von der rechten Seite erhalten wir $SA'S^{-1} = A$. Wählen wir $T = S^{-1}$, so erhalten wir $A = T^{-1}A'T$ und damit $A' \approx A$.

(T) Es gelte $A \approx A'$ und $A' \approx A''$. Dann gibt es invertierbare Matrizen S und T , so dass $A' = S^{-1}AS$, $A'' = T^{-1}A'T$ gilt. Durch Ersetzung von A' in der zweiten Gleichung durch den in der ersten Gleichung gegebenen Wert erhalten wir $A'' = T^{-1}S^{-1}AST = (ST)^{-1}A(ST)$. Also gilt $A \approx A''$. $\square$

Satz 5.1.5. [Bedeutung der Ähnlichkeit] *Zwei Matrizen A und A' in $M_n(\mathbb{K})$ sind genau dann ähnlich, wenn eine lineare Abbildung $\varphi: V \rightarrow V$ auf einem n -dimensionalen Vektorraum V und zwei Basen B und B' von V existieren, so dass $A = [\varphi]_B$ und $A' = [\varphi]_{B'}$ gilt.*

Beweis. Sind $A = [\varphi]_B$ und $A' = [\varphi]_{B'}$ Matrizen einer linearen Abbildung $\varphi : V \rightarrow V$ bzgl. zweier Basen B und B' in V , so gibt es eine invertierbare Matrix $S = [\text{id}]_{B'}^B$ (die Transformationsmatrix) mit $A' = S^{-1}AS$, d.h. $A \approx A'$.

Umgekehrt sei $A \approx A'$. Sei $V := \mathbb{K}^n$ und $\varphi : \mathbb{K}^n \rightarrow \mathbb{K}^n$ definiert durch $\varphi(v) := Av$. Dann gilt $[\varphi]_B = A$ bzgl. der Standardbasis $B = (e_1, \dots, e_n)$ von $\mathbb{K}^n$. Nach Annahme gilt $A' = S^{-1}AS$ für eine invertierbare Matrix S . Die Spalten $v_1, \dots, v_n$ von S bilden eine Basis B' von $\mathbb{K}^n$ mit $S = [\text{id}]_{B'}^B$. Die Matrix von φ bzgl. der Basis B' ist dann $A' = S^{-1}AS = [\varphi]_{B'}$ aufgrund der Transformationsformel. Also stellen A und A' dieselbe lineare Abbildung bzgl. der Basen B und B' dar. $\square$

Wir können nun das am Beginn dieses Kapitels gestellte Problem umformulieren: Finde für eine gegebene Matrix A eine ähnliche Matrix A' , deren Gestalt möglichst einfach ist. Wann ist eine Matrix "möglichst einfach"? Die einfachsten Matrizen sind Diagonalmatrizen

$$A = \text{diag}(\lambda_1, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Die Matrix von φ bzgl. der Basis B hat genau dann die Diagonalgestalt $\text{diag}(\lambda_1, \dots, \lambda_n)$, wenn gilt:

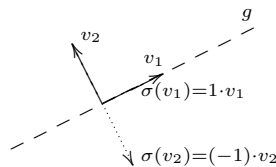
$$\varphi(v_j) = \lambda_j v_j \quad \text{für } j = 1, \dots, n.$$

Dies führt zu folgender **Frage**: *Gibt es eine Basis $v_1, \dots, v_n$ von V , so dass $\varphi(v_j) = \lambda_j v_j$ für geeignete Zahlen $\lambda_j \in \mathbb{K}$ ($j = 1, \dots, n$) gilt und wie können wir solche Vektoren v_j und Zahlen λ_j finden?*

Beispiele 5.1.6. In einigen Fällen können wir eine solche Basis aus geometrischen Überlegungen finden: Sei $V = \mathbb{R}^2$ und g eine Gerade durch 0.

(a) Sei σ die (orthogonale) Spiegelung an der Geraden g . Wir wählen einen Vektor v_1 in Richtung von g und einen Vektor v_2 orthogonal zu g . Dann folgt

$$\sigma(v_1) = 1 \cdot v_1 \quad \text{und} \quad \sigma(v_2) = (-1) \cdot v_2.$$



Die Matrix von σ bzgl. der Basis v_1, v_2 ist $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

(b) Sei π die Orthogonalprojektion auf g . Für die obige Basis v_1, v_2 erhalten wir

$$\sigma(v_1) = 1 \cdot v_1, \quad \sigma(v_2) = 0 = 0 \cdot v_2.$$

Also ergibt sich die Matrix von A bzgl. v_1, v_2 zu $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

(c) Es gibt andere geometrische Abbildungen, die keine Eigenvektoren besitzen. Zum Beispiel die Rotation $\rho_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ um den Nullpunkt um den Winkel $\theta \in]0, \pi[$ besitzt in der Tat keinen Eigenvektor zu einem reellen Eigenwert $\lambda \in \mathbb{R}$.

Wir können die Gleichung $\varphi(v) = \lambda v$ auch auf die folgende Art schreiben:

$$\begin{aligned} \varphi(v) = \lambda v &\iff \varphi(v) - \lambda v = 0 \\ &\iff \varphi(v) - \lambda \text{id}(v) = 0 \\ &\iff (\varphi - \lambda \text{id})(v) = 0 \\ &\iff v \in \ker(\varphi - \lambda \text{id}). \end{aligned}$$

Dabei ist $\text{id} = \text{id}_V$ die identische Abbildung auf V .

Wir suchen zunächst Vektoren $v \in V$, die $\varphi(v) = \lambda v$ für bestimmte Skalare $\lambda \in \mathbb{K}$ erfüllen. Wir wollen geeignete Namen für solche Vektoren v und Skalare λ einführen.

5.2 Eigenvektoren und Eigenwerte

Definition 5.2.1. [Eigenvektoren und Eigenwerte]

(a) Sei $\varphi \in \text{End}(V)$ ein Endomorphismus von V . Eine Zahl $\lambda \in \mathbb{K}$ heißt *Eigenwert* von φ , wenn es einen Vektor $0 \neq v \in V$ gibt, so dass

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

gilt und jeder Vektor $v \neq 0$, der diese Gleichung erfüllt, heißt *Eigenvektor* von φ zum Eigenwert λ .¹

Der Unterraum

$$V_\lambda := V_\lambda(\varphi) := \ker(\varphi - \lambda \text{id}_V) = \{v \in V : \varphi(v) = \lambda v\}$$

heißt *Eigenraum* zum Eigenwert λ . Da der Kern einer linearen Abbildung immer ein linearer Teilraum ist, ist auch V_λ ein Untervektorraum von V . Die Dimension

$$g_\lambda := \dim V_\lambda \geq 1$$

heißt die *geometrische Vielfachheit* des Eigenwertes λ . Eine Zahl $\lambda \in \mathbb{K}$ ist also genau dann Eigenwert, wenn $V_\lambda(\varphi) \neq \{0\}$ ist, also $g_\lambda \geq 1$. In diesem Fall ist $V_\lambda(\varphi) \setminus \{0\}$ die Menge der Eigenvektoren von φ zum Eigenwert λ .

Die lineare Abbildung φ heißt *diagonalisierbar*, wenn V eine Basis aus Eigenvektoren von φ besitzt, also wenn eine Basis B existiert, so dass die zugehörige Matrix $[\varphi]_B$ Diagonalgestalt hat.

(b) Ist $A \in M_n(\mathbb{K})$ eine $(n \times n)$ -Matrix, so betrachten wir die lineare Abbildung

$$L_A \in \text{End}(\mathbb{K}^n), \quad x \mapsto Ax.$$

Eine Zahl $\lambda \in \mathbb{K}$ heißt *Eigenwert* von A , wenn λ Eigenwert von L_A ist, also wenn ein $0 \neq x \in \mathbb{K}^n$ mit

$$Ax = \lambda x$$

existiert. Entsprechend heißt

$$\ker(L_A - \lambda \text{id}) = \{x \in \mathbb{K}^n : Ax = \lambda x\}$$

Eigenraum von A zum Eigenwert λ . Alle Aussagen über Eigenvektoren und Eigenwerte von linearen Abbildungen lassen sich in diesem Sinn auf Matrizen übertragen.

Die Matrix A heißt *diagonalisierbar*, wenn L_A diagonalisierbar ist, also wenn in $\mathbb{K}^n$ eine Basis aus Eigenvektoren von A existiert.

Mit dieser Terminologie können wir die Frage in 5.1.1 anders formulieren: Wann ist die lineare Abbildung $\varphi : V \rightarrow V$ diagonalisierbar?

Bemerkung 5.2.2. Wir fassen noch einmal die wichtigsten Aspekte zusammen.

λ ist ein Eigenwert von φ

$\iff$ Die Gleichung $(\varphi - \lambda \text{id})(v) = 0$ hat eine Lösung $v \neq 0$

$\iff \ker(\varphi - \lambda \text{id}) \neq \{0\}$

¹Darüber, ob man 0 in der Definition eines Eigenvektors mit einschliessen sollte, scheiden sich die Geister. In den Lehrbüchern von S. Lang und N. Bourbaki ist 0 ein Eigenvektor für jeden Eigenwert λ . Eine Zahl λ heißt dann Eigenwert, wenn der zugehörige Eigenraum positive Dimension besitzt. Im Grunde entspricht dies auch meiner Sichtweise, aber die meisten deutschsprachigen Bücher zur Linearen Algebra sehen das anders. Um in dieser nebensächlichen Frage keine unnötige Verwirrung zu erzeugen, schliessen wir uns der deutschsprachigen LA-Literatur an.

Analog werden Eigenvektoren charakterisiert:

$$\begin{aligned} & v \text{ ist ein Eigenvektor von } \varphi \text{ bzgl. } \lambda \\ \iff & v \text{ ist eine nichtverschwindende Lösung von } (\varphi - \lambda \text{id})v = 0 \\ \iff & 0 \neq v \in \ker(\varphi - \lambda \text{id}) \end{aligned}$$

Bemerkung 5.2.3. Für $A \in M_n(\mathbb{K})$ ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von A , wenn das homogene System linearer Gleichungen

$$(A - \lambda \mathbf{1})v = 0 \tag{H}$$

eine nichttriviale Lösung $v \neq 0$ besitzt. Die nichttrivialen Lösungen von (H) sind dann die Eigenvektoren zum Eigenwert λ .

In der Tat, können wir die Gleichung $Av = \lambda v$ auch auf folgende Art schreiben:

$$\begin{aligned} Av = \lambda v & \iff Av - \lambda v = 0 \\ & \iff Av - \lambda v = 0 \\ & \iff (A - \lambda \mathbf{1})v = 0 \end{aligned}$$

wobei $\mathbf{1} = \mathbf{1}_n$ die $n \times n$ -Einheitsmatrix bezeichnet.

Wir legen uns nun die Frage vor, wie wir Eigenwerte und Eigenvektoren einer linearen Abbildung bzw. einer Matrix berechnen. Der folgende Satz ist hierbei von zentraler Bedeutung.

Satz 5.2.4. Ist $\varphi \in \text{End}(V)$ und $\dim V < \infty$, so ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von φ , wenn $\det(\varphi - \lambda \text{id}) = 0$ ist.

Beweis. Da wir einen endlichdimensionalen Vektorraum V betrachten, gilt:

$$\begin{aligned} \ker(\varphi - \lambda \text{id}) = \{0\} & \iff \varphi - \lambda \text{id} \text{ ist injektiv} \\ & \iff \varphi - \lambda \text{id} \text{ ist bijektiv} \\ & \iff \det(\varphi - \lambda \text{id}) \neq 0 \end{aligned} \quad \square$$

Folgerung 5.2.5. Für $A \in M_n(\mathbb{K})$ gilt:

$$\begin{aligned} \lambda \text{ ist Eigenwert von } A & \iff A - \lambda \mathbf{1} \text{ ist nicht invertierbar} \\ & \iff \boxed{\det(A - \lambda \mathbf{1}) = 0} \end{aligned}$$

Die letzte Formel heißt *charakteristische Gleichung* von A .

Beweis. Wir wissen, dass die Matrix $A - \lambda \mathbf{1}$ genau dann nicht invertierbar ist, wenn $\det(A - \lambda \mathbf{1}) = 0$ gilt. Wir können nun Satz 5.2.4 auf die lineare Abbildung $L_A: \mathbb{K}^n \rightarrow \mathbb{K}^n, x \mapsto Ax$ anwenden. Die Behauptung folgt dann aus

$$\det(A - \lambda \mathbf{1}) = \det(L_A - \lambda \text{id}),$$

denn $A - \lambda \mathbf{1}$ ist die Matrix von $L_A - \lambda \text{id}$ bzgl. der Standardbasis von $\mathbb{K}^n$. $\square$

Beispiele 5.2.6. (a) Wie in Beispiel 5.1.6(a) betrachten wir die orthogonale Spiegelung σ an der Geraden g :

Eigenwerte	$\lambda_1 = 1$	$\lambda_2 = -1$
Eigenräume	$V_{\lambda_1} = \mathbb{R}v_1$	$V_{\lambda_2} = \mathbb{R}v_2$
	alle Vektoren auf g	alle Vektoren senkrecht zu g

(b) Für die Orthogonalprojektion π auf g aus Beispiel 5.1.6(b) gilt:

$$\left. \begin{array}{l} \text{Eigenwerte} \\ \text{Eigenvektoren} \\ \text{Eigenräume} \end{array} \right\} \begin{array}{ll} \lambda_1 = 1 & \lambda_2 = 0 \\ & \text{wie oben} \end{array}$$

In beiden Beispielen erhalten wir zwei Eigenwerte mit der geometrischen Vielfachheit 1.

(c) Wir betrachten die Matrix

$$A := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R}).$$

Für $\lambda \in \mathbb{R}$ erhalten wir die Matrizen

$$A - \lambda \mathbf{1} = \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix}$$

mit

$$\det(A - \lambda \mathbf{1}) = \lambda^2 + 1 \neq 0.$$

Also sind alle Matrizen $A - \lambda \mathbf{1}$ invertierbar, denn die charakteristische Gleichung von A besitzt keine Nullstelle. Folglich hat die reelle Matrix A keinen Eigenwert.

Wir sehen insbesondere, dass nicht jede Matrix einen Eigenwert besitzt.

(d) Wie betrachten die komplexe Matrix

$$A := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{C}).$$

Ihre charakteristische Gleichung

$$0 = \det(A - \lambda \mathbf{1}) = \lambda^2 + 1$$

besitzt dann die Nullstellen $\lambda_1 = i$ und $\lambda_2 = -i$, die zueinander komplex konjugiert sind. Lösen der linearen Gleichungssysteme

$$(A - \lambda_j \mathbf{1})v = 0$$

führt auf die Eigenvektoren

$$v_1 = \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

Beispiel 5.2.7. (für $\mathbb{K} = \mathbb{R}$): Wir untersuchen die reelle Matrix $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ auf Eigenvektoren

und Eigenwerte. Dann ist $A - \lambda \mathbf{1} = \begin{pmatrix} 1 - \lambda & 3 \\ 3 & 1 - \lambda \end{pmatrix}$.

Eigenwerte: Die charakteristische Gleichung ist also

$$0 = \det(A - \lambda \mathbf{1}) = \begin{vmatrix} 1 - \lambda & 3 \\ 3 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - 9$$

und damit $\lambda_{1/2} = 1 \pm 3$, d.h.

$$\lambda_1 = 4, \quad \lambda_2 = -2.$$

Eigenvektoren: Wir müssen beide Eigenwerte berücksichtigen.

- $\lambda_1 = 4$:

$$(A - \lambda_1 \mathbf{1})v = \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix} v = 0 \quad \implies \quad v \in \mathbb{R} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- $\lambda_2 = -2$:

$$(A - \lambda_2 \mathbf{1})v = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix} v = 0 \quad \implies \quad v \in \mathbb{R} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Also sind $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ Eigenvektoren von A . Diese bilden eine Basis B' von $\mathbb{R}^2$. Also ist A diagonalisierbar.

Bemerkung: Die lineare Abbildung $\varphi_A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, die durch die Matrix A bzgl. der Standardbasis B gegeben ist, besitzt die Matrix

$$A' = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$$

bzgl. der Basis $B' = (v_1, v_2)$. In der Tat ist $Av_1 = 4v_1$ und $Av_2 = -2v_2$.

Die Matrizen A und A' sind durch die Transformationsformel

$$A' = S^{-1}AS$$

miteinander verbunden. Dabei ist S die Transformationsmatrix $S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Die Spalten der Transformationsmatrix S sind die Koordinaten der neuen Basisvektoren v_1, v_2 bzgl. der Standardbasis (e_1, e_2) .

Bemerkung 5.2.8. Eine Matrix $A \in M_n(\mathbb{K})$ ist genau dann diagonalisierbar, wenn es in $\mathbb{K}^n$ eine Basis $v_1, v_2, \dots, v_n$ aus Eigenvektoren zu Eigenwerten $\lambda_1, \lambda_2, \dots, \lambda_n$ von A gibt. Ist S die Transformationsmatrix mit den Eigenvektoren $v_1, \dots, v_n$ als Spalten, so gilt

$$S^{-1}AS = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Folgerung 5.2.9. Eine Matrix A ist genau dann diagonalisierbar, wenn sie einer Diagonalmatrix A' ähnlich ist, d.h. wenn es eine invertierbare Matrix S gibt, so dass $A' := S^{-1}AS$ eine Diagonalmatrix ist.

Beispiele 5.2.10. $[\mathbb{K} = \mathbb{R}, n = 2]$

(a) Die Matrix $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \in M_2(\mathbb{R})$ aus Beispiel 5.2.7 ist diagonalisierbar: Für die Eigenwerte

$\lambda_1 = 4$ und $\lambda_2 = -2$ haben wir die Eigenvektoren: $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, die eine Basis von $\mathbb{R}^2$ bilden. Wir erhalten

$$S^{-1}AS = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix} \quad \text{mit} \quad S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

wobei v_1, v_2 die Spalten der Matrix S bilden.

(b) Die Matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in M_2(\mathbb{C})$ aus Beispiel 5.2.6 hat keine reellen Eigenwerte. Sie kann also über dem Körper der reellen Zahlen nicht diagonalisiert werden. A hat aber komplexe $\lambda_1 = 1 + i$ und $\lambda_2 = 1 - i$ mit den Eigenvektoren

$$v_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Also erhalten wir

$$S^{-1}AS = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix} \quad \text{mit} \quad S = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}.$$

Satz 5.2.11. Die Matrix $A \in M_n(\mathbb{K})$ habe den Eigenvektor $v \in \mathbb{K}^n$ zu dem Eigenwert λ .

(i) v ist auch Eigenvektor von

- αA , $\alpha \in \mathbb{K}$, zum Eigenwert $\alpha\lambda$.
- $\alpha \mathbf{1}_n + A$, $\alpha \in \mathbb{K}$, zum Eigenwert $\alpha + \lambda$.
- A^n , $n \in \mathbb{N}$, zum Eigenwert λ^n .
- A^{-1} , falls invertierbar ist, zum Eigenwert λ^{-1} .

(ii) Für $\mathbb{K} = \mathbb{C}$ ist $\bar{v} \in \mathbb{C}^n$ Eigenvektor zum Eigenwert $\bar{\lambda}$ für die Matrix $\bar{A}$. Insbesondere treten nicht reelle Eigenwerte reeller Matrizen immer in konjugiert komplexen Paaren auf.

Beweis. (i) folgt sofort aus der Eigenwertgleichung $Av = \lambda v$ durch evidente Umformungen.

(ii) Aus $Av = \lambda v$ folgt sofort $\bar{A}\bar{v} = \bar{\lambda}\bar{v}$ durch komplexe Konjugation der Einträge. $\square$

Beispiel 5.2.12. Sei $P: \mathbb{K}^n \rightarrow \mathbb{K}^n$ eine Projektion. Dann ist $P^2 = P$ und aus Satz 5.2.11 folgt sofort für jeden Eigenwert λ von P die Relation $\lambda^2 = \lambda$, also $\lambda \in \{0, 1\}$. Der Eigenraum zum Eigenwert 1 ist das Bild $\text{im}(P) = \ker(\mathbf{1} - P)$ und $\ker(P)$ ist der Eigenraum zum Eigenwert 0. Aus $\mathbb{K}^n = \ker(P) \oplus \text{im}(P)$ folgt daher die Diagonalisierbarkeit von P .

Lineare Unabhängigkeit von Eigenvektoren

Satz 5.2.13. Seien $\lambda_1, \lambda_2, \dots, \lambda_r$ paarweise verschiedene Eigenwerte von $\varphi \in \text{End}(V)$ und $v_1, \dots, v_r$ zugehörige Eigenvektoren. Dann sind $v_1, \dots, v_r$ linear unabhängig.

Beweis. Mittels Induktion nach k zeigen wir, dass $v_1, \dots, v_k$ linear unabhängig sind.

Für $k = 1$ ist nichts zu zeigen. Wir nehmen nun an, dass die Vektoren $v_1, \dots, v_{k-1}$ linear unabhängig sind. Sei

$$\mu_1 v_1 + \dots + \mu_k v_k = 0.$$

Wenden wir φ auf diese Gleichung an, so erhalten wir

$$0 = \varphi(\mu_1 v_1 + \dots + \mu_k v_k) = \mu_1 \varphi(v_1) + \dots + \mu_k \varphi(v_k) = \mu_1 \lambda_1 v_1 + \dots + \mu_k \lambda_k v_k.$$

Ziehen wir das λ_k -fache der ersten Gleichung von der zweiten ab, so ergibt sich

$$0 = \mu_1(\lambda_1 - \lambda_k)v_1 + \dots + \mu_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1},$$

und aus der linearen Unabhängigkeit der Vektoren $v_1, \dots, v_{k-1}$ erhalten wir

$$\mu_j(\lambda_j - \lambda_k) = 0, \quad j = 1, \dots, k-1.$$

Wegen $\lambda_j \neq \lambda_k$ für $j < k$ folgt hieraus $\mu_j = 0$. Damit ist auch $\mu_k v_k = 0$, und somit $\mu_k = 0$, da v_k als Eigenvektor von 0 verschieden ist. $\square$

Folgerung 5.2.14. Seien $\lambda_1, \lambda_2, \dots, \lambda_r$ paarweise verschiedene Eigenwerte von φ . Dann ist die Summe der zugehörigen Eigenräume direkt:

$$V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_r} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}.$$

Beweis. Sei

$$u_1 + \dots + u_{r-1} + u_r = 0$$

für $u_j \in V_{\lambda_j}$. Wir müssen zeigen, dass $u_j = 0$ für alle $j = 1, \dots, r$ gilt. Da die u_j entweder 0 oder Eigenvektoren zum Eigenwert λ_j sind, sind die von 0 verschiedenen Summanden nach Satz 5.2.13 linear unabhängig. Da ihre Summe 0 ergibt, müssen also alle Summanden verschwinden. $\square$

Bemerkung 5.2.15. Ein Endomorphismus φ eines n -dimensionalen Vektorraumes V kann höchstens n verschiedene Eigenwerte besitzen. Hat nämlich φ verschiedene Eigenwerte $\lambda_1, \dots, \lambda_r$ mit zugehörigen Eigenvektoren $v_1, \dots, v_r$, so sind diese nach Satz 5.2.13 linear unabhängig. Also gilt $r \leq n$.

Satz 5.2.16. Hat $A \in M_n(\mathbb{K})$ n verschiedene Eigenwerte, dann ist A diagonalisierbar.

Beweis. Seien $\lambda_1, \dots, \lambda_n$ verschiedene Eigenwerte von A und $v_1, \dots, v_n$ zugehörige Eigenvektoren. Diese sind linear unabhängig (Satz 5.2.13) und bilden daher eine Basis von $\mathbb{K}^n$. Also ist A diagonalisierbar. $\square$

Bemerkung 5.2.17. Seien $\lambda_1, \dots, \lambda_r$ verschiedene Eigenwerte von φ und $g_1, \dots, g_r$ deren geometrische Vielfachheiten. Wir finden dann jeweils eine g_i -elementige Basis von V_{λ_i} . Nehmen wir diese Basen zusammen, so erhalten wir eine Basis des Teilraums $V_{\lambda_1} \oplus \dots \oplus V_{\lambda_r}$. Gilt $g_1 + g_2 + \dots + g_r = n$, dann ist dieser Teilraum gleich V . Also hat V eine Basis von Eigenvektoren von φ und die Matrix von φ bzgl. dieser Basis ist eine Diagonalmatrix:

$$\begin{matrix} d_1 \{ \\ d_r \{ \end{matrix} \left(\begin{array}{ccc|ccc|ccc} \lambda_1 & & 0 & & & & & & \\ & \ddots & & & & & & & \\ 0 & & \lambda_1 & & & & & & \\ & & & \ddots & & & & & \\ & & & & \ddots & & & & \\ & & 0 & & & \lambda_r & 0 & & \\ & & & & & & \ddots & & \\ & & & & & 0 & & \lambda_r & \end{array} \right)$$

Bemerkung 5.2.18. Die obigen Eigenschaften gelten entsprechend für Matrizen $A \in M_n(\mathbb{K})$, da man sie durch Anwendung auf die lineare Abbildung $L_A: \mathbb{K}^n \rightarrow \mathbb{K}^n, x \mapsto Ax$ übertragen kann: Seien $\lambda_1, \lambda_2, \dots, \lambda_r$ paarweise verschiedene Eigenwerte einer $n \times n$ -Matrix A und seien $g_1, g_2, \dots, g_r$ ihre geometrischen Vielfachheiten.

a) Die Summe der Eigenräume von A ist direkt:

$$V_{\lambda_1} + V_{\lambda_2} + \dots + V_{\lambda_r} = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \dots \oplus V_{\lambda_r}.$$

b) Sind $v_1, v_2, \dots, v_r$ Eigenvektoren zu den paarweise verschiedenen Eigenwerten $\lambda_1, \lambda_2, \dots, \lambda_r$, so sind diese Eigenvektoren linear unabhängig. Hieraus folgern wir insbesondere $r \leq n$, d.h. eine $n \times n$ -Matrix A hat höchstens n verschiedene Eigenwerte.

Aufgabe 5.2.19. Sei $A \in M_n(\mathbb{C})$ eine Matrix mit reellen Einträgen. Zeige: Ist $v \in \mathbb{C}^n$ ein Eigenvektor von A zum Eigenwert $\lambda \in \mathbb{C}$, so ist $\bar{v}$ ein Eigenvektor von A zum Eigenwert $\bar{\lambda}$. Hierbei steht $\bar{v} \in \mathbb{C}^n$ für den Vektor, der durch Konjugieren der Einträge von v entsteht.

5.3 Formale Polynome

In diesem Abschnitt lernen wir den Begriff eines formalen Polynoms mit Koeffizienten in einem Körper $\mathbb{K}$ kennen. Dies sind Ausdrücke der Gestalt

$$p(X) = a_0 + a_1X + \dots + a_nX^n,$$

wobei wir X nicht als Element von $\mathbb{K}$ interpretieren, sondern als eine *Unbestimmte*. In diesem Sinn wird p keine Funktion $\mathbb{K} \rightarrow \mathbb{K}$ sein, sondern eine formale Linearkombination der Symbole X^n , $n \in \mathbb{N}_0$.

Definition 5.3.1. Sei $\mathbb{K}$ ein Körper.

(a) Wir schreiben $\mathbb{K}[X]$ für die Menge der Ausdrücke der Gestalt

$$p(X) = \sum_{n=0}^N a_n X^n = a_0 + a_1 X + \dots + a_N X^N, \quad a_0, \dots, a_N \in \mathbb{K}, N \in \mathbb{N}_0,$$

die wir *formale Polynome in der Unbestimmten X* nennen. Wir betrachten zwei solche Ausdrücke

$$\sum_{n=0}^N a_n X^n = a_0 + a_1 X + \dots + a_N X^N \quad \text{und} \quad \sum_{n=0}^M b_n X^n = b_0 + b_1 X + \dots + b_M X^M$$

für $N \leq M$ als gleich, wenn

$$a_0 = b_0, \dots, a_N = b_N, \quad b_{N+1} = \dots = b_M = 0,$$

gilt (Koeffizientenvergleich). Jedes Element von $\mathbb{K}[X]$ wird also eindeutig durch eine Folge $(a_n)_{n \in \mathbb{N}_0}$ repräsentiert, bei der alle bis auf endlich viele Glieder 0 sind.

(b) Durch

$$\sum_{n=0}^N a_n X^n + \sum_{n=0}^N b_n X^n := \sum_{n=0}^N (a_n + b_n) X^n$$

und

$$\lambda \sum_{n=0}^N a_n X^n := \sum_{n=0}^N \lambda a_n X^n$$

wird auf $\mathbb{K}[X]$ die Struktur eines K -Vektorraums definiert. Aus dieser Definition ergibt sich sofort, dass die *formalen Monome* $(X^n)_{n \in \mathbb{N}_0}$ eine Basis des K -Vektorraums $\mathbb{K}[X]$ bilden.

Darüber hinaus definieren wir eine *Multiplikation*

$$\begin{aligned} & (a_0 + a_1 X + \dots + a_N X^N) \cdot (b_0 + a_1 X + \dots + b_M X^M) \\ & := a_0 b_0 + (a_0 b_1 + a_1 b_0) X + \dots + \left(\sum_{i+j=n} a_i b_j \right) X^n + \dots + a_N b_M X^{N+M}. \end{aligned}$$

(c) Ist $p(X) = a_0 + \dots + a_N X^N \in \mathbb{K}[X]$ mit $a_N \neq 0$, so nennen wir $\deg(p(X)) := N$ den *Grad von $p(X)$* . Die Zahl a_N heißt *Leitkoeffizient von $p(X)$* . Für das Nullpolynom setzen wir $\deg(0) := -\infty$. Hiermit ist insbesondere $\deg(X^n) = n$ für alle $n \in \mathbb{N}_0$.

Für Elemente von $\mathbb{K}[X]$ werden wir oft $p(X)$ schreiben, um auf die Unbestimmte X hinzuweisen. Anders als für Funktionen steht $p(X)$ also nicht für “den Wert von p an der Stelle X ”, sondern für das Polynom selbst.

Satz 5.3.2. *Der Raum $\mathbb{K}[X]$ der Polynome in einer Unbestimmten X mit Koeffizienten in $\mathbb{K}$ ist ein kommutativer Ring mit Einselement $1 = X^0$. Für $p(X), q(X) \in \mathbb{K}[X]$ gilt*

$$\deg(p(X)q(X)) = \deg(p(X)) + \deg(q(X)). \quad (5.2)$$

Beweis. Der Beweis durch Nachrechnen ist eine leichte Übung. □

Bemerkung 5.3.3. (a) Aus (5.2) folgt insbesondere, dass das Produkt zweier Polynome $p(X), q(X) \neq 0$ auch von 0 verschieden ist. Man sagt auch, $\mathbb{K}[X]$ sei nullteilerfrei.

Insbesondere folgt hieraus auch, dass Faktorisierungen eindeutig sind: Aus $p(X) = f(X)q(X) = \tilde{f}(X)q(X)$ folgt $f(X) = \tilde{f}(X)$ wenn $q(X) \neq 0$ ist, denn $(f(X) - \tilde{f}(X))q(X) = 0$.

Bemerkung 5.3.4. Man kann jedem formalen Polynom

$$p(X) = a_0 + a_1 X + \dots + a_n X^n \in \mathbb{K}[X]$$

die Polynomfunktion

$$\tilde{p}: \mathbb{K} \rightarrow \mathbb{K}, \quad x \mapsto a_0 + a_1 x + \dots + a_n x^n$$

im Raum $\mathbb{K}[x]$ der Polynomfunktionen $\mathbb{K} \rightarrow \mathbb{K}$ zuordnen. Die Abbildung

$$\Phi: \mathbb{K}[X] \rightarrow \mathbb{K}[x], \quad p \mapsto \tilde{p}$$

ist im allgemeinen aber nicht injektiv, denn verschiedene formale Polynome können die gleiche Polynomfunktion beschreiben.

Ist zum Beispiel $\mathbb{K} = \{0, 1\}$ der zweielementige Körper, so gilt $x^2 = x$ für beide Elemente $0, 1 \in \mathbb{K}$. Insbesondere ist also $\Phi(X^n) = \Phi(X)$ für alle $n \geq 1$. Zum Beispiel stellen die Polynome

$$p_1(X) = X \quad \text{und} \quad p_2(X) = X^2$$

die gleiche Polynomfunktion $\mathbb{K} \rightarrow \mathbb{K}$ dar.

Eine genauere Analyse der Situation zeigt, dass die lineare Abbildung Φ genau dann injektiv ist, wenn der Körper $\mathbb{K}$ unendlich ist, denn jedes Element $q(X) \in \ker \Phi$ ist ein Polynom, dessen zugehörige Funktion $q: \mathbb{K} \rightarrow \mathbb{K}$ verschwindet. Wie wir unten sehen werden, hat ein von Null verschiedenes Polynom vom Grad n maximal n verschiedene Nullstellen (Satz 5.3.7), was für unendliche Körper $\mathbb{K}$ zeigt, dass Φ injektiv ist. Interessiert man sich also nur für unendliche Körper wie $\mathbb{Q}$, $\mathbb{R}$ oder $\mathbb{C}$, so benötigt man das Konzept eines formalen Polynoms in der Linearen Algebra nicht, da man in diesem Fall nicht zwischen den Räumen $\mathbb{K}[x]$ und $\mathbb{K}[X]$ unterscheiden muss.

Definition 5.3.5. [Nullstellen von Polynomen] Ist $p(X) = a_0 + \dots + a_N X^N \in \mathbb{K}[X]$, so heißt $\lambda \in \mathbb{K}$ *Nullstelle* oder *Wurzel* von $p(X)$, wenn

$$p(\lambda) := a_0 + a_1 \lambda + \dots + a_N \lambda^N = 0$$

in $\mathbb{K}$ gilt. Wir ersetzen hierbei in $p(X)$ die Unbestimmte X durch das Element $\lambda \in \mathbb{K}$.

Lemma 5.3.6. Ist $p(X)$ ein Polynom vom Grad $d \geq 1$ und λ eine Nullstelle von $p(X)$, so existiert ein eindeutig bestimmtes Polynom $q(X)$ vom Grad $d - 1$ mit

$$p(X) = (X - \lambda)q(X).$$

Beweis. Zunächst führt die binomische Formel auf

$$X^n = (X - \lambda + \lambda)^n = \sum_{k=0}^n \binom{n}{k} \lambda^{n-k} (X - \lambda)^k,$$

so dass wir $p(X)$ auch schreiben können als

$$\begin{aligned} p(X) &= \sum_{n=0}^d a_n X^n = \sum_{n=0}^d \sum_{k=0}^n a_n \binom{n}{k} \lambda^{n-k} (X - \lambda)^k \\ &= \sum_{k=0}^d \underbrace{\left(\sum_{n=k}^d a_n \binom{n}{k} \lambda^{n-k} \right)}_{=: b_k} (X - \lambda)^k = \sum_{k=0}^d b_k (X - \lambda)^k, \end{aligned}$$

für gewisse $b_0, \dots, b_d \in K$. Wegen $0 = p(\lambda) = b_0$ erhalten wir die Faktorisierung

$$p(X) = (X - \lambda) \underbrace{\sum_{k=1}^d b_k (X - \lambda)^{k-1}}_{=: q(X)}$$

mit $\deg(q(X)) = d - 1$. □

Satz 5.3.7. Ein Polynom $0 \neq p(X) \in \mathbb{K}[X]$ vom Grad n besitzt höchstens n verschiedene Nullstellen.

Beweis. Wir zeigen die Behauptung durch Induktion nach dem Grad n .

Induktionsanfang: Ist $n = 0$, so ist $p(X) = a_0 \neq 0$. Also hat $p(X)$ keine Nullstellen.

Induktionsschritt: Wir nehmen an, dass die Behauptung für $n - 1$ richtig ist. Ist $\lambda \in \mathbb{K}$ eine Nullstelle von $p(X)$, so erhalten wir eine Faktorisierung $p(X) = (X - \lambda)q(X)$ mit $\deg(q(X)) = n - 1$. Nach Induktionsannahme hat $q(X)$ maximal $n - 1$ Nullstellen, und da $x - \lambda \neq 0$ für alle $x \neq \lambda$ in $\mathbb{K}$ gilt, hat $p(X)$ daher maximal n Nullstellen. $\square$

Definition 5.3.8. Sei $p(X)$ ein Polynom vom Grad d und $\lambda \in \mathbb{K}$ eine Nullstelle. Wir schreiben

$$p(X) = (X - \lambda)p_1(X)$$

für ein $p_1(X) \in \mathbb{K}[X]$ vom Grad $d - 1$. Ist $p_1(\lambda) \neq 0$, so heißt λ *einfache Nullstelle* von $p(X)$. Ist $p_1(\lambda) = 0$, so finden wir ein Polynom $p_2(X) \in \mathbb{K}[X]$ mit

$$p_1(X) = (X - \lambda)p_2(X),$$

also

$$p(X) = (X - \lambda)^2 p_2(X).$$

Iteration dieses Prozesses liefert eine Zahl $s > 0$ und ein Polynom $p_s(X) \in \mathbb{K}[X]$ mit

$$p(X) = (X - \lambda)^s p_s(X) \quad \text{und} \quad p_s(\lambda) \neq 0.$$

Wir nennen λ dann eine *Nullstelle der Vielfachheit s* .

Bemerkung 5.3.9. Es seien $\lambda_1, \dots, \lambda_r$ die verschiedenen Nullstellen des Polynoms $p(X) \in \mathbb{K}[X]$ mit den Vielfachheiten $a_1, \dots, a_r$. Dann haben wir

$$p(X) = (X - \lambda_1)^{a_1} p_1(X)$$

und die Zahlen $\lambda_2, \dots, \lambda_r$ sind keine Nullstellen von $(X - \lambda_1)^{a_1}$, also Nullstellen von p_1 . Für $i > 1$ sei f_i die Vielfachheit der Nullstelle λ_i von p_1 . Dann ist $p_2(X) = (X - \lambda_i)^{f_i} s(X)$ mit $s(\lambda_i) \neq 0$, also

$$p(X) = (X - \lambda_i)^{f_i} r(X),$$

wobei $r(X) = (X - \lambda_1)^{a_1} s(X)$ nicht in λ_i verschwindet. Also ist $a_i = f_i$. Induktiv erhalten wir so eine Faktorisierung

$$p(X) = (X - \lambda_1)^{a_1} \cdots (X - \lambda_r)^{a_r} q(X),$$

wobei $q(X)$ keine Nullstellen in $\mathbb{K}$ besitzt.

Theorem 5.3.10. (Fundamentalsatz der Algebra) *Jedes komplexe Polynom in $\mathbb{C}[X]$ ohne Nullstellen ist konstant und jedes Polynom $p(X) \in \mathbb{C}[X]$ besitzt eine Zerlegung in Linearfaktoren:*

$$p(X) = c(X - \lambda_1)^{a_1} \cdots (X - \lambda_r)^{a_r}$$

mit einer Konstanten $c \in \mathbb{C}$.

Der Beweis dieses Satzes lässt sich nicht mit rein algebraischen Hilfsmitteln erbringen. Er benötigt Methoden aus der Analysis und wird in der Vorlesung Funktionentheorie erbracht.

Bemerkung 5.3.11. Auch wenn man weiß, dass komplexe, nicht konstante Polynome immer Nullstellen besitzen, kann das Berechnen von Nullstellen konkreter Polynome sehr schwierig sein. Für Polynome vom Grad 1 ist es trivial, und für Polynome vom Grad 2, 3 und 4 gibt es direkte Lösungsformeln, die quadratische, kubische und quartische Wurzeln verwenden.

Definition 5.3.12. Einen Körper $\mathbb{K}$ mit der Eigenschaft, dass jedes Polynom $p(X) \in \mathbb{K}[X]$ in Linearfaktoren zerfällt, nennt man *algebraisch abgeschlossen*. Wegen Lemma 5.3.6 ist dies äquivalent dazu, dass jedes Polynom $p(X)$ mit $\deg(p(X)) \geq 1$ eine Nullstelle besitzt.

Der Fundamentalsatz der Algebra besagt also, dass der Körper $\mathbb{C}$ der komplexen Zahlen algebraisch abgeschlossen ist. Die Körper $\mathbb{R}$ und $\mathbb{Q}$ sind dagegen nicht algebraisch abgeschlossen, da das Polynom $X^2 + 1$ keine Nullstellen besitzt.

5.4 Das charakteristische Polynom einer Matrix

Definition 5.4.1. [Das charakteristische Polynom] Für eine $n \times n$ -Matrix $A \in M_n(\mathbb{K}) = \mathbb{K}^{(n,n)}$ betrachten wir den Ausdruck $A - X\mathbf{1}_n$ als Element von $M_n(\mathbb{K}[X])$, also als eine Matrix mit Einträgen in dem Ring $\mathbb{K}[X]$. Die Determinante $\det(A - X\mathbf{1}_n)$ dieser Matrix ist nun ein Element von $\mathbb{K}[X]$:

$$\chi_A(X) := \det(A - X\mathbf{1}_n) = \begin{vmatrix} a_{11} - X & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - X & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - X \end{vmatrix} \in \mathbb{K}[X].$$

Dieses Polynome heißt *charakteristisches Polynom* der Matrix A .

Beispiel 5.4.2. Für $A \in M_2(\mathbb{K})$ ist

$$\begin{aligned} \chi_A(X) &= \begin{vmatrix} a_{11} - X & a_{12} \\ a_{21} & a_{22} - X \end{vmatrix} = (a_{11} - X)(a_{22} - X) - a_{12}a_{21} \\ &= X^2 - (a_{11} + a_{22})X + a_{11}a_{22} - a_{12}a_{21} \\ &= X^2 - \operatorname{tr}(A)X + \det(A), \end{aligned}$$

wobei $\operatorname{tr}(A) = a_{11} + a_{22}$ die Spur der Matrix A ist.

Allgemein gilt:

Satz 5.4.3. Das charakteristische Polynom $\chi_A(X)$ von A ist ein Polynom vom Grad n mit Koeffizienten aus $\mathbb{K}$. Weiter gilt

$$\chi_A(X) = (-1)^n X^n + (-1)^{n-1}(\operatorname{tr} A)X^{n-1} + \dots + \det(A),$$

wobei $\operatorname{tr}(A) = \sum_{j=1}^n a_{jj}$ die Spur der Matrix A heißt.

Beweis. Berechnen wir $\det(A - X\mathbf{1})$ mittels der Leibniz-Formel so erhalten wir

$$\begin{aligned} \chi_A(X) &= (a_{11} - X)(a_{22} - X) \dots (a_{nn} - X) + \text{Produkte mit höchstens } n-2 \\ &\quad \text{Faktoren } (a_{ii} - X) \\ &= (-1)^n X^n + (-1)^{n-1}(a_{11} + \dots + a_{nn})X^{n-1} + \text{Terme niedrigeren Grades} \\ &= (-1)^n X^n + (-1)^{n-1}(\operatorname{tr} A)X^{n-1} + \text{Terme niedrigeren Grades} \end{aligned}$$

Weiter haben wir für den konstanten Term $\chi_A(0) = \det(A)$. □

Bemerkung 5.4.4. Nach Folgerung 5.2.5 sind die Eigenwerte von A genau die Nullstellen λ des charakteristischen Polynoms $\chi_A(X)$. Da ein Polynom vom Grad n höchstens n verschiedene Nullstellen hat (Satz 5.3.7), erhalten wir damit einen neuen Beweis für die Aussage in Bemerkung 5.2.18(b), dass eine $n \times n$ -Matrix höchstens n verschiedene Eigenwerte besitzt.

Definition 5.4.5. [Algebraische Vielfachheit] Wir nennen für einen Eigenwert λ von A die Vielfachheit a_λ der Nullstelle λ des charakteristischen Polynoms $\chi_A(X)$ die *algebraische Vielfachheit des Eigenwerts* λ .

Satz 5.4.6. Ähnliche Matrizen haben das gleiche charakteristische Polynom. Insbesondere haben sie die gleichen Eigenwerte, die gleiche Determinante und die gleiche Spur.

Beweis. Ist $B = S^{-1}AS$, so ist auch $B - X\mathbf{1} = S^{-1}(A - X\mathbf{1})S$ und daher

$$\chi_B(X) = \det(B - X\mathbf{1}) = \det(S^{-1})\chi_A(X)\det(S) = \chi_A(X). \quad \square$$

Der folgende Satz folgt unmittelbar aus $\det(A^\top) = \det(A)$:

Satz 5.4.7. Ist $A^\top \in M_n(\mathbb{K})$ die Transponierte der Matrix A , so gilt $\chi_{A^\top} = \chi_A$. Insbesondere haben A und $A^\top$ die gleichen Eigenwerte.

Satz 5.4.8. Für jeden Eigenwert λ einer Matrix $A \in M_n(\mathbb{K})$ haben wir die Ungleichungen

$$1 \leq g_\lambda \leq a_\lambda \leq n$$

für die geometrische Vielfachheit g_λ und die algebraische Vielfachheit a_λ .

Beweis. Sei $k = g_\lambda$ und $v_1, \dots, v_k$ linear unabhängige Eigenvektoren zum Eigenwert λ . Wir ergänzen diese Vektoren zu einer Basis $v_1, \dots, v_n$ von $\mathbb{K}^n$. Für die Matrix S deren Spalten die Vektoren $v_1, \dots, v_k$ sind, erhalten wir die ähnliche Matrix

$$A' := S^{-1}AS = \begin{pmatrix} \lambda \mathbf{1}_k & B \\ 0 & D \end{pmatrix}.$$

Wegen Satz 5.4.6 ist daher

$$\chi_A(X) = \chi_{A'}(X) = \det \begin{pmatrix} (\lambda - X)\mathbf{1}_k & B \\ 0 & D - X\mathbf{1}_{n-k} \end{pmatrix} = (\lambda - X)^k \chi_D(X).$$

Hieraus folgt $g_\lambda = k \leq a_\lambda$. □

Beispiele 5.4.9. $[\mathbb{K} = \mathbb{R}, n = 2]$ (a) Wir betrachten die Matrix $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Zur Berechnung der Eigenwerte betrachten wir das charakteristische Polynom

$$\chi_A(X) = \det(A - X\mathbf{1}) = \begin{vmatrix} 1 - X & 2 \\ 0 & 1 - X \end{vmatrix} = (1 - X)^2.$$

Also ist 1 der einzige Eigenwert, und er hat die algebraischen Vielfachheit $a_1 = 2$.

Eigenvektoren zum einzigen Eigenwert 1 erhalten wir als Lösungen von

$$(A - 1 \cdot \mathbf{1})v = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} v = 0 \quad \implies \quad v \in \mathbb{R} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Also ist die geometrische Vielfachheit des Eigenwertes $\lambda_1 = 1$ gleich $g_1 = 1$. In diesem Beispiel ist $g_1 < a_1$; es gibt also keine Basis von Eigenvektoren.

(b) Für die Matrix

$$C = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \in M_2(\mathbb{R}),$$

die die Drehung der Ebene $\mathbb{R}^2$ um den Winkel $\varphi \in \mathbb{R}$ beschreibt, ist

$$\chi_C[X] = \det \begin{pmatrix} \cos(\varphi) - X & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) - X \end{pmatrix} = (\cos(\varphi) - X)^2 + \sin(\varphi)^2 = X^2 - 2\cos(\varphi)X + 1.$$

Reelle Eigenwerte hat diese Matrix also genau dann, wenn $\cos(\varphi)^2 = 1$ ist, also wenn $\varphi \in \mathbb{Z}\pi$, d.h., φ ist ein ganzzahliges Vielfaches von π . Für $\varphi \in 2\mathbb{Z}\pi$ ist $C = \mathbf{1}$ mit dem zweifachen reellen Eigenwert 1 und für $\varphi \in (2\mathbb{Z} + 1)\pi$ ist $C = -\mathbf{1}$ mit dem zweifachen reellen Eigenwert -1 .

Ist φ kein Vielfaches von π , so ist C über $\mathbb{R}$ nicht diagonalisierbar, da es keine Eigenwerte gibt. Allerdings können wir C auch als komplexe Matrix in $M_2(\mathbb{C})$ auffassen und finden die Eigenwerte

$$\lambda_{1/2} = \cos(\varphi) \pm \sqrt{\cos(\varphi)^2 - 1} = \cos(\varphi) \pm \sqrt{-\sin(\varphi)^2} = \cos(\varphi) \pm i\sin(\varphi).$$

Da diese beiden Eigenwerte wegen $\sin(\varphi) \neq 0$ verschieden sind, ist C über dem Körper $\mathbb{C}$ der komplexen Zahlen diagonalisierbar.

(c) Allgemein erhalten wir für eine reelle Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R})$$

das charakteristische Polynome

$$\chi_A(X) = (a - X)(d - X) - bc = X^2 - (a + d)X + ad - bc = X^2 - \operatorname{tr}(A)X + \det(A)$$

mit den komplexen Nullstellen

$$\lambda_{1/2} = \frac{\operatorname{tr}(A)}{2} \pm \sqrt{\frac{\operatorname{tr}(A)^2}{4} - \det(A)}.$$

Wir sehen also, dass A genau dann reelle Eigenwerte hat, wenn

$$4 \det(A) \leq \operatorname{tr}(A)^2$$

gilt. Ist sogar $4 \det(A) < \operatorname{tr}(A)^2$, so besitzt A zwei verschiedene reelle Eigenwerte und für $4 \det(A) = \operatorname{tr}(A)^2$ erhalten wir den Eigenwert $\operatorname{tr}(A)/2$ mit der algebraischen Vielfachheit 2.

Satz 5.4.10. (Bedingungen für Diagonalisierbarkeit von Matrizen) *Eine $n \times n$ -Matrix $A \in M_n(\mathbb{K})$ ist genau dann diagonalisierbar über $\mathbb{K}$, wenn ihr charakteristisches Polynom $\chi_A(X)$ in ein Produkt von Linearfaktoren*

$$\chi_A(X) = (\lambda_1 - X)^{a_1} \cdots (\lambda_r - X)^{a_r}, \quad \lambda_1, \dots, \lambda_r \in \mathbb{K},$$

zerfällt und falls für jeden Eigenwert λ_j die geometrische Vielfachheit g_j gleich der algebraischen Vielfachheit a_j ist.

Beweis. Seien $\lambda_1, \dots, \lambda_r$ die verschiedenen Eigenwerte von A . Ist A diagonalisierbar und sind $g_1, \dots, g_r$ die geometrischen Vielfachheiten von $\lambda_1, \dots, \lambda_r$, so existiert eine invertierbare Matrix S (deren Spalten die Eigenvektoren von A sind), so dass

$$A' := S^{-1}AS = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & & 0 \\ & \ddots & \\ 0 & & \boxed{\begin{matrix} \lambda_r & 0 \\ & \ddots \\ 0 & \lambda_r \end{matrix}} \end{pmatrix}$$

$\begin{matrix} g_1 \left\{ \right. \\ \\ g_r \left\{ \right. \end{matrix}$

gilt. Dann folgt aus Satz 5.4.6:

$$\chi_A(X) = \chi_{A'}(X) = \det(A' - X\mathbf{1}) = \prod_{j=1}^r (\lambda_j - X)^{g_j}.$$

Insbesondere folgt $a_j = g_j$, die algebraischen und geometrischen Vielfachheiten stimmen also überein.

Jetzt nehmen wir an, dass das charakteristische Polynom $\chi_A(X)$ in Linearfaktoren zerfällt:

$$\chi_A(X) = (\lambda_1 - X)^{a_1} \cdots (\lambda_r - X)^{a_r}$$

und die geometrische Vielfachheit g_j des Eigenwerts λ_j mit der algebraischen Vielfachheit a_j übereinstimmt. Dann ist

$$n = \deg(\chi_A(X)) = a_1 + \dots + a_r = g_1 + \dots + g_r,$$

also folgt aus Bemerkung 5.2.17, dass A diagonalisierbar ist. □

Folgerung 5.4.11. *Ist $\mathbb{K}$ algebraisch abgeschlossen, so ist $A \in M_n(\mathbb{K})$ genau dann diagonalisierbar, wenn für jeden Eigenwert λ die algebraische und die geometrische Vielfachheit übereinstimmen.*

Beweis. Da das charakteristische Polynom $\chi_A(X) \in \mathbb{K}[X]$ wegen der algebraischen Abgeschlossenheit von $\mathbb{K}$ in Linearfaktoren zerfällt, folgt die Behauptung aus Satz 5.4.10. $\square$

Da $\mathbb{C}$ wegen dem Fundamentalsatz der Algebra algebraisch abgeschlossen ist, gilt insbesondere:

Folgerung 5.4.12. *Eine komplexe Matrix $A \in M_n(\mathbb{C})$ ist genau dann diagonalisierbar, wenn für jeden Eigenwert $\lambda \in \mathbb{C}$ die algebraische und die geometrische Vielfachheit übereinstimmen.*

Bemerkung 5.4.13. [Spezialfall: $\mathbb{K} = \mathbb{R}$] Ist eine reelle $n \times n$ -Matrix A diagonalisierbar, so ist sie natürlich auch als komplexe Matrix diagonalisierbar, und alle Eigenwerte sind reell.

Ist umgekehrt, $A \in M_n(\mathbb{R})$ eine Matrix mit reellen Einträgen, die als komplexe Matrix diagonalisierbar ist (Folgerung 5.4.11) und nur reelle Eigenwerte besitzt, so kann man zeigen, dass A auch als reelle Matrix diagonalisierbar ist. Hierzu argumentiert man wie folgt. Sei $\lambda \in \mathbb{R}$ ein Eigenwert von A . Dann ist $A - \lambda \mathbf{1} \in M_n(\mathbb{R})$ eine reelle Matrix. Sei $g_\lambda^{\mathbb{C}}$ die geometrische Vielfachheit von λ als Eigenwert der komplexen Matrix $A \in M_n(\mathbb{C})$ und $g_\lambda^{\mathbb{R}}$ die geometrische Vielfachheit von λ als Eigenwert der reellen Matrix. Dann ist

$$g_\lambda^{\mathbb{R}} = n - \text{rank}(A - \lambda \mathbf{1}) = g_\lambda^{\mathbb{C}},$$

denn der Gauß–Jordan–Algorithmus zeigt, dass der Rang einer reellen Matrix sich nicht ändert, wenn man sie als komplexe Matrix betrachtet. Damit erhalten wir

$$\sum_i g_{\lambda_i}^{\mathbb{R}} = \sum_i g_{\lambda_i}^{\mathbb{C}} = n,$$

woraus mit Bemerkung 5.2.17 folgt, dass A reell diagonalisierbar ist.

5.4.14. [Anwendungen der Diagonalisierbarkeit] Wir beachten zunächst, dass für beliebige $n \times n$ -Matrizen A, B und für jede invertierbare $n \times n$ -Matrix S die folgenden Beziehungen gelten:

$$S^{-1}(A + B)S = S^{-1}AS + S^{-1}BS \quad (\text{i})$$

$$S^{-1}(\lambda A)S = \lambda \cdot S^{-1}AS \quad (\text{ii})$$

$$S^{-1}(AB)S = (S^{-1}AS) \cdot (S^{-1}BS) \quad (\text{iii})$$

Das bedeutet, dass die Abbildung

$$\Psi_S : M_n(\mathbb{K}) \rightarrow M_n(\mathbb{K}), \quad A \mapsto S^{-1}AS,$$

ein Automorphismus der Algebra $M_n(\mathbb{K})$ ist, mit der Inversen

$$\Psi_{S^{-1}} : M_n(\mathbb{K}) \rightarrow M_n(\mathbb{K}), \quad A \mapsto SAS^{-1}.$$

Im Spezialfall $A = B$ in (iii) erhalten wir $S^{-1}A^2S = (S^{-1}AS)^2$; mittels Induktion folgt:

$$S^{-1}A^nS = (S^{-1}AS)^n. \quad (\text{iv})$$

Setzen wir $B = S^{-1}AS$, so gilt

$$A^n = SB^nS^{-1} \quad (\text{v})$$

Beispiel 5.4.15. Wie in Beispiel 5.2.7 sei $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. Dann gilt $S^{-1}AS = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$ mit

$$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{und} \quad S^{-1} = -\frac{1}{2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Nun können wir die Potenzen A^n explizit berechnen:

$$\begin{aligned}
 A^n &= S(S^{-1}AS)^n S^{-1} = S \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}^n S^{-1} \\
 &= S \begin{pmatrix} 4^n & 0 \\ 0 & (-2)^n \end{pmatrix} S^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 4^n & 0 \\ 0 & (-2)^n \end{pmatrix} \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
 &= \begin{pmatrix} 4^n & (-2)^n \\ 4^n & -(-2)^n \end{pmatrix} \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
 &= \frac{1}{2} \begin{pmatrix} 4^n + (-2)^n & 4^n - (-2)^n \\ 4^n - (-2)^n & 4^n + (-2)^n \end{pmatrix} = 2^{n-1} \begin{pmatrix} 2^n + (-1)^n & 2^n - (-1)^n \\ 2^n - (-1)^n & 2^n + (-1)^n \end{pmatrix}.
 \end{aligned}$$

Aufgabe 5.4.16. Für $a_0, \dots, a_{n-1} \in \mathbb{K}$ betrachten wir die Matrix

$$A := \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & -a_2 \\ 0 & 0 & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix} \in M_n(\mathbb{K}).$$

- (a) Zeigen Sie $\chi_A(X) = (-1)^n(X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0)$.
 (b) Für jeden Eigenwert λ hat der zugehörige Eigenraum die Dimension $g_\lambda = 1$.

5.5 Trigonalisierung

Es gibt auch über dem Körper der komplexen Zahlen nichtdiagonalisierbare Matrizen wie z.B.

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},$$

siehe Beispiel 5.4.9. In diesem Fall ist das charakteristische Polynom gegeben durch $\chi_A(X) = (1 - X)^2$. Es gibt also genau einen Eigenwert $\lambda = 1$, dessen algebraische Vielfachheit $a_1 = 2$ echt größer als seine geometrische Vielfachheit $g_1 = 1$ ist, denn $g_1 = 2 - \text{rank}(A - \mathbf{1}) = 1$.

5.5.1 Trigonalisierung komplexer Matrizen

Wir stellen uns die Frage, ob jede Matrix einer Matrix ähnlich ist, die *eine obere Dreiecksmatrix* ist. Wir werden später sehen, dass über dem Körper der komplexen Zahlen die *Jordansche Normalform* die bestmögliche Lösung ergibt. Im Augenblick halten wir Folgendes fest:

Lemma 5.5.1. *Jede komplexe Matrix $A \in M_n(\mathbb{C})$ hat einen Eigenwert.*

Beweis. Nach dem Fundamentalsatz der Algebra hat das charakteristische Polynom $\chi_A(X) = \det(A - X\mathbf{1})$ eine Nullstelle, also A einen Eigenwert. $\square$

Satz 5.5.2. [Trigonalisierung komplexer Matrizen] *Jede komplexe $n \times n$ -Matrix A ist einer oberen Dreiecksmatrix ähnlich, d.h. für jedes $A \in M_n(\mathbb{C})$ gibt es eine invertierbare Matrix $S \in \text{GL}_n(\mathbb{C})$, so dass*

$$A' := S^{-1}AS = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

eine obere Dreiecksmatrix ist.

Beweis. Der Beweis wird mittels Induktion nach n geführt. Für $n = 1$ ist A schon in Dreiecksgestalt und wir können $S = \mathbf{1}_n$ wählen. Für $n > 1$ sei v ein Eigenvektor von A zu einem Eigenwert λ (Lemma 5.5.1). Wir erweitern v zu einer Basis von $\mathbb{C}^n$, deren Vektoren wir, v zuerst, als Spalten in eine Matrix T schreiben. Die Matrix T ist eine invertierbare Matrix mit v als erster Spalte und $T^{-1}AT$ ist die Matrix von A bzgl. dieser neuen Basis, das heißt dass $T^{-1}AT$ die Form

$$T^{-1}AT = \begin{pmatrix} \lambda & * & \cdots & * \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}$$

für eine Matrix $B \in M_{n-1}(\mathbb{C})$ hat. Anwendung der Induktionsannahme auf B ergibt eine Matrix $V \in \text{GL}_{n-1}(\mathbb{C})$, für die $B' := V^{-1}BV$ obere Dreiecksgestalt hat. Nun sei

$$V' := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & V & \\ 0 & & & \end{pmatrix}$$

Damit ist

$$(V')^{-1}T^{-1}ATV' = \begin{pmatrix} \lambda & * & \cdots & * \\ 0 & & & \\ \vdots & & V^{-1}BV & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} \lambda & * & \cdots & * \\ 0 & & & \\ \vdots & & B' & \\ 0 & & & \end{pmatrix}$$

eine obere Dreiecksmatrix und $S := TV'$ ist die gesuchte Transformationsmatrix. $\square$

Folgerung 5.5.3. [Trigonalisierung von Endomorphismen] *Sei $\varphi : V \rightarrow V$ ein Endomorphismus eines n -dimensionalen Vektorraumes V über dem Körper $\mathbb{C}$ der komplexen Zahlen. Dann gibt es eine Basis $B = (v_1, v_2, \dots, v_n)$ in $\mathbb{C}^n$, so dass die Matrix $[\varphi]_B$ von φ bzgl. dieser Basis obere Dreiecksgestalt hat.*

Beweis. Ist B irgendeine Basis von V und $A := [\varphi]_B$ die Matrix von φ bzgl. B , so folgt aus der Ähnlichkeit von A zu einer oberen Dreiecksmatrix A' die Existenz einer Basis B' mit $A' = [\varphi]_{B'}$ (Satz 5.1.5). $\square$

Bemerkung 5.5.4. (Trigonalisierung über allgemeinen Körpern) (a) Wir nennen $A \in M_n(\mathbb{K})$ *trigonalisierbar*, wenn A ähnlich zu einer oberen Dreiecksmatrix B ist. Aus $\chi_A(X) = \chi_B(X) = \prod_{j=1}^n (b_{jj} - X)$ folgt dann sofort, dass das charakteristische Polynom von X in Linearfaktoren zerfällt.

(b) Zerfällt $\chi_A(X)$ in Linearfaktoren, so besitzt A einen Eigenwert λ und wie in Beweis von Satz 5.5.2 sehen wir, dass A ähnlich zu einer Matrix der Gestalt

$$A' = \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix}, \quad B \in M_{n-1}(\mathbb{K}),$$

ist. Dann ist aber

$$\chi_A(X) = \chi_{A'}(X) = (\lambda - X)\chi_B(X)$$

und somit zerfällt auch $\chi_B(X)$ in Linearfaktoren.

Induktiv sieht man so ein, dass eine Matrix $A \in M_n(\mathbb{K})$ genau dann trigonalisierbar ist, wenn ihr charakteristisches Polynom χ_A in Linearfaktoren zerfällt. Insbesondere ist das immer der Fall, wenn $\mathbb{K}$ algebraisch abgeschlossen ist.

5.5.2 Blocktrigonalisierung reeller Matrizen

Notwendig für die Trigonalisierbarkeit einer Matrix A ist, dass sie einen Eigenwert besitzt (sogar, dass χ_A in Linearfaktoren zerfällt). Hieraus folgt sofort, dass nicht jede reelle Matrix trigonalisierbar ist. Trotzdem wollen wir an dieser Stelle nicht aufgeben und zeigen, dass sich reelle Matrizen auf eine etwas gröbere Dreiecksgestalt transformieren lassen.

Definition 5.5.5. Sei $\varphi: V \rightarrow V$ linear. Dann heißt ein Unterraum $U \subseteq V$ *invariant unter φ* oder φ -invariant, wenn $\varphi(U) \subseteq U$ gilt.

Beispiel 5.5.6. Für jeden Eigenvektor $v \in V$ von $\varphi \in \text{End}(V)$ ist $\mathbb{K}v$ ein invarianter Unterraum der Dimension 1. Ist umgekehrt $U = \mathbb{K}v$ ein eindimensionaler Unterraum, so ist $\varphi(v) \in \mathbb{K}v$ und damit $\varphi(v) = \lambda v$ für ein $\lambda \in \mathbb{K}$. Also ist v ein Eigenvektor. Die Eigenvektoren erzeugen also genau die eindimensionalen invarianten Unterräume.

Bemerkung 5.5.7. Besitzt die lineare Abbildung $\varphi: V \rightarrow V$ des n -dimensionalen Vektorraums V einen invarianten Unterraum U der Dimension k , so kann man eine Basis $B = (v_1, \dots, v_n)$ von V so wählen, dass $v_1, \dots, v_k$ eine Basis von U ist. Für $j \leq k$ ist dann $\varphi(v_j) \in U = \text{span}(v_1, \dots, v_k)$. Die Matrix $[\varphi]_B$ von φ bzgl. B hat also die Gestalt

$$[\varphi]_B = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \quad \text{mit} \quad A_{11} \in M_k(\mathbb{K}), A_{12} \in M_{k, n-k}(\mathbb{K}), A_{22} \in M_{n-k}(\mathbb{K}).$$

Definition 5.5.8. (Blockmatrizen) Ist $n = n_1 + \dots + n_k$ und $m = m_1 + \dots + m_r$, so kann man jede Matrix $A \in M_{n,m}(\mathbb{K})$ auch als sogenannte $(k \times r)$ -Blockmatrix $A = (A_{ij})$ schreiben. Hierbei fasst die Matrix $A_{ij} \in M_{n_i, m_j}(\mathbb{K})$ alle Einträge $a_{k,\ell}$ zusammen, für die

$$n_1 + \dots + n_{i-1} < k \leq n_1 + \dots + n_i, \quad m_1 + \dots + m_{j-1} < \ell \leq m_1 + \dots + m_j$$

gilt. Das sieht komplizierter aus als es ist und erweist sich dann als sehr praktisch, wenn viele dieser Blöcke verschwinden. In den meisten Fällen ist $k = r = 2$ und man hat eine Blockstruktur der Gestalt

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

wie wir es in Bemerkung 5.5.7 gesehen haben.

Eine Blockmatrix $A = (A_{ij})$ heißt *obere Blockdreiecksmatrix*, wenn $A_{ij} = 0$ für $i > j$ gilt und *unter Blockdreiecksmatrix*, wenn $A_{ij} = 0$ für $i < j$ gilt.

Eigenräume liefern invariante Unterräume der Dimension 1. Das nächstbeste sind invariante Unterräume der Dimension 2 und so etwas existiert auch im Reellen. Hier ist die zentrale Idee:

Bemerkung 5.5.9. Sei $A \in M_n(\mathbb{R})$ und $\lambda = \alpha + i\beta \in \mathbb{C}$ ein komplexer Eigenwert von A (der nach dem Fundamentalsatz der Algebra immer existiert!). Wir schreiben einen zugehörigen Eigenvektor als

$$z = x + iy \in \mathbb{C}^n \quad \text{mit} \quad x, y \in \mathbb{R}^n.$$

Aus

$$Az = Ax + iAy \quad \text{und} \quad \lambda z = (\alpha x - \beta y) + i(\alpha y + \beta x)$$

erhalten wir durch Vergleich von Real- und Imaginärteil die beiden Relationen

$$Ax = \alpha x - \beta y \quad \text{und} \quad Ay = \beta x + \alpha y.$$

Sind x und y linear abhängig, so spannen sie einen eindimensionalen A -invarianten Unterraum auf. Ist μ der zugehörige reelle Eigenwert, so ist $Ax = \mu x$, $Ay = \mu y$ und daher auch $Az = \mu z$. Insbesondere ist dann $\lambda = \mu \in \mathbb{R}$, d.h., $\beta = 0$.

Wir nehmen daher an, dass $\beta \neq 0$ ist, d.h., $\lambda \notin \mathbb{R}$. Dann sind x und y linear unabhängig und spannen einen zweidimensionalen Unterraum $U := \mathbb{R}x + \mathbb{R}y$ auf, der unter A invariant ist. Die Einschränkung der linearen Abbildung $L_A v = Av$ auf U hat bzgl. der Basis $B := (x, y)$ die Matrix

$$[L_A|_U]_B = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}.$$

Schreiben wir $\lambda = r(\cos \theta + i \sin \theta)$ in Polarkoordinaten, erhalten wir

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} = r \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

Das ist die Matrix einer sogenannten *Drehstreckung*, also ein Vielfaches einer Drehmatrix. Hierbei ist r der Streckungsfaktor und θ der Drehwinkel (im Uhrzeigersinn!).

Damit haben wir ein Analogon von Lemma 5.5.1 im Reellen und können den folgenden Satz beweisen.

Satz 5.5.10. [Blocktrigonalisierung reeller Matrizen] *Jede reelle $n \times n$ -Matrix A ist ähnlich zu einer oberen Blockdreiecksmatrix*

$$A' = (A'_{ij}) = \begin{pmatrix} A'_{11} & A'_{12} & \cdots & A'_{1k} \\ 0 & A'_{22} & \cdots & A'_{2k} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & A'_{kk} \end{pmatrix},$$

deren Diagonalblöcke entweder die Gestalt $A'_{jj} = (\mu)$ mit einem reellen Eigenwert μ von A haben, oder die Gestalt

$$A'_{jj} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix},$$

wobei $\lambda = \alpha + i\beta \in \mathbb{C} \setminus \mathbb{R}$ ein komplexer Eigenwert ist.

Beweis. Wir argumentieren genau wie im komplexen Fall mit vollständiger Induktion nach der Dimension n .

Wie wir in Bemerkung 5.5.9 gesehen haben, besitzt A entweder einen reellen Eigenwert μ und einen zugehörigen Eigenvektor v_1 , oder es existiert ein komplexer Eigenwert $\lambda = \alpha + i\beta \in \mathbb{C} \setminus \mathbb{R}$ und linear unabhängige Vektoren $v_1, v_2 \in \mathbb{R}^n$ mit

$$Av_1 = \alpha v_1 - \beta v_2 \quad \text{und} \quad Av_2 = \beta v_1 + \alpha v_2.$$

Wir ergänzen diese Vektoren zu einer Basis $B = (v_1, \dots, v_n)$ von $\mathbb{R}^n$. Dann ist A ähnlich zu einer Blockmatrix der Gestalt

$$C = \begin{pmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{pmatrix},$$

wobei C_{11} bereits die gewünschte Form besitzt. Nach Induktionsvoraussetzung gilt unsere Behauptung für die kleinere Matrix C_{22} . Wegen $\chi_A = \chi_{C_{11}} \chi_{C_{22}}$ (Aufgabe 5.5.11) sind alle Eigenwerte von C_{22} auch solche von A . Durch Konjugation mit einer Transformationsmatrix der Blockgestalt

$$\tilde{S} = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & S \end{pmatrix},$$

für die $S^{-1}C_{22}S$ die gewünschte Gestalt besitzt, sieht man nun, dass $\tilde{S}^{-1}C\tilde{S} \approx C \approx A$ die gewünschte Gestalt hat. $\square$

Aufgabe 5.5.11. Sei $A \in M_n(\mathbb{K})$, $1 \leq k < n$, so dass A die folgende Blockstruktur besitzt:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \quad \text{mit} \quad A_{11} \in M_k(\mathbb{K}), A_{12} \in M_{k,n-k}(\mathbb{K}), A_{22} \in M_{n-k}(\mathbb{K}).$$

Zeigen Sie, dass das charakteristische Polynom wie folgt faktorisiert:

$$\chi_A(X) = \chi_{A_{11}}(X) \chi_{A_{22}}(X).$$

5.6 Ideale im Polynomring

In diesem Abschnitt bauen wir unsere Kenntnisse über den Ring $\mathbb{K}[X]$ der formalen Polynome über einem Körper $\mathbb{K}$ noch etwas aus. Im Zentrum steht dabei zunächst die Einsicht, dass man für Polynome, genau wie für ganze Zahlen, eine Division von $p(X)$ durch $q(X)$ mit Rest in eindeutigerweise definieren kann, wenn man verlangt, dass der Grad des Restpolynoms kleiner als der Grad von $q(X)$ ist. Mittels Polynomdivision studieren wir dann die Struktur sogenannter Ideale in dem Ring $\mathbb{K}[X]$ und zeigen, dass sie alle von einem Element erzeugt werden, also sogenannte Hauptideale sind.

Satz 5.6.1. [Polynomdivision] Sind $p, q \in \mathbb{K}[X]$ mit $q \neq 0$, so existieren eindeutig bestimmte Polynome $s, r \in \mathbb{K}[X]$ mit

$$p = s \cdot q + r, \quad \deg(r) < \deg(q).$$

Das Polynom r heißt *Rest des Polynoms p nach Division durch q* . Das Polynom q heißt *Teiler von p* , wenn $r = 0$ ist. In diesem Fall schreiben wir auch $s = \frac{p}{q}$.

Beweis. Eindeutigkeit: Haben wir zwei Darstellungen

$$p = sq + r = \tilde{s}q + \tilde{r}$$

mit $\deg(r), \deg(\tilde{r}) < m$, so erhalten wir durch Differenzbildung

$$(s - \tilde{s})q = \tilde{r} - r.$$

Ist $s \neq \tilde{s}$, so erhalten wir für die Grade auf beiden Seiten

$$\deg((s - \tilde{s})q) = \deg(s - \tilde{s}) + \deg(q) = \deg(r - \tilde{r}) < \deg(q),$$

im Widerspruch zu Satz 5.3.2.

Existenz: Sei

$$p(X) = a_0 + a_1X + \dots + a_nX^n \quad \text{und} \quad q(X) = b_0 + b_1X + \dots + b_mX^m$$

mit $a_n \neq 0 \neq b_m$. Wir zeigen die Existenz von s und r über Induktion nach dem Grad n von p .

1. Fall: Ist $n < m$, so setzen wir $s := 0$ und $r := p$ und erhalten $p = sq + r$ mit $\deg(r) = n < m$.
2. Fall: Ist $n \geq m$, so betrachten wir das Polynom

$$\begin{aligned} \tilde{p}(X) &:= p(X) - \frac{a_n}{b_m} X^{n-m} q(X) = (a_n X^n + a_{n-1} X^{n-1} + \dots) - \frac{a_n}{b_m} X^{n-m} (b_m X^m + \dots) \\ &= (a_{n-1} - \frac{a_n}{b_m} b_{m-1}) X^{n-1} + \dots, \end{aligned}$$

also ist $\deg(\tilde{p}) < n$. Gemäß der Induktionsannahme existieren eindeutige Polynome $\tilde{s}$ und $\tilde{r}$ mit $\deg(\tilde{r}) < n$ und

$$\tilde{p} = \tilde{s}q + \tilde{r}.$$

Dann ist

$$p(X) = \tilde{p}(X) + \frac{a_n}{b_m} X^{n-m} q(X) = \left(\tilde{s}(X) + \frac{a_n}{b_m} X^{n-m} \right) q(X) + \tilde{r}(X),$$

und die Behauptung folgt mit

$$s(X) := \tilde{s}(X) + \frac{a_n}{b_m} X^{n-m}, \quad r(X) := \tilde{r}(X). \quad \square$$

Beispiel 5.6.2. Für

$$p(X) := X^4 + X^3 + X^2 + 1 \quad \text{und} \quad q(X) = X^3 - X$$

erhalten wir sukzessive

$$p_1(X) := p(X) - Xq(X) = (X^4 + X^3 + X^2 + 1) - X(X^3 - X) = X^3 + 2X^2 + 1,$$

$$p_2(X) := p_1(X) - q(X) = 2X^2 + X + 1.$$

Durch Rückwärtseinsetzen ergibt sich damit $p = sq + r$ mit

$$r(X) := 2X^2 + X + 1 \quad \text{und} \quad s(X) := X + 1.$$

Definition 5.6.3. Eine Teilmenge $I \subseteq \mathbb{K}[X]$ heißt *Ideal*, wenn

$$I + I \subseteq I \quad \text{und} \quad \mathbb{K}[X]I \subseteq I$$

gelten, d.h. für $f, g \in I$ und $h \in \mathbb{K}[X]$ sind $f + g, fh \in I$. Hieraus folgt insbesondere, dass I ein Untervektorraum ist.

Beispiel 5.6.4. (a) Ist $\varphi: \mathbb{K}[X] \rightarrow R$ ein Homomorphismus von Ringen, so ist $\ker \varphi := \varphi^{-1}(0)$ ein Ideal. In der Tat folgt aus $\varphi(f) = \varphi(g) = 0$ und $h \in \mathbb{K}[X]$:

$$\varphi(f + g) = \varphi(f) + \varphi(g) = 0 \quad \text{und} \quad \varphi(fh) = \varphi(f)\varphi(h) = 0\varphi(h) = 0.$$

(b) Ist $p(X) \in \mathbb{K}[X]$, so ist

$$(f) := f(X)\mathbb{K}[X] := \{f(X)q(X) : q(X) \in \mathbb{K}[X]\}$$

ein Ideal. Es besteht aus allen Vielfachen von $f(X)$ bzw. aus allen Polynomen, die durch $f(X)$ (ohne Rest) geteilt werden. Auch hier ist die Verifikation trivial:

$$f(X)q_1(X) + f(X)q_2(X) = f(X)(q_1(X) + q_2(X)), \quad f(X)q_1(X)q_2(X) = f(X)(q_1(X)q_2(X)).$$

Definition 5.6.5. Ideale der Gestalt $(f) \subseteq \mathbb{K}[X]$ heißen *Hauptideale*.

Lemma 5.6.6. Jedes Ideal $I \subseteq \mathbb{K}[X]$ ist ein Hauptideal. Ist $I \neq \{0\}$, so enthält I genau ein Element f mit minimalem Grad und Leitkoeffizient 1 und es gilt $I = (f)$.

Man nennt $\mathbb{K}[X]$ daher auch einen *Hauptidealring*.

Beweis. Ist $I = \{0\}$, so nehmen wir $f := 0$ und $I = (0) = 0 \cdot \mathbb{K}[X]$ gilt trivialerweise.

Ist $I \neq \{0\}$, so ist

$$\{n \in \mathbb{N}_0 : (\exists f \in I \setminus \{0\}) \deg(f) = n\}$$

nicht leer, enthält also ein minimales Element d . Sei $f \in I$ mit $\deg(f) = d$ und Leitkoeffizient 1. Ist $g \in I$, so erhalten wir mittels Polynomdivision Polynome $q, r \in \mathbb{K}[X]$ mit

$$g = f \cdot q + r, \quad \deg(r) < \deg(f) = d.$$

Wegen $f, g \in I$ ist auch $f \cdot q \in I$ und daher $r = g - fq \in I$. Aus $\deg(r) < d$ folgt daher $r = 0$, also $g = fq \in f \cdot \mathbb{K}[X]$. Da I ein Ideal ist, gilt $f \cdot \mathbb{K}[X] \subseteq I$ trivialerweise. Also ist $I = (f)$. Ist $\deg(g) = d$, so ist $\deg(q) = 1$, also $q \in \mathbb{K}$ und somit q der Leitkoeffizient von g . Ist dieser 1, so erhalten wir $g = f$. $\square$

Definition 5.6.7. (a) Sind $p_1, \dots, p_n \in \mathbb{K}[X]$ Polynome, so definieren wir den *größten gemeinsamen Teiler (ggT)* von $p_1, \dots, p_n$ durch

$$\text{ggT}(p_1, \dots, p_n) := f \quad \text{für} \quad (p_1) + \dots + (p_n) = (f),$$

wobei f den Leitkoeffizienten 1 haben soll.

(b) Wir nennen $p_1, \dots, p_n$ *teilerfremd*, wenn $\text{ggT}(p_1, \dots, p_n) = 1$ ist, also wenn Polynome $r_1, \dots, r_n \in \mathbb{K}[X]$ existieren mit

$$1 = r_1 p_1 + \dots + r_n p_n.$$

Aufgabe 5.6.8. Zeigen Sie die Relation

$$\text{ggT}(p_1, p_2, \dots, p_n) := \text{ggT}(p_1, \text{ggT}(p_2, \dots, p_n))$$

für $p_1, \dots, p_n \in \mathbb{K}[X]$.

Aufgabe 5.6.9. (a) Sind p_1 und p_2 teilerfremd, so ist

$$\text{ggT}(p_1, p_2 p_3) = \text{ggT}(p_1, p_3).$$

Hinweis: $p_3 = ap_1 + bp_2$ für geeignete $a, b \in \mathbb{K}[X]$.

(b) Ist $\text{ggT}(q, p_j) = 1$ für $j = 1, \dots, n$, so ist auch

$$\text{ggT}(q, p_1 p_2 \cdots p_n) = 1.$$

(c) Sind $p_1, p_2, \dots, p_n$ paarweise teilerfremd, so gilt

$$\text{ggT}(q_1, \dots, q_n) = 1$$

für die Polynome

$$q_j := p_1 \cdots p_{j-1} p_{j+1} \cdots p_n, \quad j = 1, \dots, n.$$

Hinweis: Aus (b) folgt, dass q_j und p_j teilerfremd sind. Jeder Teiler d von allen q_j ist daher teilerfremd zu allen p_j , also auch zu allen q_j (wegen (b)).

5.7 Der Satz von Cayley–Hamilton

Definition 5.7.1. [Anwendung eines Polynoms auf eine Matrix] Sei $A \in M_n(\mathbb{K})$ eine Matrix und

$$P = a_0 + a_1 X + \dots + a_n X^n \in \mathbb{K}[X]$$

ein formales Polynom in der Unbestimmten X . In diesem Abschnitt schreiben wir Elemente von $\mathbb{K}[X]$ ohne die Variable X , da dies hier praktischer ist. Wir definieren eine Matrix

$$P(A) := a_0 \mathbf{1} + a_1 A + \dots + a_n A^n.$$

D.h. wir ersetzen die Unbestimmte X durch A .

Ist zum Beispiel $P = X^2 + 2X + 1$, so erhalten wir $P(A) = A^2 + 2A + \mathbf{1}$. Ist $Q = X^2 + 4X + 4$, so gilt $Q(A) = A^2 + 4A + 4\mathbf{1}$. Man rechnet leicht nach, dass die Abbildung

$$\Phi_A: \mathbb{K}[X] \rightarrow M_n(\mathbb{K}), \quad P \mapsto P(A)$$

ein Homomorphismus von Ringen ist, also eine lineare Abbildung, die zusätzlich mit der Multiplikation verträglich ist:

$$P(A)Q(A) = (PQ)(A) \quad \text{für alle } P, Q \in \mathbb{K}[X].$$

Definition 5.7.2. [Anwendung eines Polynoms auf eine lineare Abbildung] Genauso wie Matrizen, kann man auch lineare Abbildungen in Polynome einsetzen: Sei $\varphi: V \rightarrow V$ eine lineare Abbildung auf einem $\mathbb{K}$ -Vektorraum und

$$P = a_0 + a_1 X + \dots + a_n X^n \in \mathbb{K}[X]$$

ein Polynom, so definieren wir

$$P(\varphi) := a_0 \text{id}_V + a_1 \varphi + \dots + a_n \varphi^n.$$

Bemerkung 5.7.3. [Eine Matrix als Nullstelle eines Polynoms] Bei der erneuten Betrachtung von Matrizen beachten wir, dass für eine gegebene Matrix A und ein Polynom P der Fall $P(A) = 0$ eintreten kann. Ist z.B. wie zuvor $P = X^2 - 2X + 1 = (X - 1)^2$ und $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, so gilt $P(A) = (A - \mathbf{1})^2 = 0$. Andererseits ist $A - \mathbf{1} \neq 0$. Für das Polynom $Q := X - 1$ gilt also $Q(A) \neq 0$. Dieses Phänomen tritt nicht auf, wenn wir Nullstellen der Polynome in $\mathbb{K}$ betrachten, denn Q und P haben die gleichen Nullstellen. Wenden wir Q und P auf Matrizen an, so haben sie verschiedene “Nullstellen.”

Matrizen haben Nullteiler; Polynome nicht!

Gilt $P(A) = 0$ für ein Polynom $P \in \mathbb{K}[X]$, so sagen wir P *annulliert* A . Von Nullstellen sprechen wir nur, wenn wir Zahlen, d.h. Elemente von $\mathbb{K}$, in P einsetzen. Ist λ Wurzel eines Polynoms P , dann annulliert P natürlich $\lambda \mathbf{1}$, denn es gilt

$$P(\lambda \mathbf{1}) = P(\lambda) \mathbf{1} = 0.$$

Die obigen Beispiele zeigen allerdings, dass das Annullieren nicht immer von so einfacher Natur ist.

Bemerkung 5.7.4. (a) Ist $A \in M_n(\mathbb{K})$, so ist

$$I_A := \{p \in \mathbb{K}[X] : p(A) = 0\}$$

ein Ideal. Es ist klar, dass I_A ein Untervektorraum ist. Um einzusehen, dass I_A ein Ideal ist, sei $f \in I_A$ und $g \in \mathbb{K}[X]$. Dann ist

$$(fg)(A) = f(A)g(A) = 0 \cdot g(A) = 0,$$

also $fg \in I_A$.

(b) Ist $\varphi \in \text{End}(V)$, so ist

$$I_\varphi := \{p \in \mathbb{K}[X] : p(\varphi) = 0\}$$

ein Ideal. Das sieht man genauso, wie in (a) für Matrizen.

Es ist ein überraschendes Ergebnis, dass jede Matrix durch ihr charakteristisches Polynom annulliert wird.

Satz 5.7.5. [Cayley–Hamilton] Sei $A \in M_n(\mathbb{K})$ und $\chi_A(X) = \det(A - X\mathbf{1})$ ihr charakteristisches Polynom. Die Einsetzung von A in das charakteristische Polynom $\chi_A(X)$ ergibt die Nullmatrix, d.h.

$$\chi_A(A) = 0.$$

Beweis. Wir erinnern uns an die Relation

$$\mathbf{Ad}(A)^t \cdot A = \det A \cdot \mathbf{1}_n$$

für die adjungierte Matrix $\mathbf{Ad}(A)$, deren Einträge gegeben sind durch

$$\mathbf{Ad}(A)_{ij} = (-1)^{i+j} \det(A_{ij}),$$

wobei $A_{ij} \in M_{n-1}(\mathbb{K})$ die Matrix ist, die man durch streichen der i .Zeile unter j .Spalte von A erhält. Damit erhalten wir für $\chi_A(X) = a_0 + a_1X + \dots + a_nX^n$.

$$\mathbf{Ad}(A - X\mathbf{1})^t \cdot (A - X\mathbf{1}) = \det(A - X\mathbf{1}) \cdot \mathbf{1} = \chi_A(X) \cdot \mathbf{1} = \left(\sum_{j=0}^n a_j X^j \right) \mathbf{1}, \quad (*)$$

wenn wir diese Relation auf die Matrix $B := A - X\mathbf{1} \in M_n(\mathbb{K}[X])$ über dem Ring $\mathbb{K}[X]$ anwenden.

Die Einträge $b'_{ij} = (-1)^{i+j} \det((A - X\mathbf{1})_{ij})$ der Matrix $\mathbf{Ad}(A - X\mathbf{1})$ sind Polynome in X vom Grad $\leq n-1$, da $B_{ij} = A_{ij} - X\mathbf{1}_{ij}$ eine $(n-1) \times (n-1)$ -Matrix ist. Wir erhalten also

$$\mathbf{Ad}(A - X\mathbf{1})^t = C_0 + C_1X + \cdots + C_{n-1}X^{n-1} \quad \text{mit} \quad C_0, C_1, \dots, C_{n-1} \in M_n(\mathbb{K}),$$

d.h. die Matrix C_j enthält in der Position (k, ℓ) jeweils den Koeffizienten von X^j des Polynoms, das an der Position (k, ℓ) der Matrix steht.

Setzen wir in (*) ein, erhalten wir durch Vergleich der Koeffizienten in $\mathbb{K}[X]$:

$$\begin{array}{rclcl} C_0 A & & = & a_0 \mathbf{1} & / \cdot \mathbf{1} \\ C_1 A & - & C_0 & = & a_1 \mathbf{1} & / \cdot A \\ C_2 A & - & C_1 & = & a_2 \mathbf{1} & / \cdot A^2 \\ \vdots & & \vdots & & & \\ C_{n-1} A & - & C_{n-2} & = & a_{n-1} \mathbf{1} & / \cdot A^{n-1} \\ & & -C_{n-1} & = & a_n \mathbf{1} & / \cdot A^n \\ \hline 0 & & = & \chi_A(A) & & \end{array}$$

Die letzte Zeile erhält man durch Aufsummieren der linken und rechten Seiten nach deren Multiplikation mit den auf der rechten Seite angezeigten Faktoren. $\square$

Aufgabe 5.7.6. Es ist instruktiv, sich zu überlegen, warum der folgende “einfache” Beweis des Satzes von Cayley-Hamilton nicht schlüssig ist:

$$\chi_A(A) = \det(A - A \cdot \mathbf{1}) = \det(A - A) = 0.$$

Beispiele 5.7.7. (für $\mathbb{K} = \mathbb{R}$) (a) Wie in Beispiel 5.2.7 sei $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. Das charakteristische Polynom ist $\chi_A(X) = (1 - X)^2 - 9$. Damit folgt aus dem Satz von Cayley-Hamilton

$$(\mathbf{1} - A)^2 - 9 \cdot \mathbf{1} = 0,$$

d.h. es ist $A^2 - 2A + 8 \cdot \mathbf{1} = 0$ und damit $A^2 = 2A + 8 \cdot \mathbf{1}$. (Probe!)

Als Konsequenz kann man jede Potenz von A als Linearkombination der Matrizen A und $\mathbf{1} = A^0$ darstellen. Es gilt z.B.

$$A^3 = AA^2 = A(2A + 8 \cdot \mathbf{1}) = 2A^2 + 8A = 2(2A + 8 \cdot \mathbf{1}) + 8A = 12A + 16 \cdot \mathbf{1}.$$

(b) Wir betrachten wie in Beispiel 5.2.6 die Matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Das charakteristische Polynom ist $\chi_A(X) = (1 - X)^2 + 1$. Nach dem Satz von Cayley-Hamilton folgt

$$\chi_A(A) = (\mathbf{1} - A)^2 + \mathbf{1} = 0$$

und daraus $A^2 - 2A + 2 \cdot \mathbf{1} = 0$. Damit können wir $A^2 = 2(A - \mathbf{1})$ folgern. (Probe!)

Satz 5.7.8. Für jede Matrix $A \in M_n(\mathbb{K})$ existiert genau ein Polynom μ_A minimalen Grades mit Leitkoeffizient 1, für das

$$\mu_A(A) = 0$$

gilt. Dieses Polynom μ_A teilt jedes anderen Polynom f mit $f(A) = 0$.

Das Polynom μ_A heißt *Minimalpolynom der Matrix A*.

Beweis. Nach dem Satz von Cayley-Hamilton gilt $\chi_A(A) = 0$. Das Ideal

$$I_A := \{f \in \mathbb{K}[X] : f(A) = 0\}$$

(Bemerkung 5.7.4) ist also von $\{0\}$ verschieden. Nach Lemma 5.6.6 ist I_A ein Hauptideal, das von dem eindeutig bestimmten Polynom $\mu_A \in I_A$ minimalen Grades erzeugt wird. Hieraus folgt die Behauptung. $\square$

Bemerkung 5.7.9. Nicht alle lineare Abbildungen haben annullierende Polynome, aber wir werden im Rahmen unsrer Diskussion der Jordanschen Normalform sehen, dass uns die Polynome, die eine lineare Abbildung φ annullieren schon sehr viel über φ verraten.

Ist z.B. $V = \mathcal{F}(\mathbb{N}_0, \mathbb{R}) \cong \mathbb{R}^{\mathbb{N}_0}$ der unendlich-dimensionale Vektorraum der Folgen von reellen Zahlen, dann wird die Verschiebungsabbildung $\varphi : V \rightarrow V$ mit $\varphi(a_0, a_1, a_2, \dots) = (0, a_0, a_1, \dots)$ durch kein Polynom annulliert. Ist $e_n = (\delta_{nm})_{m \in \mathbb{N}_0} \in V$ die Folge, die genau eine 1 an der n -ten Stelle hat und sonst Nullen, so gilt für $P = a_0 + a_1X + \dots + a_nX^n$ die Beziehung

$$P(\varphi)(e_0) = a_0e_0 + a_1e_1 + \dots + a_ne_n,$$

so dass aus der linearen Unabhängigkeit der e_i sofort folgt, dass φ von keinem Polynom $P \neq 0$ annulliert wird. Lineare Abbildungen auf endlichdimensionalen Vektorräumen annullieren alle das charakteristische Polynom ihrer Matrix.

Ebenso wie für das charakteristische Polynom haben ähnliche Matrizen das gleiche Minimalpolynom:

Satz 5.7.10. *Ähnliche Matrizen haben das gleiche Minimalpolynom.*

Beweis. Seien $A, B \in M_n(\mathbb{K})$ ähnlich. Da das Minimalpolynom μ_A eindeutig durch das Ideal I_A definiert ist, reicht es $I_A = I_B$ zu zeigen. Sei dazu $S \in \text{GL}_n(\mathbb{K})$ mit $A = S^{-1}BS$. Damit ergibt sich $A^k = (S^{-1}BS)^k = S^{-1}B^kS$ für alle $k \in \mathbb{N}_0$ und daher $p(A) = S^{-1}p(B)S$ für alle $p \in \mathbb{K}[X]$. Insbesondere folgt hieraus, dass $p(A) = 0$ zu $p(B) = 0$ äquivalent ist, also $I_A = I_B$. $\square$

5.7.11. [Berechnung des Minimalpolynoms] Wir wissen bereits, dass das Minimalpolynom μ_A einer Matrix A jedes andere Polynom teilt, das A annulliert. Nach dem Satz von Cayley–Hamilton teilt das Minimalpolynom das charakteristische Polynom χ_A .

Im Fall $\mathbb{K} = \mathbb{C}$ können wir das charakteristische Polynom in Linearfaktoren zerlegen:

$$\chi_A(X) = (\lambda_1 - X)^{a_1} \dots (\lambda_r - X)^{a_r},$$

wobei a_j die algebraische Vielfachheit von λ_j ist. Da μ_A das Polynom χ_A teilt, ist jede Nullstelle von μ_A auch eine von χ_A und wir erhalten die Gestalt

$$\mu_A(X) = (X - \lambda_1)^{k_1} \dots (X - \lambda_r)^{k_r}$$

mit $k_i \leq a_i$, da μ_A ein Teiler von χ_A ist. Um μ_A zu berechnen, probiert man nun aus, welche Polynome der Gestalt

$$(X - \lambda_1)^{k_1} \dots (X - \lambda_r)^{k_r}$$

die Matrix A annullieren und hat $k_1, \dots, k_r$ jeweils so zu bestimmen, dass sie minimal sind.

Für $\mathbb{K} = \mathbb{R}$ ist die Situation komplizierter, denn hier muss man quadratische Faktoren verwenden.

Man hat ein wenig mehr Information, denn der folgende Satz zeigt für $\mathbb{K} = \mathbb{C}$, dass obige Exponenten k_i alle mindestens 1 sein müssen.

Satz 5.7.12. *Sei A eine Matrix. Dann haben das charakteristische Polynom χ_A und das Minimalpolynom μ_A die gleichen Nullstellen.*

Beweis. Aus dem Satz von Cayley–Hamilton und Satz 5.7.8 folgt, dass μ_A das charakteristische Polynom χ_A teilt. Insbesondere ist jede Nullstelle von μ_A auch eine von χ_A , also ein Eigenwert.

Sei umgekehrt $\lambda \in \mathbb{K}$ eine Nullstelle von χ_A , also ein Eigenwert. Weiter sei $v \in \mathbb{K}^n$ ein Eigenvektor zu λ . Dann folgt aus $A^k v = \lambda^k v$ für alle $k \in \mathbb{N}$ die Beziehung

$$0 = \mu_A(A)v = \mu_A(\lambda) \cdot v.$$

Also ist $\mu_A(\lambda) = 0$. $\square$

Beispiel 5.7.13. Sei

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix} \in M_3(\mathbb{R}).$$

Man berechnet leicht das charakteristische Polynom von A als $\chi_A(X) = (2 - X)^3$. Das Minimalpolynom ist $\mu_A(X) = (X - 2)^2$. In der Tat ist

$$A - 2 \cdot \mathbf{1} = \begin{pmatrix} -1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hieraus folgt schnell $(A - 2 \cdot \mathbf{1})^2 = 0$, und wegen $A - 2 \cdot \mathbf{1} \neq 0$ ist $\mu_A(X) = (X - 2)^2$.

Aufgabe 5.7.14. Sei $A \in M_n(\mathbb{K})$ mit $A^d = 0$ für ein $d \in \mathbb{N}$. Zeige: Die Matrix A ist ähnlich zu einer strikt oberen Dreiecksmatrix $B = (b_{ij})$, d.h., $b_{ij} = 0$ für $i \geq j$. Hinweis: Man hat eine Basis geeignet zu wählen. Dazu betrachte man die Unterräume $V_k := \text{im}(A^{d-k})$ mit

$$V_0 = \{0\} \subseteq V_1 \subseteq \dots \subseteq V_d = \mathbb{K}^n$$

und wähle eine Basis von $\mathbb{K}^n$ so, dass man mit einer Basis von V_1 beginnt, dann zu einer Basis von V_2 ergänzt, etc.

Aufgabe 5.7.15. 1. Sei $A \in M_n(\mathbb{K})$ mit $A^d = 0$. Dann ist $\chi_A(X) = (-1)^n X^n$. Was ist das Minimalpolynom von A ? Hinweis: Der Fall $\mathbb{K} = \mathbb{C}$ ist etwas leichter.

2. Sei $A \in M_n(\mathbb{K})$ und $\lambda \in \mathbb{K}$ mit $(A - \lambda \cdot \mathbf{1})^d = 0$. Dann ist $\chi_A(X) = (\lambda - X)^n$. Was ist das Minimalpolynom von A ?

Aufgabe 5.7.16. Sei $0 \neq A \in M_n(\mathbb{K})$. Zeigen Sie, dass der Grad d des Minimalpolynoms mit der Dimension des Unterraums

$$\text{span}\{A^k : k \in \mathbb{N}_0\} \subseteq M_n(\mathbb{K})$$

übereinstimmt. Folgern Sie hieraus, dass die Dimension dieses Unterraums maximal n ist.

Kapitel 6

Die Jordansche Normalform

In diesem Kapitel befassen wir uns mit folgender Frage: wie können wir entscheiden, ob zwei $n \times n$ -Matrizen A und B ähnlich sind, d.h. ob es eine invertierbare Matrix S gibt, so dass $S^{-1}AS = B$ ist. Wir wissen bereits, dass wenn A und B beide diagonalisierbar sind, es genügt zu prüfen, dass beide Matrizen dieselben Eigenwerte mit den gleichen Vielfachheiten haben. Wie wir aber gesehen haben, sind nicht alle Matrizen diagonalisierbar. Wir werden zeigen, dass jede komplexe Matrix einer Matrix in sogenannter *Jordanscher Normalform*¹ ähnlich ist, also einer Matrix mit nur wenigen Einträgen außerhalb der Diagonalen und für die Ähnlichkeit genauso schnell festgestellt werden kann wie für Diagonalmatrizen.

Unabhängig von Ähnlichkeitstests können Jordansche Normalformen auch zur Berechnung von Matrixpotenzen sinnvoll sein und damit auch zur Lösung von linearen Rekursionen und Differentialgleichungen.

6.1 Jordankästchen

In diesem Abschnitt ist $\mathbb{K}$ ein beliebiger Körper.

Definition 6.1.1. Ein *Jordanblock* oder eine *elementare Jordanmatrix* ist eine $m \times m$ -Matrix der Form

$$J_{m,\lambda} := \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in M_m(\mathbb{K})$$

Einige Spezialfälle von Jordanblöcken:

$$\begin{aligned} J_{1,\lambda} &= (\lambda), & J_{2,\lambda} &= \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, & J_{3,\lambda} &= \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \\ J_{2,0} &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, & J_{3,0} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Lemma 6.1.2. Ein Jordanblock $J := J_{m,\lambda}$ mit $m > 1$ ist nicht diagonalisierbar. Sein Eigenraum zum Eigenwert λ ist eindimensional und das Minimalpolynom ist $\mu_J(X) = (X - \lambda)^m$.

¹Camille Jordan (1838–1922), frz. Mathematiker in Paris. Er schrieb in den 1870er Jahren das erste Lehrbuch, das die Galoissche Theorie der Polynomgleichungen behandelte. Hierdurch machte er die Galoisschen Ideen der mathematischen Fachwelt zugänglich, was der noch jungen Gruppentheorie wichtige Impulse gab. Unter anderem regte es Sophus Lie an, über eine Galoissche Theorie der Differentialgleichungen nachzudenken, auf der die moderne Theorie der Symmetrien von Differentialgleichungen beruht.

Beweis. (a) Das charakteristische Polynom $\chi_J(X)$ von $J := J_{m,\lambda}$ ist $\det(J - X\mathbf{1}) = (\lambda - X)^m$, da J eine obere Dreiecksmatrix ist. Also ist λ der einzige Eigenwert; er hat die algebraische Vielfachheit $a_\lambda = m$.

Die Eigenvektoren x zum Eigenwert λ erfüllen die Gleichung

$$0 = (J - \lambda \cdot \mathbf{1})x = (x_2, x_3, \dots, x_n, 0)^\top,$$

das bedeutet $x_2 = 0, x_3 = 0, \dots, x_m = 0$ und $x_1 \in \mathbb{K}$ ist beliebig. Also ist die geometrische Vielfachheit g_λ des Eigenwertes λ gleich 1. Der zugehörige Eigenraum $\mathbb{K}e_1$ ist eindimensional.

Es gibt also keine Basis von Eigenvektoren für J , d.h. J ist nicht diagonalisierbar für $n \geq 2$.

(b) Da das Minimalpolynom μ_J ein Teiler von χ_J ist, hat es die Gestalt $(X - \lambda)^k$ für ein $k \in \{1, \dots, m\}$. Für die Standardbasis $e_1, \dots, e_m$ von $\mathbb{K}^m$ gilt

$$(J - \lambda \cdot \mathbf{1})^k e_m = e_{m-k}.$$

Für $k < m$ ergibt sich insbesondere $(J - \lambda \cdot \mathbf{1})^k \neq 0$ und weiter $(J - \lambda \cdot \mathbf{1})^m = 0$ folgt. Folglich ist das Minimalpolynom $\mu_J(X) = (X - \lambda)^m$. $\square$

Definition 6.1.3. Wir sagen, eine Matrix A besitzt *Jordansche Normalform*, wenn sie folgende Blockgestalt besitzt:

$$A = \begin{pmatrix} \boxed{J_1} & & & \\ & \boxed{J_2} & & \\ & & \ddots & \\ & & & \boxed{J_t} \end{pmatrix}$$

wobei jedes der $J_1, J_2, \dots, J_t$ ein Jordanblock ist.

Hauptsatz über die Jordansche Normalform

Satz 6.1.4. Sei V ein n -dimensionaler Vektorraum über dem algebraisch abgeschlossenen Körper $\mathbb{K}$ und $\varphi \in \text{End}(V)$. Dann existiert eine Basis $B = (v_1, v_2, \dots, v_n)$ von V , so dass die Matrix $[\varphi]_B$ von φ bzgl. B Jordansche Normalform besitzt. Insbesondere gilt dies für $\mathbb{K} = \mathbb{C}$.

Solch eine Basis B heißt *Jordanbasis* für φ .

Satz 6.1.5. Sei $A \in M_n(\mathbb{K})$ und $\mathbb{K}$ algebraisch abgeschlossen. Dann ist A ähnlich zu einer Matrix A' in Jordanscher Normalform. Sind A' und A'' ähnlich und in Jordanscher Normalform, so unterscheiden sie sich nur durch eine Permutation der Jordanblöcke. In diesem Sinn ist A' eindeutig bis auf eine Permutation der Jordanblöcke.

Bemerkung 6.1.6. Im Lichte der Folgerung 6.1.5 erhalten wir den folgenden Test für die Ähnlichkeit von Matrizen mit Einträgen in einem algebraisch abgeschlossenen Körper: um zu entscheiden, ob es ein S gibt, so dass $S^{-1}AS = B$ ist, berechne die Jordansche Normalform von A und B und prüfe, ob sie bis auf eine Permutation der Blöcke übereinstimmen.

Wie bereits bemerkt, haben wir damit auch die Möglichkeit, einige der Anwendungen der Diagonalisierbarkeit auf nichtdiagonalisierbare Matrizen auszudehnen: es gibt z.B. relativ leichte Formeln für A^n und $\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$, falls A in Jordanscher Normalform vorliegt (siehe Abschnitt 6.7).

Beispiel 6.1.7. Jede Diagonalmatrix ist in Jordanscher Normalform mit $t = n$: Die Jordanblöcke enthalten lediglich den Eintrag λ_i :

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Weitere Beispiele von Matrizen in Jordanscher Normalform:

$$\begin{pmatrix} \boxed{1} & & & \\ & \boxed{0} & & \\ & & \boxed{2} & \\ & & & \boxed{0} \\ & & & & \boxed{3} \end{pmatrix}, \begin{pmatrix} \boxed{1} & \boxed{1} & & \\ & \boxed{0} & \boxed{1} & \\ & & & \boxed{2} & \boxed{1} \\ 0 & & & & \boxed{0} & \boxed{2} \end{pmatrix}, \begin{pmatrix} \boxed{2} & \boxed{1} & \boxed{0} & & \\ & \boxed{0} & \boxed{2} & \boxed{1} & \\ & & \boxed{0} & \boxed{0} & \boxed{2} & \\ & & & \boxed{0} & \boxed{0} & \boxed{2} \end{pmatrix}, \begin{pmatrix} \boxed{2} & \boxed{1} & & & \\ & \boxed{0} & \boxed{2} & & \\ & & & \boxed{2} & \\ 0 & & & & \boxed{2} \end{pmatrix}.$$

Die folgenden Matrizen sind *nicht* in Jordanscher Normalform:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Bemerkung 6.1.8. Ist A in Jordanscher Normalform, so sind die Einträge λ_i in der Diagonalen die Eigenwerte von A , jeder mit der Anzahl seiner algebraischen Vielfachheit $a_i := a_{\lambda_i}$. Verschiedene Jordanblöcke können zu demselben Eigenwert λ_i gehören: **Die Zahl der Jordanblöcke zum Eigenwert λ_i ist die geometrische Vielfachheit $g_i := g_{\lambda_i}$ des Eigenwertes λ_i .**

Bemerkung 6.1.9. Aus Lemma 6.1.2 erhalten wir leicht, dass das Minimalpolynom einer Matrix A in Jordanscher Normalform, die aus Blöcken $J_{m_i, \lambda}$ besteht, gegeben ist durch

$$(X - \lambda)^p \quad \text{mit} \quad p := \max\{m_1, \dots, m_t\}.$$

Ein Polynom $Q \in \mathbb{K}[X]$ annulliert nämlich genau dann die Matrix A , wenn es alle Jordanblöcke $J_{m_i, \lambda}$ annulliert, was nach Lemma 6.1.2 heißt, dass es von $(X - \lambda)^p$ geteilt wird.

Besteht A aus Jordanblöcken zu verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_r$ und ist p_i die maximale Größe eines Jordanblocks zu dem Eigenwert λ_i , so sieht man mit der gleichen Argumentation wie im Beweis von Lemma 6.1.2, dass das Minimalpolynom von A gegeben ist durch

$$\mu_A(X) = \prod_{i=1}^r (X - \lambda_i)^{p_i}.$$

Aufgabe 6.1.10. Wir betrachten eine Matrix der Gestalt

$$J_{m, \lambda}(x) := \begin{pmatrix} \lambda & x & & 0 \\ & \lambda & \ddots & \\ & & \ddots & x \\ 0 & & & \lambda \end{pmatrix} \in M_m(\mathbb{K}).$$

Zeigen Sie: Ist $x \neq 0$, so ist $J_{m, \lambda}(x)$ ähnlich zu $J_{m, \lambda}(1) = J_{m, \lambda}$.

6.2 Hauptvektoren und verallgemeinerte Eigenräume

Bemerkung 6.2.1. [Einleitende Betrachtungen] Angenommen, $B = (v_1, \dots, v_m, v_{m+1}, \dots, v_n)$ ist eine Basis des Vektorraums V , so dass die Matrix $A = [\varphi]_B$ von $\varphi \in \text{End}(V)$ bzgl. B Jordansche Normalform besitzt, wobei die ersten m Basisvektoren $v_1, \dots, v_m$ zum ersten Jordanblock

$$J_1 = J_{m, \lambda}$$

gehören. Dann ergibt sich

$$\begin{array}{ll} \varphi(v_1) = \lambda v_1 & Av_1 = \lambda v_1 \\ \varphi(v_2) = v_1 + \lambda v_2 & Av_2 = v_1 + \lambda v_2 \\ \vdots & \vdots \\ \varphi(v_m) = v_{m-1} + \lambda v_m & Av_m = v_{m-1} + \lambda v_m, \end{array}$$

oder anders geschrieben

$$\begin{array}{ll} (\varphi - \lambda \text{id})v_1 = 0 & (A - \lambda \cdot \mathbf{1})v_1 = 0 \\ (\varphi - \lambda \text{id})v_2 = v_1 & (A - \lambda \cdot \mathbf{1})v_2 = v_1 \\ \vdots & \vdots \\ (\varphi - \lambda \text{id})v_m = v_{m-1} & (A - \lambda \cdot \mathbf{1})v_m = v_{m-1}. \end{array} \quad (6.1)$$

Definition 6.2.2. [Jordanketten] Eine Folge $v_1, v_2, \dots, v_m$ von Vektoren in V heißt *Jordankette* für $\varphi \in \text{End}(V)$, wenn (6.1) mit $v_1 \neq 0$ erfüllt ist. In diesem Fall ist v_1 ein Eigenvektor zum Eigenwert λ . Weiterhin gilt für $j = 0, 1, \dots, m-1$

$$(\varphi - \lambda \text{id})^j v_m = v_{m-j} \quad \text{und} \quad (\varphi - \lambda \text{id})^m v_m = 0.$$

Eine Folge $v_1, v_2, \dots, v_m$ von Vektoren in $\mathbb{K}^n$ heißt *Jordankette* der Matrix $A \in M_n(\mathbb{K})$ zum Eigenwert λ , wenn (6.1) mit $v_1 \neq 0$ erfüllt ist.

Lemma 6.2.3. *Bilden die Vektoren $v_1, \dots, v_m$ eine Jordankette für φ und den Eigenwert λ , so sind sie linear unabhängig.*

Beweis. Zum Erreichen eines Widerspruchs nehmen wir $\sum_{i=1}^m \alpha_i v_i = 0$ an, wobei nicht alle $\alpha_i = 0$ seien. Sei $k = \max\{i: \alpha_i \neq 0\}$. Beachte, dass $k \geq 2$ gelten muss. Dann ist

$$(\varphi - \lambda \text{id})^{k-1} v_k = v_1 \quad \text{und} \quad (\varphi - \lambda \text{id})^{k-1} v_j = 0 \quad \text{für} \quad j < k,$$

also

$$0 = (\varphi - \lambda \text{id})^{k-1} \left(\sum_{i=1}^m \alpha_i v_i \right) = \alpha_k v_1,$$

woraus $\alpha_k = 0$ folgt. Dieser Widerspruch zeigt die lineare Unabhängigkeit der v_i . $\square$

Definition 6.2.4. [Verallgemeinerte Eigenvektoren] Sei $\varphi \in \text{End}(V)$. Ein Vektor $0 \neq v \in V$ heißt *verallgemeinerter Eigenvektor* (oder auch *Hauptvektor*) von φ zum Eigenwert λ von φ , falls es eine natürliche Zahl m gibt mit

$$(\varphi - \lambda \text{id})^m v = 0.$$

Das kleinste m mit dieser Eigenschaft heißt die *Stufe* des verallgemeinerten Eigenvektors v . Man beachte, dass Eigenvektoren verallgemeinerte Eigenvektoren der Stufe 1 sind.

Die Stufe m ist kleiner oder gleich $\dim V$: Ist nämlich v ein verallgemeinerter Eigenvektor der Stufe m , so ist

$$v_1 := (\varphi - \lambda \text{id})^{m-1} v, \quad \dots \quad v_{m-1} := (\varphi - \lambda \text{id})v, \quad v_m := v$$

eine Jordankette, also linear unabhängig und damit $m \leq \dim V$. (Ähnlich wie im Beweis von Satz 5.4.8 zeigt man, dass $m \leq a_\lambda$ gilt, wobei a_λ die algebraische Vielfachheit von λ ist).

Wir schreiben

$$V^\lambda(\varphi) := \bigcup_{m \in \mathbb{N}} \ker(\varphi - \lambda \text{id})^m$$

für die Menge der verallgemeinerten Eigenvektoren von φ zum Eigenwert λ , zusammen mit 0. Der Raum $V^\lambda(\varphi)$ heißt *verallgemeinerter Eigenraum von φ zum Eigenwert λ* .

Analog definiert man für eine Matrix $A \in M_n(\mathbb{K})$ verallgemeinerte Eigenvektoren als verallgemeinerte Eigenvektoren der linearen Abbildung $L_A: \mathbb{K}^n \rightarrow \mathbb{K}^n, x \mapsto Ax$ und überträgt die Begriffe entsprechend.

Das folgende Lemma wird im folgenden permanent benutzt werden.

Vertauschungslemma

Lemma 6.2.5. *Vertauschen φ und $\psi \in \text{End}(V)$ miteinander, so sind die verallgemeinerten Eigenräume $V^\lambda(\varphi)$ und die Eigenräume $V_\lambda(\varphi)$ invariant unter ψ , d.h. aus $v \in V^\lambda(\varphi)$ folgt $\psi(v) \in V^\lambda(\varphi)$. Insbesondere ist $V^\lambda(\varphi)$ invariant unter φ .*

Beweis. Sei $v \in V$ mit $(\varphi - \lambda \text{id})^m v = 0$. Da ψ mit φ vertauscht, vertauscht es auch mit $\varphi - \lambda \text{id}$ und somit auch mit der Potenz $(\varphi - \lambda \text{id})^m$. Damit erhalten wir

$$(\varphi - \lambda \text{id})^m(\psi(v)) = \psi \circ (\varphi - \lambda \text{id})^m(v) = 0.$$

Hieraus folgt die Behauptung. □

Wir werden nun den Hauptsatz beweisen, indem wir zuerst zeigen, dass für $\varphi \in \text{End}(V)$, $\mathbb{K}$ algebraisch abgeschlossen und $\dim V < \infty$ der Raum V die direkte Summe der verallgemeinerten Eigenräume ist.

Lemma 6.2.6. *Sei V ein Vektorraum und $E_1, \dots, E_n \in \text{End}(V)$ Elemente mit*

$$E_1 + \dots + E_n = \text{id}_V \quad \text{und} \quad E_i E_j = \delta_{ij} E_i.$$

Dann ist V die direkte Summe der Unterräume $V_i := \text{im}(E_i)$ und $\ker(E_i) = \sum_{j \neq i} V_j$.

Beweis. Zunächst erhalten wir für jedes $v \in V$ die Relation

$$v = \text{id}_V(v) = \sum_{j=1}^m E_j(v) \in V_1 + \dots + V_m, \quad \text{also} \quad V = V_1 + \dots + V_m.$$

Wegen $E_j E_i = 0$ für $j \neq i$ ist $V_i \subseteq \ker E_j$ und für $v \in V_i$ existiert ein $w \in V$ mit $v = E_i(w)$, was $E_i(v) = E_i^2(w) = E_i(w) = v$ zur Folge hat.

Seien nun $v_j \in V_j$ mit $\sum v_j = 0$. Dann gilt also

$$0 = E_i(0) = E_i\left(\sum_j v_j\right) = E_i(v_i) + \sum_{j \neq i} E_i(v_j) = E_i(v_i) = v_i.$$

Folglich ist die Summe $V = V_1 + \dots + V_m$ direkt. □

Das folgende Zerlegungslemma liefert eine Möglichkeit, zu einer gegebenen linearen Abbildung $\varphi \in \text{End}(V)$ eine direkte Zerlegung von V zu finden, die unter φ invariant ist:

Zerlegungslemma

Lemma 6.2.7. *Sei V ein Vektorraum, $\varphi \in \text{End}(V)$ und $f \in \mathbb{K}[X]$ ein Polynom mit $f(\varphi) = 0$. Ist $f = f_1 f_2 \dots f_n$ eine Zerlegung in paarweise teilerfremde Faktoren, so gilt*

$$V = V_1 \oplus \dots \oplus V_n \quad \text{für} \quad V_j := \ker(f_j(\varphi)), j = 1, \dots, n.$$

Die Unterräume V_j sind φ -invariant und für $i \neq j$ ist $f_j(\varphi)|_{V_i}$ invertierbar.

Beweis. Aus Aufgabe 5.6.9(c) folgt für die Polynome $q_j := f_1 \cdots f_{j-1} f_{j+1} \cdots f_n$ die Relation $\text{ggT}(q_1, \dots, q_n) = 1$. Also existieren $r_1, \dots, r_n \in \mathbb{K}[X]$ mit

$$1 = r_1 q_1 + \cdots + r_n q_n. \quad (6.2)$$

Sei $E_j := (r_j q_j)(\varphi) \in \text{End}(V)$. Wenden wir die Relation (6.2) auf den Endomorphismus φ an, so erhalten wir

$$E_1 + \cdots + E_n = \text{id}_V.$$

Die Abbildungen E_i vertauschen mit allen linearen Abbildungen, die mit φ vertauschen, denn sie sind Polynome in φ . Für $i \neq j$ ist f ein Teiler von $q_i q_j$, also

$$E_i E_j = (r_1 q_1 r_2 q_2)(\varphi) = (r_1 r_2 q_1 q_2)(\varphi) = 0.$$

Daher ist auch

$$E_j = E_j \mathbf{1} = E_j (E_1 + \cdots + E_n) = E_j^2.$$

Aus Lemma 6.2.6 folgt daher

$$V = V_1 \oplus \cdots \oplus V_n \quad \text{mit} \quad V_j = \text{im}(E_j).$$

Da die E_j mit φ vertauschen, sind die Unterräume V_j unter φ invariant. Für $i \neq j$ erhalten wir aus der Relation

$$\text{id}_{V_j} = E_j|_{V_j} = (r_j q_j)(\varphi)|_{V_j} = f_i(\varphi)|_{V_j} (r_j q_j / f_i)(\varphi)|_{V_j} = (r_j q_j / f_i)(\varphi)|_{V_j} f_i(\varphi)|_{V_j},$$

dass $f_i(\varphi)|_{V_j}$ invertierbar ist mit der Inversen $(r_j q_j / f_i)(\varphi)|_{V_j}$. Andererseits ist

$$f_j(\varphi)V_j = f_j(\varphi)(q_j r_j)(\varphi)V = f(\varphi)r_j(\varphi)V = \{0\},$$

also

$$V_j \subseteq \ker(f_j(\varphi)).$$

Da $f_j(\varphi)$ alle Summanden der direkten Summe $V = V_1 \oplus \cdots \oplus V_n$ invariant lässt und auf V_i , $i \neq j$, invertierbar ist, erhalten wir $V_j = \ker(f_j(\varphi))$. Aus der Konstruktion folgt sofort $f_j(\varphi)|_{V_j} = f_j(\varphi)|_{V_j} = 0$. $\square$

Satz 6.2.8. (Zerlegung in verallgemeinerte Eigenräume) *Sei V ein endlichdimensionaler Vektorraum über dem algebraisch abgeschlossenen Körper $\mathbb{K}$ und $\varphi \in \text{End}(V)$ mit den paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_m$. Dann ist*

$$V = V^{\lambda_1}(\varphi) \oplus \cdots \oplus V^{\lambda_m}(\varphi),$$

d.h. V ist direkte Summe der verallgemeinerten Eigenräume von φ .

Beweis. Da $\mathbb{K}$ algebraisch abgeschlossen ist, zerfällt das charakteristische Polynom $\chi_\varphi(X)$ in Linearfaktoren:

$$\chi_\varphi(X) = (X - \lambda_1)^{a_1} (X - \lambda_2)^{a_2} \cdots (X - \lambda_m)^{a_m}$$

und nach dem Satz von Cayley–Hamilton gilt $\chi_\varphi(\varphi) = 0$. Wir betrachten nun für $f := \chi_\varphi$ die Zerlegung $f = f_1 \cdots f_m$ mit $f_j(X) := (X - \lambda_j)^{a_j}$. Jedes Polynom, das f_i und f_j für $i \neq j$ teilt, hat keine Nullstellen, ist also konstant, d.h. f_i und f_j sind teilerfremd.

Wir können daher Lemma 6.2.7 anwenden und erhalten eine φ -invariante Zerlegung

$$V = V_1 \oplus \cdots \oplus V_m \quad \text{mit} \quad V_j = \ker(f_j(\varphi)) = \ker((\varphi - \lambda_j \text{id})^{a_j}) \subseteq V^{\lambda_j}(\varphi),$$

so dass $f_j(\varphi)$ auf allen anderen Räumen V_i , $i \neq j$, invertierbar ist. Ist

$$v = v_1 + \cdots + v_m \in V^{\lambda_j}(\varphi), \quad v_j \in V_j,$$

so ist

$$0 = f_j(\varphi)^N v = \sum_{i=1}^m f_j(\varphi)^N v_i$$

für ein N . Wegen der Direktheit der Summe ist dann auch $f_j(\varphi)^N v_i = 0$ für alle $i \neq j$, also $v_i = 0$, da $f_j(\varphi)|_{V_i}$ invertierbar ist. Daher ist $v \in V_j$ und somit haben wir die Gleichheit $V_j = V^{\lambda_j}(\varphi)$. $\square$

Aufgabe 6.2.9. Sei $\varphi \in \text{End}(V)$ diagonalisierbar. Zeigen Sie, dass $\psi \in \text{End}(V)$ genau dann mit φ vertauscht, wenn ψ alle Eigenräume $V_\lambda(\varphi)$ invariant lässt.

Aufgabe 6.2.10. Zeigen Sie, dass die Stufe m eines Hauptvektors $v \in V^\lambda(\varphi)$ der Ungleichung $m \leq a_\lambda$ genügt, wobei a_λ die algebraische Vielfachheit von λ ist.

6.3 Die Jordan-Zerlegung

Die Zerlegung in verallgemeinerte Eigenräume hat eine wichtige Konsequenz: die Jordan-Zerlegung $\varphi = \varphi_s + \varphi_n$ wobei φ_s diagonalisierbar ist, φ_n nilpotent und beide miteinander vertauschen. Hierbei nennen wir einen Endomorphismus $\varphi \in \text{End}(V)$ *nilpotent*, wenn $\varphi^N = 0$ für ein $N \in \mathbb{N}$ gilt.

Lemma 6.3.1. Ist $\varphi \in \text{End}(V)$ diagonalisierbar und $W \subseteq V$ invariant unter φ , so ist auch $\varphi|_W$ diagonalisierbar.

Beweis. Wir haben zu zeigen, dass W von Eigenvektoren von φ aufgespannt wird, wenn dies für V der Fall ist. Sei dazu $w \in W$. Da φ diagonalisierbar ist, können wir w schreiben als

$$w = v_1 + \dots + v_n \quad \text{wobei} \quad v_i \in V_{\lambda_i}(\varphi)$$

für paarweise verschiedene Eigenwerte λ_i von φ gilt. Es reicht nun zu zeigen, dass alle Summanden v_j in W enthalten sind.

Wir zeigen durch Induktion nach n , dass für eine Summe $v_1 + \dots + v_n$ von Eigenvektoren zu paarweise verschiedenen Eigenwerten, die in W enthalten ist, auch alle Summanden in W enthalten sind. Für $n = 1$ ist nichts zu zeigen. Ist $n > 1$, so erhalten wir

$$W \ni \varphi(v_1 + \dots + v_n) - \lambda_n(v_1 + \dots + v_n) = \lambda_1 v_1 + \dots + \lambda_n v_n - \lambda_n(v_1 + \dots + v_n) = \sum_{j=1}^{n-1} (\lambda_j - \lambda_n) v_j.$$

Aus der Induktionsannahme folgt daher $(\lambda_j - \lambda_n)v_j \in W$ für $j < n$, also auch $v_j \in W$, da $\lambda_j \neq \lambda_n$ ist. Damit ist auch

$$v_n = (v_1 + \dots + v_n) - (v_1 + \dots + v_{n-1}) \in W. \quad \square$$

Lemma 6.3.2. Sei V ein $\mathbb{K}$ -Vektorraum und $\varphi, \psi \in \text{End}(V)$ mit $\varphi\psi = \psi\varphi$. Dann gilt:

- (i) Sind φ und ψ diagonalisierbar, so auch $\varphi + \psi$.
- (ii) Sind φ und ψ nilpotent, so auch $\varphi + \psi$.

Beweis. (i) Nach Lemma 6.2.5 sind alle Eigenräume $V_\lambda(\varphi)$ unter ψ invariant und nach Lemma 6.3.1 ist $\psi|_{V_\lambda(\varphi)}$ diagonalisierbar. Also ist jeder Eigenvektor von φ Summe gemeinsamer Eigenvektoren von φ und ψ . Da andererseits $V = \sum_\lambda V_\lambda(\varphi)$ ist, folgt daraus, dass V von gemeinsamen Eigenvektoren von φ und ψ aufgespannt wird. Insbesondere ist $\varphi + \psi$ diagonalisierbar.

(ii) Sei $\varphi^m = 0 = \psi^n$. Dann folgt aus der binomischen Formel

$$(\varphi + \psi)^k = \sum_{i+j=k} \binom{k}{i} \varphi^i \psi^j.$$

Ist $k \geq n + m - 1$, so ist entweder $i \geq m$ oder $j \geq n$, d.h. alle Summanden verschwinden. Folglich ist $(\varphi + \psi)^k = 0$. $\square$

Jordan-Zerlegung

Theorem 6.3.3. Sei $\mathbb{K}$ algebraisch abgeschlossen und V ein endlichdimensionaler $\mathbb{K}$ -Vektorraum. Zu jedem $\varphi \in \text{End}(V)$ existieren eine diagonalisierbare lineare Abbildung φ_s und eine nilpotente lineare Abbildung φ_n , so dass

- (i) $\varphi = \varphi_s + \varphi_n$.
- (ii) Die Eigenräume von φ_s sind genau die verallgemeinerten Eigenräume von φ , d.h. $V_\lambda(\varphi_s) = V^\lambda(\varphi)$ für alle $\lambda \in \mathbb{K}$.
- (iii) Vertauscht $\psi \in \text{End}(V)$ mit φ , so auch mit φ_s und φ_n .
- (iv) (Eindeutigkeit der Jordan-Zerlegung) Sind $\varphi'_s, \varphi'_n \in \text{End}(V)$, φ'_s diagonalisierbar und φ'_s nilpotent mit $\varphi'_s \varphi'_n = \varphi'_n \varphi'_s$ und $\varphi = \varphi'_s + \varphi'_n$, so ist $\varphi_s = \varphi'_s$ und $\varphi_n = \varphi'_n$.

Beweis. (i),(ii) Seien $V_j = V^{\lambda_j}(\varphi)$ die verallgemeinerten Eigenräume von φ zu den paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_m$. Wir wissen schon aus Satz 6.2.8, dass

$$V = V_1 \oplus \dots \oplus V_m$$

eine direkte Summe ist. Seien $E_j: V \rightarrow V$ die Projektionen auf die Komponenten V_j und

$$\varphi_s := \sum_{j=1}^m \lambda_j E_j.$$

Dann ist $V_j = V^{\lambda_j}(\varphi) = V_{\lambda_j}(\varphi_s)$, denn für $v \in V_k$ ist

$$\varphi_s(v) = \sum_{j=1}^m \lambda_j E_j v = \lambda_k v.$$

Für ein allgemeines Element $v = \sum_{j=1}^m v_j \in V$ mit $v_j \in V_j$ ist

$$\varphi(\varphi_s(v)) = \varphi\left(\sum_{j=1}^m \lambda_j v_j\right) = \sum_{j=1}^m \lambda_j \varphi(v_j) = \varphi_s\left(\sum_{j=1}^m \varphi(v_j)\right) = \varphi_s(\varphi(v)),$$

da $\varphi(v_j) \in V_j$ ist. Daher vertauscht φ mit φ_s . Insbesondere vertauscht φ mit $\varphi_n := \varphi - \varphi_s$. Dieser Endomorphismus lässt auch alle V_j invariant, und

$$\varphi_n|_{V_j} = (\varphi - \lambda_j \text{id})|_{V_j} \quad (6.3)$$

ist nilpotent, da $V_j = V^{\lambda_j}(\varphi)$ ist.

(iii) Vertauscht ψ mit φ , so lässt ψ die verallgemeinerten Eigenräume V_j von φ invariant (Lemma 6.2.5), vertauscht also mit allen E_j (Aufgabe 6.2.9), und somit auch mit φ_s . Also vertauscht ψ auch mit $\varphi_n = \varphi - \varphi_s$.

(iv) Da φ'_n und φ'_s miteinander vertauschen, vertauschen sie auch mit $\varphi = \varphi'_s + \varphi'_n$. Wegen (iii) vertauschen sie daher auch mit φ_n und φ_s . Aus $\varphi = \varphi_s + \varphi_n = \varphi'_s + \varphi'_n$ und Lemma 6.3.2 folgt daher, dass

$$\varphi_s - \varphi'_s = \varphi'_n - \varphi_n$$

sowohl nilpotent als auch diagonalisierbar ist. Da alle Eigenwerte 0 sind, ist $0 = \varphi_s - \varphi'_s = \varphi'_n - \varphi_n$, also $\varphi_n = \varphi'_n$ und $\varphi_s = \varphi'_s$. $\square$

Definition 6.3.4. Die Zerlegung $\varphi = \varphi_s + \varphi_n$ heißt *Jordan-Zerlegung* von $\varphi \in \text{End}(V)$. Die Abbildung φ_s heißt *diagonalisierbare* und φ_n *nilpotente Jordan-Komponente* von φ .

Beispiel 6.3.5. Ist $M = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ ein Jordan-Kästchen, so ist die Jordan-Zerlegung von M natürlich gegeben durch

$$M = \underbrace{\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}}_{M_s} + \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{M_n}.$$

Die Matrix $M = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ dargegeben ist diagonalisierbar und somit $M = M_s$. In diesem Fall ist also

$$M = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$$

nicht die Jordan-Zerlegung, obwohl der erste Summand diagonalisierbar und der zweite nilpotent ist. Man verifiziert leicht, dass beide Summanden nicht miteinander vertauschen.

6.4 Der Beweis des Hauptsatzes

In diesem Abschnitt ist V zunächst ein Vektorraum über einem beliebigen Körper $\mathbb{K}$ sofern nicht explizit die Einschränkung gemacht wird, dass $\mathbb{K}$ algebraisch abgeschlossen ist. Weiter ist $\varphi \in \text{End}(V)$ ein Endomorphismus des endlichdimensionalen $\mathbb{K}$ -Vektorraums V .

Lemma 6.4.1. *Sei $\mathcal{C} = (v_1^1, \dots, v_{\ell_1}^1, v_1^2, \dots, v_{\ell_2}^2, \dots, v_1^s, \dots, v_{\ell_s}^s)$ eine Familie von s Jordanketten zum Eigenwert λ von φ , d.h.*

$$(\varphi - \lambda \text{id})v_{j+1}^i = v_j^i \quad \text{für } 1 < j \leq \ell_i, \quad \text{und} \quad (\varphi - \lambda \text{id})v_1^i = 0.$$

Sind die Eigenvektoren $v_1^1, \dots, v_1^s$ linear unabhängig, so sind alle Vektoren in $\mathcal{C}$ linear unabhängig.

Beweis. Sei $\sum_{i,j} \alpha_{ij} v_j^i = 0$. Wir müssen zeigen, dass alle α_{ij} verschwinden. Nehmen wir also an, dies sei nicht der Fall. Sei k das Maximum der Stufen j , für die ein $\alpha_{ij} \neq 0$ ist. Wenden wir $(\varphi - \lambda \text{id})^{k-1}$ an, so erhalten wir eine lineare Relation zwischen den Eigenvektoren

$$0 = \sum \alpha_{ik} (\varphi - \lambda \text{id})^{k-1} v_k^i = \sum \alpha_{ik} v_1^i.$$

Hieraus folgt $\alpha_{ik} = 0$ für alle i , im Widerspruch zur Wahl von k . $\square$

Satz 6.4.2. *Für jeden Eigenwert λ von φ besitzt der verallgemeinerte Eigenraum $V^\lambda(\varphi)$ eine Jordanbasis.*

Beweis. Um die Notation zu vereinfachen, setzen wir $W := V^\lambda(\varphi)$ und $\psi := (\varphi - \lambda \text{id})|_W$. Dann ist $\psi^N = 0$ für ein $N \in \mathbb{N}$, das wir minimal annehmen dürfen, denn nach Lemma 6.4.1 ist die Länge von Jordan-Ketten durch die Dimension von W beschränkt. Insbesondere ist also $N \leq \dim W$.

Wir beweisen die Behauptung durch Induktion nach N . Für $N = 1$ ist $\psi = 0$ und alle Jordanketten haben die Länge 1. Jede Basis ist also eine Jordanbasis.

Sei jetzt $N > 1$ und $U := \text{im}(\psi)$. Dann ist trivialerweise $\psi(U) \subseteq U$ und $\psi^{N-1}(U) \subseteq \psi^{N-1}(\psi(W)) = \{0\}$. Die Induktionsvoraussetzung liefert also eine Basis von U , die aus Jordanketten für die Einschränkung $\psi|_U$ besteht:

$$b_j, \psi b_j, \dots, \psi^{r_j} b_j, \quad j = 1, \dots, k, \quad \text{mit} \quad \psi^{r_j+1} b_j = 0 \quad \text{und} \quad \psi^{r_j} b_j \neq 0.$$

Zuerst verlängern wir unsere Ketten etwas: Jedes b_j können wir als $\psi(w_j)$ für ein $w_j \in W$ schreiben. Wir erhalten so Jordanketten

$$w_j, \psi w_j = b_j, \dots, \psi^{r_j+1} w_j = \psi^{r_j} b_j, \quad j = 1, \dots, k,$$

in W .

Die Vektoren $\psi^{r_j+1} w_j = \psi^{r_j} b_j$ liegen in $\ker \psi$ und sind gemäß unsrer Konstruktion linear unabhängig. Wir ergänzen sie nun durch Vektoren $w_{k+1}, \dots, w_n$ zu einer Basis von $\ker \psi$. Jeder dieser neuen Vektoren bildet eine Jordankette der Länge 1. Wir haben nun n Jordanketten in W , für die die jeweiligen Endvektoren eine Basis von $\ker \psi$ bilden. Insbesondere sind sie linear unabhängig und aus Lemma 6.4.1 folgt, dass die Gesamtheit der so konstruierten Jordanketten linear unabhängig ist. Die Dimension des Unterraums, den sie aufspannen, ist

$$\dim \psi(W) + k + (n - k) = \dim \psi(W) + n = \dim \psi(W) + \dim(\ker \psi) = \dim W,$$

wobei wir für die letzte Gleichheit die Dimensionsformel verwendet haben. Also haben wir eine Jordanbasis von W gefunden. $\square$

Beweis. (des Hauptsatzes 6.1.4 über die Jordansche Normalform:) Aus Satz 6.4.2 folgt die Existenz einer Jordanbasis für jeden verallgemeinerten Eigenraum V^{λ_i} (hierbei ist keine Annahme über den Körper notwendig). Da wir nach Satz 6.2.8 eine direkte Zerlegung $V = V^{\lambda_1} \oplus \dots \oplus V^{\lambda_m}$ haben, erhalten wir eine Jordanbasis für V , indem wir Jordanbasen für die verallgemeinerten Eigenräume kombinieren. Ist B eine Jordanbasis von V , so ist die Matrix $[\varphi]_B$ von φ bzgl. B in Jordanscher Normalform. Das folgt aus der Art, wie φ auf Basisvektoren wirkt. $\square$

Auffinden der Jordanschen Normalform einer Matrix A : Fassen wir die entscheidenden Schritte, die auf eine Jordanbasis führen, zusammen.

1. Bestimme die Eigenwerte λ_i von A mit ihren algebraischen Vielfachheiten $a_i := a_{\lambda_i}$.
2. Bestimme für jeden Eigenwert λ_i eine Basis des verallgemeinerten Eigenraums V^{λ_i} , die aus Jordanketten besteht. Hierzu sei $\lambda := \lambda_j$ und $W := V^\lambda$. Hierbei kann man dem konstruktiven Beweis von Satz 6.4.2 folgen, woraus sich ein rekursives Verfahren ergibt. Indem man die Potenzen $(A - \lambda \mathbf{1})^k$ berechnet, bestimmt man zunächst das minimale N mit $(A - \lambda \mathbf{1})^N W = 0$, der *Fitting-Index des Eigenwerts* λ . Das ist die Länge der längsten Jordankette. Durch

$$W_k := (A - \lambda \mathbf{1})^{N-k} W, \quad k = 1, 2, \dots, N,$$

erhalten wir eine aufsteigende Folge von Unterräumen

$$\{0\} \neq W_1 \subseteq W_2 \subseteq \dots \subseteq W_N = W,$$

Wir erinnern uns daran, dass das Bild einer Matrix der Unterraum ist, der von den Spalten aufgespannt wird.

Zuerst wählen wir eine Basis $v_1^1, \dots, v_{r_1}^1$ von W_1 und erhält so die Endpunkte der Jordanketten maximaler Länge.

Im zweiten Schritt wählt man Elemente $v_2^j \in W_2$, $j = 1, \dots, r_1$, mit $(A - \lambda \mathbf{1})v_2^j = v_1^j$ und ergänzt durch weitere Elemente

$$v_1^{r_1+1}, \dots, v_1^{r_2} \in W_2 \cap \ker(A - \lambda \mathbf{1})$$

zu einer Basis von W_2 .

Im k . Schritt wählt man Elemente $v_k^j \in W_k$, $j = 1, \dots, r_{k-1}$, mit $(A - \lambda \mathbf{1})v_k^j = v_{k-1}^j$ und ergänzt die Menge der zugehörigen Eigenvektoren v_1^j (die jeweiligen Endpunkte der Jordanketten) durch weitere Elemente

$$v_1^{r_{k-1}+1}, \dots, v_1^{r_k} \in W_k \cap \ker(A - \lambda \mathbf{1})$$

zu einer Basis von $W_k \cap \ker(A - \lambda \mathbf{1})$. Aus dem obigen Beweis wissen wir, dass man so eine Basis von W_k erhält.

Nach dem N . Schritt ist das Verfahren beendet, da $(A - \lambda \mathbf{1})^0 = \mathbf{1}$ ist.

3. Bilde die Übergangsmatrix, indem als Spalten die verallgemeinerten Eigenvektoren in der resultierenden Jordanbasis genommen werden. Diese Spalten müssen kettenweise und in aufsteigender Ordnung der Stufe gebildet werden, d.h. jede beginnt mit dem Eigenvektor.
4. Bilde die Jordansche Normalform, indem für jede Jordankette der Länge ℓ und Eigenwert λ ein Jordanblock $J_{\ell, \lambda}$ genommen wird. Achte darauf, dass die Ordnung der Jordanblöcke zur entsprechenden Ordnung der Jordanketten passt.

Beispiel 6.4.3. Sei

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

Man ermittelt sehr leicht das charakteristische Polynom von A : $\chi_A(X) = (X-1)^5$. Also ist $\lambda = 1$ der einzige Eigenwert mit algebraischer Vielfachheit $a = 5$. Weiter haben wir

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

Da die zweite und vierte Spalte linear unabhängig sind, ist $\text{rank}(A - \mathbf{1}) = 2$ und $\ker(A - \mathbf{1}) = \text{span}\{e_1, e_3, e_5\}$ ist der zugehörige Eigenraum. Wir müssen also 3-Jordanketten finden. Da $\text{im}(A - \mathbf{1}) \subseteq \ker(A - \mathbf{1})$ gilt, ist $(A - \mathbf{1})^2 = 0$. Also haben alle Jordanketten maximal die Länge 2.

Für $w := e_1 - e_3 - e_5$ erhalten wir nun $(A - \mathbf{1})e_2 = w$. Weiter ist $(A - \mathbf{1})e_4 = e_3$. Wir erhalten daher die drei Jordanketten:

$$(w, e_2), \quad (e_3, e_4) \quad \text{und} \quad (e_5).$$

Da w, e_3, e_5 linear unabhängig sind, folgt aus Lemma 6.4.1, dass (w, e_2, e_3, e_4, e_5) eine Jordanbasis ist. Mit der Transformationsmatrix

$$S := \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

erhalten wir die Jordansche Normalform von A als

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

Bemerkung 6.4.4. In Beispiel 6.4.3 ist die geometrische Vielfachheit 3 und damit gleich der Anzahl der Jordanblöcke. Die Vektoren

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, u_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

sind drei linear unabhängige Eigenvektoren von A , wobei keiner von ihnen im Bild von $A - \mathbf{1}$ ist, d.h. das Ende einer Jordankette der Länge > 1 . Allgemeiner treten solche Beispiele wie folgt auf: ist v ein Eigenvektor, so dass $(A - \lambda \mathbf{1})x = v$ keine Lösungen hat und v_1 ein Eigenvektor, der eine Jordankette der Länge > 1 terminiert, d.h. $v_1 = (A - \lambda \mathbf{1})v_2$, dann ist $v_1 + v$ ein anderer Eigenvektor, so dass $(A - \lambda \mathbf{1})x = (v_1 + v)$ keine Lösung hat.

Es bleibt zu zeigen, dass die Jordansche Normalform eindeutig bis auf Permutation der Jordanblöcke ist. Das ist gleichwertig zu der Aussage, dass sich zwei ähnliche Jordansche Normalformen nur durch eine Permutation ihrer Blöcke unterscheiden. Hierzu zeigen wir den folgende Satz, der zeigt, dass die Größe und Zahl der Jordanblöcke nur von der linearen Abbildung selbst und nicht von der Wahl einer Basis abhängt.

Satz 6.4.5. (Eindeutigkeit der Jordanschen Normalform) *Sei $\varphi: V \rightarrow V$ ein Endomorphismus des endlichdimensionalen $\mathbb{K}$ -Vektorraums V und $d_k := \dim(\ker(\varphi - \lambda \text{id})^k)$, $k \in \mathbb{N}_0$. Dann ist die Zahl c_k der Jordanketten der Länge k zum Eigenwert λ gegeben durch*

$$c_k = 2d_k - d_{k-1} - d_{k+1}. \quad (6.4)$$

Insbesondere hängt für eine Matrix $A \in M_n(\mathbb{K})$ die Zahl der Jordanketten der Länge k zum Eigenwert λ nicht von der Wahl der Jordanbasis ab.

Beweis. Sei N die Länge der längsten auftretenden Jordankette. Dann ist $(\varphi - \lambda \text{id})^N V^\lambda(\varphi) = \{0\}$, also $d_m = d_N = \dim V^\lambda(\varphi)$ für $m \geq N$. Da jede Jordankette mit einem Eigenvektor endet, wissen wir schon, dass

$$d_1 = c_1 + \dots + c_N$$

gilt. Zu d_2 liefert jede der Jordanketten der Länge ≥ 2 einen zweidimensionalen Beitrag, so dass wir

$$d_2 = c_1 + 2c_2 + 2c_3 + \dots + 2c_N$$

erhalten. Induktiv erhalten wir so

$$d_k = c_1 + 2c_2 + \dots + kc_k + kc_{k+1} + \dots + kc_N.$$

Damit erhalten wir

$$d_k - d_{k-1} = c_k + \dots + c_N$$

und somit

$$c_k = (d_k - d_{k-1}) - (d_{k+1} - d_k) = 2d_k - d_{k-1} - d_{k+1}.$$

□

Aufgabe 6.4.6. Sei $\lambda \in \mathbb{C}$ ein Eigenwert von $A \in M_n(\mathbb{C})$. Zeigen Sie, dass der Grad der Nullstelle λ im Minimalpolynom mit dem Fitting-Index $\min\{N: (A - \lambda)^N V^\lambda(A) = 0\}$ des Eigenwerts λ übereinstimmt.

6.5 Die reelle Jordansche Normalform

Der Satz über die Jordansche Normalform gilt nicht für reelle Matrizen, zumal eine reelle Matrix im allgemeinen keine Eigenwerte hat. Aus unserem Beweis der Existenz der Jordanschen Normalform kann man allerdings sehr genau ablesen, was im Reellen richtig bleibt und wo die Probleme liegen.

Sei dazu $\varphi \in \text{End}(V)$ ein Endomorphismus des reellen n -dimensionalen Vektorraums V . Ist λ ein Eigenwert von φ , so folgt die Existenz einer Basis aus Jordanketten für den verallgemeinerten Eigenraum $V^\lambda(\varphi)$ aus Satz 6.4.2, denn für dessen Beweis mussten wir keine Annahmen über den Körper machen.

Zerfällt das charakteristische Polynome von φ in Linearfaktoren, so liefert uns der Beweis von Satz 6.2.8, dass V direkte Summe der verallgemeinerten Eigenräume $V^{\lambda_j}(\varphi)$ ist, und wir erhalten eine Basis von V , die aus Jordanketten besteht.

Im allgemeinen hat das charakteristische Polynom $\chi_\varphi(X)$ die Gestalt

$$\chi_\varphi(X) = (X - \lambda_1)^{a_1} (X - \lambda_2)^{a_2} \cdots (X - \lambda_m)^{a_m} f(X),$$

wobei f keine reellen Nullstellen besitzt und $a_j := a_{\lambda_j}$ die algebraische Vielfachheit des Eigenwerts λ_j ist. Mit dem Zerlegungslemma 6.2.7 erhalten wir damit lediglich eine φ -invariant Zerlegung

$$V = \bigoplus_{j=1}^m V^{\lambda_j}(\varphi) \oplus V_{\text{rest}}.$$

Für den linken Summanden, also die Summe der verallgemeinerten Eigenräume, existiert eine Basis aus Jordanketten. Um eine geeignete Normalform für reelle Matrizen zu erhalten, müssen wir die Einschränkung $\psi := \varphi|_{V_{\text{rest}}}$ genauer betrachten. Sie hat das charakteristische Polynom

$$\chi_\psi(X) = \frac{\chi_\varphi(X)}{(X - \lambda_1)^{a_1} (X - \lambda_2)^{a_2} \cdots (X - \lambda_m)^{a_m}} = f(X),$$

das keine reellen Nullstellen besitzt (vgl. Aufgabe 5.5.11).

Ist $V_{\text{rest}} \neq \{0\}$, also $\deg(f(X)) > 0$, so hat $f(X)$ eine komplexe Nullstelle $\lambda = \alpha + i\beta$. Da das Polynom $f(X) \in \mathbb{R}[X]$ reelle Koeffizienten hat, gilt auch

$$f(\bar{\lambda}) = \overline{f(\lambda)} = 0.$$

Insbesondere ist

$$q_\lambda(X) := (X - \lambda)(X - \bar{\lambda}) = X^2 - (\lambda + \bar{\lambda})X + \lambda\bar{\lambda} = X^2 - 2\alpha X + \alpha^2 + \beta^2$$

ein reelles Polynom ohne Nullstellen, das $f(X)$ teilt. Sei c_λ maximal, so dass $q_\lambda(X)^{c_\lambda}$ ein Teiler von $f(X)$ ist. Dann ist λ keine Nullstelle des reellen Polynoms $f_1(X) := f(X)/q_\lambda(X)^{c_\lambda}$. Induktiv erhalten wir so eine Zerlegung

$$f(X) = \prod_{j=1}^s q_{\lambda_j}(X)^{c_j}, \quad c_j := c_{\lambda_j},$$

wobei $\lambda_j = \alpha_j + i\beta_j$ die komplexen Eigenwerte von φ mit positivem Imaginärteil $\beta_j > 0$ sind.

Mit dem Verfahren aus dem Beweis von Satz 6.2.8 erhalten wir nun induktiv eine φ -invariante Zerlegung

$$V_{\text{rest}} = \bigoplus_{j=1}^s W_j \quad \text{mit} \quad W_j = \ker(q_j(\psi)^{c_j}).$$

Die Unterräume W_j sind also die reellen Varianten der verallgemeinerten Eigenräume. Sie entstehen durch Zusammenfassen der Beiträge eines Paares $(\lambda, \bar{\lambda})$ zweier nichtreeller komplexer Nullstellen des charakteristischen Polynoms $\chi_\varphi(X)$.

Bemerkung 6.5.1. Um die folgende Konstruktion zu verstehen, erinnern wir uns daran, dass jeder komplexe Vektorraum V auch aufgefasst werden kann als reeller Vektorraum $V^{\mathbb{R}}$, auf dem die Multiplikation mit i einen Endomorphismus $Iv = iv$ mit $I^2 = -\text{id}$ definiert.

Ist umgekehrt V ein reeller Vektorraum und $I \in \text{End}(V)$ mit $I^2 = -\text{id}$, so erhalten wir durch die erweiterte Skalarmultiplikation

$$(a + ib)v := av + bIv$$

auf V die Struktur eines komplexen Vektorraums. Eine lineare Abbildung $\psi \in \text{End}_{\mathbb{R}}(V)$ ist genau dann komplex linear, wenn sie mit I vertauscht, also $\psi I = I\psi$ gilt.

Das folgende Lemma zeigt insbesondere, dass V_{rest} die Struktur eines komplexen Vektorraums trägt, so dass die Einschränkung von φ auf V_{rest} komplex linear wird. Damit können wir alle Resultate über Jordanketten im Komplexen in dieser Situation anwenden.

Lemma 6.5.2. *Ist $\psi \in \text{End}(W)$ ein Endomorphismus des reellen endlichdimensionalen Vektorraums W ohne reelle Eigenwerte, so existiert $I \in \text{End}(W)$ mit*

$$I^2 = -\text{id}_W \quad \text{und} \quad I\psi = \psi I.$$

Beweis. Durch Wahl einer Basis von W dürfen wir annehmen, dass $W = \mathbb{R}^n$ ist und $\psi(x) = Ax$ für $A \in M_n(\mathbb{R})$.

Wir betrachten A als komplexe Matrix. Für $\lambda \in \mathbb{C}$ und $z \in \mathbb{C}^n$ ist dann

$$\overline{(A - \lambda \mathbf{1})^k z} = (A - \bar{\lambda} \mathbf{1})^k \bar{z}.$$

Hieraus ergibt sich für die verallgemeinerten Eigenräume $V^{\lambda_j} \subseteq V := \mathbb{C}^n$ die Relation

$$\overline{V^{\lambda_j}} = V^{\bar{\lambda}_j}.$$

Für

$$V_+ := \bigoplus_{\text{Im } \lambda > 0} V^\lambda \quad \text{und} \quad V_- := \overline{V_+} := \bigoplus_{\text{Im } \lambda < 0} V^\lambda$$

gilt daher

$$V = V_+ \oplus V_-.$$

Wir betrachten die reell lineare Abbildung

$$p: V_+ \rightarrow \mathbb{R}^n, \quad p(z) := \frac{z + \bar{z}}{2} = \text{Re } z, \quad p(x + iy) = x$$

und behaupten, dass sie bijektiv ist. Aus $p(z) = 0$ folgt $z = -\bar{z} \in V_+ \cap V_- = \{0\}$. Also ist p injektiv. Ist $x \in \mathbb{R}^n$, so können wir x darstellen als $x = z + w$ mit $z \in V_+$ und $w \in V_-$. Da x reell ist, gilt dann

$$z + w = x = \bar{x} = \bar{z} + \bar{w},$$

also

$$z - \bar{w} = \bar{z} - w \in V_+ \cap V_- = \{0\}.$$

Damit ist $w = \bar{z}$, also $x = z + \bar{z} = \text{Re}(2z) \in p(V_+)$.

Wir definieren $I \in M_n(\mathbb{R})$ durch

$$Ix := p(ip^{-1}(x)).$$

Damit ergibt sich

$$I^2 x = p(ip^{-1}(p(ip^{-1}(x)))) = p(iip^{-1}(x)) = -p(p^{-1}(x)) = -x.$$

Weiter ist

$$p(Az) = \frac{Az + A\bar{z}}{2} = Ap(z) \quad \text{und} \quad Aiz = iAz.$$

Hieraus erhalten wir

$$IAx = p(ip^{-1}(Ax)) = p(iAp^{-1}(x)) = p(Aip^{-1}(x)) = Ap(ip^{-1}(x)) = AIX. \quad \square$$

Mit Lemma 6.5.2 und Bemerkung 6.5.1 erhalten wir auf W die Struktur eines komplexen Vektorraums, so dass ψ komplex linear ist. Damit gilt dies für alle Abbildungen $p(\psi)$, $p \in \mathbb{R}[X]$. Insbesondere sind alle Unterräume W invariant unter I , d.h. komplexe Unterräume. Auf W_j hat ψ als komplex lineare Abbildung die beiden Eigenwerte λ_j und $\bar{\lambda}_j$. Ist $w \in W$ eine Eigenvektor zum Eigenwert $\lambda = \alpha + i\beta$, so sind w und Iw über $\mathbb{R}$ linear unabhängig (Aufgabe 6.5.6) und es gilt

$$\psi(w) = \lambda w = \alpha w + \beta Iw, \quad \psi(Iw) = I\varphi(w) = -\beta w + \alpha Iw.$$

Ist $w_k := (\psi - \lambda)^{N-k} w$ eine Jordankette zum Eigenwert λ , so haben wir analog

$$\psi(w_k) = (\psi - \lambda \text{id})w_k + \lambda w_k = w_{k-1} + (\alpha w_k + \beta I w_k)$$

und

$$\psi(I w_k) = I w_{k-1} + (-\beta w_k + \alpha I w_k).$$

Die Einschränkung von ψ auf den Spann der Vektoren

$$w_1, I w_1, w_2, I w_2, \dots, w_N, I w_N$$

hat also bzgl. dieser Basis die Matrix

$$J_{m,\lambda}^{\mathbb{R}} := \begin{pmatrix} \boxed{\begin{matrix} \alpha & -\beta \\ \beta & \alpha \end{matrix}} & \boxed{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}} & & & & \\ & \boxed{\begin{matrix} \alpha & -\beta \\ \beta & \alpha \end{matrix}} & \boxed{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}} & & & \\ & & \ddots & \ddots & & \\ & & & & \ddots & \boxed{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}} \\ & & & & & \boxed{\begin{matrix} \alpha & -\beta \\ \beta & \alpha \end{matrix}} \\ 0 & & & & & \end{pmatrix} \in M_{2N}(\mathbb{R})$$

Wenden wir den Satz über die Jordansche Normalform auf den komplex linearen Endomorphismus ψ von W an, erhalten wir die folgende reelle Variante der Jordanschen Normalform.

Die reelle Jordansche Normalform

Satz 6.5.3. Sei V ein n -dimensionaler reeller Vektorraum und $\varphi \in \text{End}(V)$. Dann existiert eine Basis $B = (v_1, v_2, \dots, v_n)$ von V , so dass die Matrix $[\varphi]_B$ von φ bzgl. B eine Blockdiagonalmatrix ist, deren Blöcke von der Gestalt

$$J_{m,\lambda}, \quad \lambda \in \mathbb{R} \quad \text{oder} \quad J_{m,\lambda}^{\mathbb{R}}, \quad \lambda \in \mathbb{C} \setminus \mathbb{R}$$

sind.

Beispiel 6.5.4. Für eine reelle (2×2) -Matrix erhalten wir folgende Möglichkeiten für die Jordansche Normalform:

- $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, $\lambda_1 \neq \lambda_2 \in \mathbb{R}$ (zwei verschiedene reelle Eigenwerte).
- $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, $\lambda \in \mathbb{R}$, (ein reeller Eigenwerte der geometrischen Vielfachheit 2).
- $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, $\lambda \in \mathbb{R}$, (ein reeller Eigenwerte der geometrischen Vielfachheit 1).
- $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$, $\alpha, \beta \in \mathbb{R}, \beta > 0$, (kein reeller Eigenwert).

Beispiel 6.5.5. Für eine reelle (3×3) -Matrix erhalten wir folgende Möglichkeiten für die Jordansche Normalform. Hierzu erinnern wir uns daran, dass es immer einen reellen Eigenwert gibt, da das charakteristische Polynom vom Grad 3 eine reelle Nullstelle besitzt.

- $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$, (diagonalisierbare Matrizen).
- $\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$, (zwei reelle Eigenwerte, nicht diagonalisierbar).
- $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$, (ein reeller Eigenwert, geometrische Vielfachheit $g_\lambda = 1$).
- $\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \alpha & -\beta \\ 0 & \beta & \alpha \end{pmatrix}$, (nur ein reeller Eigenwert, χ zerfällt nicht).

Aufgabe 6.5.6. Sei V ein komplexer Vektorraum und $v_1, \dots, v_n \in V$. Dann sind diese Vektoren genau dann über $\mathbb{C}$ linear unabhängig, wenn $v_1, iv_1, v_2, iv_2, \dots, v_n, iv_n$ über $\mathbb{R}$ linear unabhängig sind.

6.6 Matrixpotenzen

Ist J eine Jordansche Normalform, dann können wir $J = D + N$ schreiben, wobei D eine Diagonalmatrix und N eine Matrix mit verschwindender Diagonale und einigen Einsen oberhalb der Diagonale ist. Ist J ein Jordanblock, dann gilt $D = \lambda \mathbf{1}$ und natürlich $DN = ND$. Ist J eine Jordansche Normalform aus mehreren Blöcken, die möglicherweise zu verschiedenen Eigenwerten gehört, dann gilt ebenfalls $DN = ND$, wie man anhand von Blockmultiplikation sehen kann.

Es gilt also

$$J^\ell = (D + N)^\ell = \sum_{i=0}^{\ell} \binom{\ell}{i} D^{\ell-i} N^i = D^\ell + \ell D^{\ell-1} N + \dots$$

Nur die ersten Terme dieser Summe sind von Null verschieden, genauer diejenigen, bei denen i kleiner als die Größe des größten Jordanblocks in J ist. Die Summe kann also mit vertretbarem Aufwand berechnet werden. Ist z.B. J die Jordansche Normalform aus Beispiel 6.4.3, so gilt

$$J^\ell = \begin{pmatrix} 1 & \ell & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \ell & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Natürlich folgt wie gewöhnlich aus $S^{-1}AS = J$ auch $A^\ell = SJ^\ell S^{-1}$.

Beispiel 6.6.1. Die Lösung einer linearen Rekursion erster Ordnung

$$f(0) = g, \quad f(k+1) = Af(k)$$

mit $f: \mathbb{N} \rightarrow \mathbb{C}^n, g \in \mathbb{C}^n, A \in M_n(\mathbb{C})$ ergibt sich als $f(n) = A^n g$ ergibt. Ist A nicht diagonalisierbar, dann können wir die Jordansche Normalform benutzen, um A^n zu berechnen. Betrachte z.B. das

folgende System einer linearen Rekursionen erster Ordnung:

$$\begin{aligned} a_{k+1} &= 3a_k + b_k \\ b_{k+1} &= 3b_k \\ c_{k+1} &= -b_k + 3c_k + d_k \\ d_{k+1} &= 3d_k \\ e_{k+1} &= -b_k + 3e_k \\ a_0 = b_0 = c_0 = d_0 = e_0 &= 1 \end{aligned}$$

Schreiben wir $f(k) = (a_k \ b_k \ c_k \ d_k \ e_k)^\top$ und $g = (1 \ 1 \ 1 \ 1 \ 1)^\top$, so folgt

$$f(k+1) = (A + 2\mathbf{1})f(k), f(0) = g$$

wobei A die Matrix aus Beispiel 6.4.3 ist. Wegen $S^{-1}AS = J$ gilt auch $S^{-1}(A + 2\mathbf{1})S = J + 2\mathbf{1}$, woraus wir die Jordansche Normalform von $A + 2E$ ohne weitere Berechnung erhalten.

Aus dem Obigen erkennen wir

$$(J + 2\mathbf{1})^k = \begin{pmatrix} 3^k & k3^{k-1} & 0 & 0 & 0 \\ 0 & 3^k & 0 & 0 & 0 \\ 0 & 0 & 3^k & k3^{k-1} & 0 \\ 0 & 0 & 0 & 3^k & 0 \\ 0 & 0 & 0 & 0 & 3^k \end{pmatrix},$$

woraus wir folgende Lösung in geschlossener Form erhalten:

$$f(k) = S(J + 2\mathbf{1})^k S^{-1}g$$

oder

$$\begin{aligned} a(k) = e(k) &= 3^{k-1}k + 3^k \\ b(k) = c(k) = d(k) &= 3^k, \end{aligned}$$

Bemerkung 6.6.2. Lineare Rekursionen treten insbesondere in der folgenden Situation auf: Wir betrachten ein kontinuierliches dynamische Systeme, dessen Zustand $x(t) \in \mathbb{R}^n$ (in Abhängigkeit von der Zeit t) sich durch die lineare Differentialgleichung

$$\dot{x}(t) = Ax(t), \quad A \in M_n(\mathbb{R}),$$

beschreiben lässt. Wir werden im nächsten Abschnitt sehen, wie sich solche Gleichungen exakt lösen lassen. Oft ist es aber praktikabler, die kontinuierliche zeitliche Entwicklung durch eine zeitliche diskrete Entwicklung zu approximieren. Man interessiert sich dann nur noch für Zeitpunkte $t_k = t_0 + k\Delta t$, wobei $\Delta t > 0$ das Maß für die Feinheit der diskretisierten Zeitachse ist. Entsprechend erhalten wir für die Entwicklung des diskreten dynamischen Systems, dessen Zustände durch die Vektoren $x^{(k)} := x(t_k)$ gegeben, die näherungsweise Relation

$$\frac{1}{\Delta t}(x(t_{k+1}) - x(t_k)) \sim \dot{x}(t_k) = Ax(t_k).$$

Dies führt zur Rekursionsformel

$$x^{(k+1)} = (\mathbf{1} + \Delta t A)x^{(k)}$$

und man kann erwarten, dass für den Anfangswert $x^{(0)} := x(t_0)$ die so erhaltene Folge $(x^{(k)})_{k \in \mathbb{N}}$ Näherungswerte für die Werte $x(t_k)$ liefert, die man bei einer kontinuierlichen Systementwicklung erhalten hätte.

Konkret haben wir

$$x^{(k)} = (\mathbf{1} + \Delta t A)^k x^{(0)}.$$

Das Problem besteht also in der möglichst expliziten Berechnung der Potenzen der Matrix $(\mathbf{1} + \Delta t A)^k$.

Bemerkung 6.6.3. Wir betrachten die lineare Rekursion

$$x^{(k+1)} = Ax^{(k)}, \quad k \in \mathbb{N}_0.$$

Ist der Anfangswert $x^{(0)}$ ein Eigenvektor von A zum Eigenwert λ , so ergibt sich die einfache Lösung

$$x^{(k)} = \lambda^k x^{(0)}.$$

Ist $x^{(0)}$ ein Hauptvektor der Stufe 2, d.h., $(A - \lambda \mathbf{1})^2 x^{(0)} = 0$, so erhalten wir

$$A^k x^{(0)} = (\lambda \mathbf{1} + (A - \lambda \mathbf{1}))^k x^{(0)} = \lambda^k x^{(0)} + k \lambda^{k-1} (A - \lambda \mathbf{1}) x^{(0)}$$

und allgemein für $(A - \lambda \mathbf{1})^m x^{(0)} = 0$:

$$\begin{aligned} A^k x^{(0)} &= (\lambda \mathbf{1} + (A - \lambda \mathbf{1}))^k x^{(0)} = \sum_{j=0}^k \binom{k}{j} \lambda^{k-j} (A - \lambda \mathbf{1})^j x^{(0)} \\ &= \lambda^k x^{(0)} + k \lambda^{k-1} (A - \lambda \mathbf{1}) x^{(0)} + \binom{k}{2} \lambda^{k-2} (A - \lambda \mathbf{1})^2 x^{(0)} + \dots \end{aligned}$$

Die Lösung ist also eine endliche Summe von Termen der Gestalt $\lambda^{k-j} \binom{k}{j} v_j$.

Im allgemeinen zerlegt man den Anfangswert $x^{(0)}$ in seine Komponenten $x^{(0)} = \sum_{\ell=1}^r x_\ell^{(0)}$ zerlegen, wobei $x_\ell^{(0)}$ ein (im allgemeinen komplexer) Hauptvektor zum Eigenwert λ_ℓ ist. Für die Lösung erhalten wir dann eine entsprechende Summe von Ausdrücken der Gestalt $\lambda_\ell^{k-j} \binom{k}{j} v_{j,\ell}$.

Beispiel 6.6.4. Mit diesen Techniken lassen sich insbesondere die Fibonacci-Zahlen $(F_n)_{n \in \mathbb{N}_0}$ durch einen expliziten Ausdruck berechnen. Sie genügen der Rekursionsformel

$$F_{k+2} = F_{k+1} + F_k \quad \text{mit} \quad F_0 = 0, F_1 = 1.$$

Da die Rekursionsformel die zwei vorhergehenden Folgenglieder enthält, modellieren wir die Entwicklung durch einen Vektor

$$x^{(k)} := \begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix} \in \mathbb{R}^2, \quad k \in \mathbb{N}_0.$$

Die Rekursionformel wird dann zu der linearen Gleichung

$$x^{(k+1)} = Ax^{(k)} \quad \text{mit} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Das charakteristische Polynom dieser Matrix ist

$$\chi_A(X) = \det \begin{pmatrix} 1-X & 1 \\ 1 & -X \end{pmatrix} = X(X-1) - 1 = X^2 - X - 1.$$

Es hat die beiden reellen Nullstellen

$$\lambda_{1,2} := \frac{1 \pm \sqrt{5}}{2}$$

mit den zugehörigen Eigenvektoren

$$v_j := \begin{pmatrix} \lambda_j \\ 1 \end{pmatrix}, \quad j = 1, 2.$$

Der Startvektor $x^{(0)}$ hat in der Eigenvektorbasis die Darstellung

$$x^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\lambda_1 - \lambda_2} (v_1 - v_2).$$

Damit erhalten wir

$$x^{(k)} = A^k x^{(0)} = \frac{1}{\lambda_1 - \lambda_2} (\lambda_1^k v_1 - \lambda_2^k v_2) = \frac{1}{\lambda_1 - \lambda_2} \left(\begin{pmatrix} \lambda_1^{k+1} \\ \lambda_1^k \end{pmatrix} - \begin{pmatrix} \lambda_2^{k+1} \\ \lambda_2^k \end{pmatrix} \right).$$

Insbesondere ist

$$F_k = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^k - \left(\frac{1 - \sqrt{5}}{2} \right)^k \right) \sim \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^k,$$

da der zweite Summand sehr schnell gegen 0 konvergiert. Insbesondere erhalten wir

$$\lim_{k \rightarrow \infty} \frac{F_{k+1}}{F_k} = \frac{1 + \sqrt{5}}{2} \sim 1,618.$$

Diese Zahl ist das Verhältnis des *goldenen Schnitts*.

6.7 Die Matrixexponentialfunktion

In diesem Abschnitt lernen wir die Exponentialfunktion für Matrizen kennen. Sie ist ein sehr vielseitiges Werkzeug und spielt in der Theorie der Matrixengruppen (Lie-Gruppen). Wir werden sie dazu verwenden, die Lösungen von linearen Differentialgleichungen erster Ordnung mit konstanten Koeffizienten zu beschreiben. In diesem Abschnitt bezeichne $\mathbb{K}$ immer $\mathbb{R}$ oder $\mathbb{C}$.

Wir betrachten auf $\mathbb{K}^n$ die euklidische Norm

$$\|x\| := \|x\|_2 := \left(\sum_{j=1}^n |x_j|^2 \right)^{\frac{1}{2}}.$$

Auf dem Raum $M_{m,n}(\mathbb{K})$ der $(m \times n)$ -Matrizen betrachten wir die zugehörige Operatornorm

$$\|A\| := \sup\{\|Ax\| : x \in \mathbb{K}^n, \|x\| \leq 1\}.$$

Dass dieses Supremum endlich ist, folgt aus der Stetigkeit der Abbildung $L_A: \mathbb{K}^n \rightarrow \mathbb{K}^m, x \mapsto Ax$ und der Kompaktheit der Einheitskugel $\{x \in \mathbb{K}^n : \|x\| \leq 1\}$ (vgl. Analysis II). Ist $0 \neq y \in \mathbb{K}^n$, so ist $x := \frac{y}{\|y\|}$ ein Einheitsvektor und wir erhalten

$$\|Ay\| = \|A \frac{y}{\|y\|}\| \|y\| \leq \|A\| \|y\|.$$

Für ein Matrixprodukt BA und $\|x\| \leq 1$ ergibt sich damit

$$\|BAx\| \leq \|B\| \|Ax\| \leq \|B\| \|A\|$$

und damit die Submultiplikativität der Norm:

$$\|BA\| \leq \|B\| \|A\| \quad \text{für} \quad A \in M_{m,n}(\mathbb{K}), B \in M_{r,m}(\mathbb{K}). \quad (6.5)$$

Bemerkung 6.7.1. (a) Im Rahmen der Analysis II lernt man insbesondere, dass alle Normen auf $\mathbb{K}^n$ äquivalent sind, d.h., dass zu zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ Konstanten $c, C > 0$ existieren mit

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1 \quad \text{für alle} \quad x \in \mathbb{K}^n.$$

Hieraus folgt insbesondere, dass $\mathbb{K}^n$ bzgl. jeder Norm vollständig ist, d.h. ein Banachraum. Eine wichtige Konsequenz hieraus ist, dass eine Reihe $\sum_n v_n$ konvergiert, wenn die Reihe $\sum_n \|v_n\|$ nichtnegativer reeller Zahlen konvergiert (Majorantenkriterium).

(b) Da die Operatornorm auf $M_{m,n}(\mathbb{K})$ eine Norm definiert, gelten die Bemerkungen aus (a) insbesondere für den Raum $M_{m,n}(\mathbb{K}) \cong \mathbb{K}^{nm}$.

Satz 6.7.2. Für jedes $X \in M_n(\mathbb{K})$ konvergiert die Exponentialreihe

$$\exp(X) := e^X := \sum_{n=0}^{\infty} \frac{1}{n!} X^n.$$

- (1) Es ist $e^0 = \mathbf{1}$ und für $XY = YX$ ist $e^X \cdot e^Y = e^{X+Y}$.
 (2) Für alle $g \in \mathrm{GL}_n(\mathbb{K})$ ist $g \cdot e^X \cdot g^{-1} = e^{gXg^{-1}}$.
 (3) Für alle $X \in M_n(\mathbb{K})$ ist $e^X \in \mathrm{GL}_n(\mathbb{K})$ und $(e^X)^{-1} = e^{-X}$.

Beweis. Wegen $\|\frac{1}{n!}X^n\| \leq \frac{1}{n!}\|X\|^n$ folgt die Konvergenz der Exponentialreihe aus dem Majorantenkriterium.

(1) Die Aussage $e^0 = \mathbf{1}$ ist trivial. Gilt $XY = YX$, d.h. die linearen Abbildungen vertauschen miteinander, so ist

$$(X + Y)^n = \sum_{k=0}^n \binom{n}{k} X^k \cdot Y^{n-k}.$$

Wir erhalten nun mit der Cauchy-Produktformel:

$$\begin{aligned} e^{X+Y} &= \sum_{n=0}^{\infty} \frac{1}{n!} (X+Y)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} X^k Y^{n-k} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{X^k}{k!} \frac{Y^{n-k}}{(n-k)!} = \sum_{i=0}^{\infty} \frac{X^i}{i!} \sum_{j=0}^{\infty} \frac{Y^j}{j!} = e^X \cdot e^Y. \end{aligned}$$

(2) Für $X \in M_n(\mathbb{K})$ und $g \in \mathrm{GL}_n(\mathbb{K})$ gilt $gX^n g^{-1} = (gXg^{-1})^n$, und somit folgt aus der Stetigkeit der Matrizenmultiplikation:

$$ge^X g^{-1} = g \cdot \left(\sum_{n=0}^{\infty} \frac{1}{n!} X^n \right) \cdot g^{-1} = \sum_{n=0}^{\infty} \frac{1}{n!} gX^n g^{-1} = \sum_{n=0}^{\infty} \frac{1}{n!} (gXg^{-1})^n = e^{gXg^{-1}}.$$

(3) Aus (1) folgt $e^X \cdot e^{-X} = \mathbf{1}$, also $e^X \in \mathrm{GL}_n(\mathbb{K})$ mit $e^{-X} = (e^X)^{-1}$. □

Beispiel 6.7.3. (a) Für Diagonalmatrizen

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \text{ gilt } D^k = \begin{pmatrix} d_1^k & & 0 \\ & \ddots & \\ 0 & & d_n^k \end{pmatrix}, \text{ also } e^D = \begin{pmatrix} e^{d_1} & & 0 \\ & \ddots & \\ 0 & & e^{d_n} \end{pmatrix}.$$

(b) Für $A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$ ist $A^2 = \begin{pmatrix} 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix}$, etc. und schließlich $A^n = 0$.

Man erhält hieraus die folgende Formel

$$e^{tA} = \sum_{k=0}^{\infty} \frac{1}{k!} (tA)^k = \sum_{k=0}^{n-1} \frac{1}{k!} (tA)^k = \begin{pmatrix} 1 & t & \frac{t^2}{2} & \dots & \frac{t^{n-1}}{(n-1)!} \\ & 1 & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{t^2}{2} \\ & 0 & & 1 & t \\ & & & & 1 \end{pmatrix};$$

so ergibt sich beispielsweise für die Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix} \quad \text{als Funktionswert} \quad e^{tA} = \begin{pmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$$

(c) Für Blockdiagonalmatrizen $A = \begin{pmatrix} B & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix}$ gilt $A^n = \begin{pmatrix} B^n & \mathbf{0} \\ \mathbf{0} & C^n \end{pmatrix}$, also $e^A = \begin{pmatrix} e^B & \mathbf{0} \\ \mathbf{0} & e^C \end{pmatrix}$.

(d) Für $A = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$ (ein Jordankästchen) ist $A = \lambda \cdot \mathbf{1} + N$ mit $N = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$.

Wegen $(\lambda \mathbf{1}) \cdot N = N \cdot (\lambda \mathbf{1})$ gilt also

$$e^A = e^{\lambda \mathbf{1}} \cdot e^N = e^\lambda \cdot e^N.$$

Wegen Satz 6.7.2(2) lässt sich die Berechnung von e^A auf den Fall zurückführen, wo A in Jordan-Normalform vorliegt, d.h. folgende Form hat:

$$A = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{matrix}} & & & \mathbf{0} \\ & \ddots & & \\ & & \mathbf{0} & \ddots \\ & & & \boxed{\begin{matrix} \lambda_n & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_n \end{matrix}} \end{pmatrix}$$

Mittels (a)–(d) lässt sich e^{tA} nun direkt berechnen.

(e) Für $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ ist $A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ und somit

$$\begin{aligned} e^{tA} &= \sum_{k=0}^{\infty} \frac{1}{k!} (tA)^k \\ &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} t^{2m} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} t^{2m+1} \cdot \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos t & 0 \\ 0 & \cos t \end{pmatrix} + \begin{pmatrix} 0 & \sin t \\ -\sin t & 0 \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}. \end{aligned}$$

Satz 6.7.4. Sei $A \in M_n(\mathbb{K})$ und

$$\Phi_A : \mathbb{R} \rightarrow \mathrm{GL}_n(\mathbb{K}), \quad t \mapsto \exp(tA).$$

Dann gelten folgende Aussagen:

- (1) Die Abbildung Φ_A ist ein differenzierbarer Gruppenhomomorphismus von $(\mathbb{R}, +)$ nach $(\mathrm{GL}_n(\mathbb{K}), \cdot)$.
 (2) Für alle $t \in \mathbb{R}$ gilt $\Phi'_A(t) = A \cdot \Phi_A(t) = \Phi_A(t) \cdot A$.

Beweis. (1) Wegen $tX \cdot sX = sX \cdot tX$ für $s, t \in \mathbb{R}$ folgt aus Satz 6.7.2, dass

$$\Phi_A(t) \cdot \Phi_A(s) = \Phi_A(t+s)$$

gilt. Wegen $\Phi_A(\mathbb{R}) \subseteq \mathrm{GL}_n(\mathbb{K})$ ist Φ_A daher ein Gruppenhomomorphismus $(\mathbb{R}, +) \rightarrow \mathrm{GL}_n(\mathbb{K})$.
 Aus

$$\|e^X - \mathbf{1} - X\| = \left\| \sum_{n=2}^{\infty} \frac{X^n}{n!} \right\| \leq \sum_{n=2}^{\infty} \frac{\|X\|^n}{n!} = e^{\|X\|} - 1 - \|X\|$$

folgt

$$\left\| \frac{1}{t}(e^{tX} - \mathbf{1}) - X \right\| = \frac{\|e^{tX} - \mathbf{1} - tX\|}{|t|} \leq \frac{e^{|t|\|X\|} - 1 - |t|\|X\|}{|t|} \rightarrow 0$$

für $t \rightarrow 0$. Also ist

$$\Phi'_A(0) = \lim_{t \rightarrow 0} \frac{e^{tA} - \mathbf{1}}{t} = A.$$

Damit erhalten wir

$$\begin{aligned} \Phi'_A(t) &= \lim_{h \rightarrow 0} \frac{\Phi_A(t+h) - \Phi_A(t)}{h} = \lim_{h \rightarrow 0} \frac{\Phi_A(h)\Phi_A(t) - \Phi_A(t)}{h} \\ &= \left(\lim_{h \rightarrow 0} \frac{\Phi_A(h) - \mathbf{1}}{h} \right) \cdot \Phi_A(t) = \Phi'_A(0)\Phi_A(t) = A \cdot \Phi_A(t) \end{aligned}$$

und analog erhalten wir $\Phi'_A(t) = \Phi_A(t)A$. Folglich ist Φ_A eine differenzierbare Kurve in $M_n(\mathbb{K})$, und die Aussage (2) gilt. $\square$

Bemerkung 6.7.5. Auf $\mathbb{K}^n$ betrachten wir nun die *homogene lineare Differentialgleichung*

$$\dot{y} = Ay, \quad y(t_0) = y_0 \tag{6.6}$$

mit konstanten Koeffizienten. Wir suchen Lösungskurven $y: \mathbb{R} \rightarrow \mathbb{K}^n$.

Betrachtet man die Komponenten $y_1, \dots, y_n$ einzeln, so spricht man auch von einem System von n linearen Differentialgleichungen mit konstanten Koeffizienten:

$$\dot{y}_j(t) = \sum_{k=1}^n a_{jk} y_k(t), \quad y_j(0) = y_{0,j}.$$

Mit Hilfe der Matrixexponentialfunktion, lassen sich homogene lineare Differentialgleichungen sehr leicht lösen. Z.B. erhalten wir für die Kurve

$$y(t) := e^{(t-t_0)A} y_0 = e^{tA} e^{-t_0A} y_0$$

sofort

$$\dot{y}(t) = A e^{(t-t_0)A} y_0 = A y(t) \quad \text{und} \quad y(t_0) = y_0.$$

Dass dies die einzige Lösung des Anfangswertproblems (6.6) ist, lässt sich auch leicht einsehen. Ist $z: \mathbb{R} \rightarrow \mathbb{K}^n$ eine andere Lösung, so erhalten wir für die Kurve $w(t) := e^{-tA} z(t)$ mit der Produktregel die Ableitung

$$\dot{w}(t) = e^{-tA}(-A)z(t) + e^{-tA}\dot{z}(t) = e^{-tA}(-Az(t) + \dot{z}(t)) = 0.$$

Also ist w konstant und daher

$$z(t) = e^{tA} w(t) = e^{tA} w(t_0) = e^{tA} e^{-t_0A} z(t_0) = e^{(t-t_0)A} y_0 = y(t).$$

Bemerkung 6.7.6. Ist der Anfangswert $y_0 = y(0)$ ein Eigenvektor von A zum Eigenwert λ , so ergibt sich die einfache Lösung

$$y(t) = e^{tA}y_0 = e^{t\lambda}y_0.$$

Ist y_0 ein Hauptvektor der Stufe 2, d.h., $(A - \lambda\mathbf{1})^2y_0 = 0$, so erhalten wir

$$e^{-t\lambda}y(t) = e^{-t\lambda}e^{tA}y_0 = e^{t(A-\lambda\mathbf{1})}y_0 = y_0 + t(A - \lambda\mathbf{1})y_0$$

und allgemein für $(A - \lambda\mathbf{1})^m y_0 = 0$:

$$e^{-t\lambda}y(t) = e^{-t\lambda\mathbf{1}}e^{tA}y_0 = e^{t(A-\lambda\mathbf{1})}y_0 = y_0 + t(A - \lambda\mathbf{1})y_0 + \dots + \frac{t^{m-1}}{(m-1)!}(A - \lambda\mathbf{1})^{m-1}y_0.$$

Hieraus ergibt sich

$$y(t) = e^{t\lambda} \left(y_0 + t(A - \lambda\mathbf{1})y_0 + \dots + \frac{t^{m-1}}{(m-1)!}(A - \lambda\mathbf{1})^{m-1}y_0 \right). \quad (6.7)$$

Im allgemeinen zerlegt man den Anfangswert y_0 in seine Komponenten $y_0 = \sum_{j=1}^r y_0^j$, wobei y_0^j ein Hauptvektor zum Eigenwert λ_j ist. Für die Lösung erhalten wir dann

$$y(t) = e^{tA}y_0 = \sum_{j=1}^r e^{tA}y_0^j,$$

wobei $e^{tA}y_0^j$ die Gestalt aus (6.7) hat.

Beispiel 6.7.7. Auf $\mathbb{R}^2$ betrachten wir das System der linearen Differentialgleichungen

$$\dot{x}_1(t) = -x_2(t), \quad \dot{x}_2(t) = x_1(t),$$

In Matrixschreibweise erhalten wir

$$\dot{x}(t) = Ax(t) \quad \text{mit} \quad A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Die Lösungen dieser Differentialgleichungen haben für den Anfangsvektor $x_0 = \begin{pmatrix} a \\ b \end{pmatrix}$ die Gestalt

$$x(t) = e^{tA}x_0 = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} x_0,$$

also

$$x_1(t) = a \cos t - b \sin t \quad \text{und} \quad x_2(t) = a \sin t + b \cos t.$$

Kapitel 7

Euklidische und unitäre Räume

In den letzten Kapiteln haben wir Vektorräume über beliebigen Körpern betrachtet. Die einzigen Strukturen, die wir benutzt haben, waren Addition und skalare Multiplikation. Wir werden nun die Begriffe der Länge und der Orthogonalität und einen verallgemeinerten Winkelbegriff in Vektorräumen kennenlernen, die eine zusätzliche Struktur tragen, die durch ein Skalarprodukt gegeben ist. Dazu beschränken wir uns auf Vektorräume über den Körpern $\mathbb{R}$ und $\mathbb{C}$ der reellen und der komplexen Zahlen.

7.1 Unitäre Räume

In diesem Kapitel ist $\mathbb{K}$ immer einer der Körper $\mathbb{R}$ oder $\mathbb{C}$ und V ein $\mathbb{K}$ -Vektorraum. In diesem Sinn schreiben wir für reelle Zahlen $x \in \mathbb{R}$ auch $\bar{x} = x$.

Wir erinnern uns an die Definition einer bilinearen Abbildung $\beta: V \times V \rightarrow \mathbb{K}$. Sie soll in jedem der beiden Argumente linear sein, d.h.,

$$\beta(\lambda x + \mu y, z) = \lambda \beta(x, z) + \mu \beta(y, z) \quad \text{und} \quad \beta(x, \lambda y + \mu z) = \lambda \beta(x, y) + \mu \beta(x, z)$$

gilt für $x, y, z \in V$ und $\lambda, \mu \in \mathbb{K}$. Für $\mathbb{K} = \mathbb{C}$ lernen wir jetzt eine interessante Modifikation dieses Begriffes kennen, die es uns erlauben wird, auch im Komplexen die Grundstrukturen der euklidischen Geometrie in geeigneter Art und Weise zu beschreiben.

Definition 7.1.1. Sei V ein $\mathbb{K}$ -Vektorraum.

(a) Eine reell bilineare Abbildung $\beta: V \times V \rightarrow \mathbb{K}$ heißt *sesquilinear* oder *Sesquilinearform*, wenn gilt

$$\lambda \beta(x, y) = \beta(\lambda x, y) = \beta(x, \bar{\lambda} y) \quad \text{für} \quad \lambda \in \mathbb{K}, x, y \in V.$$

Für $\mathbb{K} = \mathbb{R}$ folgt dies sofort aus der Bilinearität, aber für $\mathbb{K} = \mathbb{C}$ ist dies eine zusätzliche Bedingung. In diesem Fall bedeutet sie, dass β im ersten Argument linear und im zweiten Argument *antilinear* ist.

(b) Eine sesquilinear Abbildung $\beta: V \times V \rightarrow \mathbb{K}$ heißt *hermitesch*, wenn

$$\overline{\beta(x, y)} = \beta(y, x) \quad \text{für} \quad x, y \in V$$

gilt. Für $\mathbb{K} = \mathbb{R}$ ist β bilinear und *symmetrisch*, d.h.,

$$\beta(x, y) = \beta(y, x) \quad \text{für} \quad x, y \in V$$

gilt. Hermitesche Sesquilinearformen über $\mathbb{R}$ sind also dasselbe wie symmetrische Bilinearformen.

(c) Eine hermitesche Sequilinearform $\beta: V \times V \rightarrow \mathbb{K}$ heißt *positiv semidefinit*, wenn

$$\beta(x, x) \geq 0 \quad \text{für} \quad x \in V$$

gilt und *positiv definit*, wenn sogar

$$\beta(x, x) > 0 \quad \text{für} \quad 0 \neq x \in V$$

gilt.

(d) Ein paar (V, β) aus einem $\mathbb{K}$ -Vektorraum V und einer positiv definite hermiteschen Form β heißt *unitärer Raum*. Für $\mathbb{K} = \mathbb{R}$ spricht man auch von einem *euklidischen Vektorraum*.

Meistens schreiben wir

$$\langle x, y \rangle := \beta(x, y)$$

für die hermitesche Form eines unitären Raums.

Die *Länge* oder *Norm* eines Vektors $z \in V$ ist definiert durch

$$\|z\| := \sqrt{\langle z, z \rangle} = \sqrt{\beta(z, z)}.$$

Jeder Vektor der Länge eins heißt *Einheitsvektor*. Jeder nichtverschwindende Vektor z kann *normiert* werden: $\frac{1}{\|z\|}z$ ist ein Einheitsvektor, denn es ist

$$\left\langle \frac{1}{\|z\|}z, \frac{1}{\|z\|}z \right\rangle = \frac{1}{\|z\|^2} \langle z, z \rangle = 1.$$

Zwei Vektoren $z, w \in V$ heißen *orthogonal*, i.Z. $z \perp w$, wenn $\langle z, w \rangle = 0$ gilt. Wegen

$$\langle z, w \rangle = \overline{\langle w, z \rangle}$$

ist *Orthogonalität* auch im Komplexen eine symmetrische Relation.

Unitäre Räume sind der Gegenstand dieses Kapitels. Hier sind die beiden zentralen Beispiele.

Beispiel 7.1.2. Das *Standardskalarprodukt* zweier Vektoren $x, y \in \mathbb{R}^n$ wurde definiert als

$$\langle x, y \rangle = \left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = x^\top y.$$

Hierdurch wird $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ zu einem euklidischen Vektorraum. Die Länge eines Vektors $x \in \mathbb{R}^n$ ist

$$\|x\| := \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Beispiel 7.1.3. Wir möchten ein positiv definites hermitesches Skalarprodukt auf $\mathbb{C}^n$ definieren. Wir können allerdings nicht dieselbe Definition wie im reellen Fall benutzen. Die Länge eines Vektors sollte eine nichtnegative reelle Zahl sein. Im Falle einer komplexen Zahl $z = x + iy$ ist der Betrag nicht durch $\sqrt{z^2}$ wie im reellen Fall gegeben, sondern durch

$$|z| = |x + iy| = \sqrt{x^2 + y^2} = \sqrt{(x + iy)(x - iy)} = \sqrt{z\bar{z}}.$$

Wir führen daher die folgende Schreibweise ein. Für jeden Vektor

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^n \quad \text{schreiben wir} \quad \bar{z} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_n \end{pmatrix}$$

für den *komplex konjugierten Vektor*. Das *Standard-(Hermitesche) Skalarprodukt* zweier Vektoren $z, w \in \mathbb{C}^n$ wird definiert als

$$\langle z, w \rangle = \left\langle \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \right\rangle := z_1 \overline{w_1} + z_2 \overline{w_2} + \cdots + z_n \overline{w_n} = z^\top \overline{w}.$$

Damit ist das Skalarprodukt zweier Vektoren in $\mathbb{C}^n$ eine komplexe Zahl. Aus der Definition folgt sofort, dass $\langle \cdot, \cdot \rangle$ hermitesch ist:

$$\langle w, z \rangle = \sum_{j=1}^n w_j \overline{z_j} = \overline{\sum_{j=1}^n z_j \overline{w_j}} = \overline{\langle z, w \rangle}$$

und es ist auch positiv definit:

$$\langle z, z \rangle = \sum_{j=1}^n |z_j|^2 > 0 \quad \text{für } z \neq 0.$$

Für die Länge eines Vektors erhalten wir

$$\|z\| := \sqrt{\langle z, z \rangle} = \sqrt{z_1 \overline{z_1} + z_2 \overline{z_2} + \cdots + z_n \overline{z_n}} = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}.$$

Bemerkung 7.1.4. Weitere Beispiele von euklidischen und unitären Vektorräumen werden wir später betrachten. Im Augenblick halten wir nur fest, dass jeder lineare Teilraum U eines euklidischen bzw. unitären Raums wieder ein euklidischer oder unitärer Vektorraum ist: Um U zu einem solchen Raum zu machen, schränken wir einfach das Skalarprodukt auf U ein.

Beachte, dass euklidische und unitäre Vektorräume unendlichdimensional sein können. Derartige Räume (insbesondere Funktionenräume) spielen eine bedeutende Rolle in der Analysis und in vielen anderen Bereichen der Mathematik.

Bemerkung 7.1.5. Ist V ein unitärer Raum, so ist V insbesondere ein komplexer Vektorraum. Wir können V aber auch als reellen Vektorraum auffassen, für den wir dann $V^{\mathbb{R}}$ schreiben. Hierzu schränken wir einfach die Skalarmultiplikation

$$\mathbb{C} \times V \rightarrow V, \quad (\lambda, v) \mapsto \lambda v$$

zu einer Abbildung

$$\mathbb{R} \times V \rightarrow V, \quad (\lambda, v) \mapsto \lambda v$$

ein. Weiter ist

$$(v, w) := \operatorname{Re} \langle v, w \rangle = \frac{1}{2}(\langle v, w \rangle + \langle w, v \rangle) = \frac{1}{2}(\langle v, w \rangle + \overline{\langle v, w \rangle})$$

ein reellwertiges Skalarprodukt mit dem $V^{\mathbb{R}}$ ein euklidischer Raum wird. (Nachweis als Übung!)

Wir können das komplexe Skalarprodukt leicht aus dem reellen rekonstruieren. Aus der Relation

$$i \operatorname{Im} \langle v, w \rangle = i \operatorname{Re}(-i \langle v, w \rangle) = i \langle -iv, w \rangle = i \langle v, iw \rangle$$

erhalten wir die *Rekonstruktionsformel*

$$\langle v, w \rangle = (v, w) + i \langle -iv, w \rangle = (v, w) + i \langle v, iw \rangle$$

(vgl. Aufgabe 7.1.3).

Im Folgenden sei V ein reeller oder komplexer unitärer Raum.

Bemerkung 7.1.6. Wir haben oben gesehen, wie man die Länge (Norm) eines Vektors durch das Skalarprodukt definiert. Es ist bemerkenswert, dass man das Skalarprodukt zweier Vektoren umgekehrt aus der Normfunktion rekonstruieren kann.

In einem euklidischen Vektorraum haben wir die *Polarisierungsidentität*:

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2)$$

und in einem unitären Vektorraum gilt die etwas kompliziertere *Polarisierungsidentität*:

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2).$$

Beide Identitäten verifiziert man leicht durch Nachrechnen (vgl. Aufgabe 7.1.1).

Satz 7.1.7. [Satz von PYTHAGORAS] Für zwei zueinander orthogonale Vektoren $x \perp y$ eines unitären Raums V gilt

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Sind $x_1, \dots, x_n \in V$ paarweise orthogonal, so gilt entsprechend

$$\left\| \sum_{j=1}^n x_j \right\|^2 = \sum_{j=1}^n \|x_j\|^2.$$

Beweis. Sei $x \perp y$. Definitionsgemäß gilt $\langle x, y \rangle = 0$, also auch $\langle y, x \rangle = \overline{\langle x, y \rangle} = 0$. Wir erhalten daher

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2.$$

Der zweite Teil der Behauptung folgt sofort durch Induktion nach der Anzahl n der Summanden. $\square$

Satz 7.1.8. [Die CAUCHY-SCHWARZsche Ungleichung] Für jede positiv semidefinite hermitesche Form $\beta: V \times V \rightarrow \mathbb{K}$ gilt die Cauchy-Schwarzsche Ungleichung

$$|\beta(x, y)|^2 \leq \beta(x, x)\beta(y, y) \quad \text{für } x, y \in V.$$

Für einen unitären Raum $(V, \langle \cdot, \cdot \rangle)$ erhalten wir insbesondere

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|.$$

Im folgenden schreiben wir kürzer “CS-Ungleichung” für diesen Sachverhalt.

Beweis. Zuerst bemerken wir, dass $\gamma(x, y) := \operatorname{Re} \beta(x, y)$ eine positiv semidefinite reell bilineare symmetrische Abbildung $\gamma: V \times V \rightarrow \mathbb{R}$ definiert. Wir erhalten daher für alle $\lambda \in \mathbb{R}$ die Ungleichung

$$0 \leq \gamma(x - \lambda y, x - \lambda y) = \gamma(x, x) - \lambda \gamma(y, x) - \lambda \gamma(x, y) + \lambda^2 \gamma(y, y) = \gamma(x, x) - 2\lambda \gamma(x, y) + \lambda^2 \gamma(y, y).$$

Ist $\gamma(y, y) = 0$, so muss also auch $\gamma(x, y) = 0$, da $\lambda \in \mathbb{R}$ beliebig ist. Ist $\gamma(y, y) \neq 0$, so erhalten wir für $\lambda := \gamma(x, y)/\gamma(y, y)$ die Relation

$$\begin{aligned} 0 &\leq \gamma(x, x) - 2\lambda \gamma(x, y) + \lambda^2 \gamma(y, y) = \gamma(x, x) - 2\gamma(x, y)^2/\gamma(y, y) + \gamma(x, y)^2/\gamma(y, y) \\ &= \gamma(x, x) - \gamma(x, y)^2/\gamma(y, y), \end{aligned}$$

also gilt

$$\gamma(x, y)^2 \leq \gamma(x, x)\gamma(y, y) \quad \text{für alle } x, y \in V.$$

Ist $\mathbb{K} = \mathbb{R}$, so ist $\gamma = \beta$ und wir haben die Cauchy-Schwarzsche Ungleichung bewiesen. Für $\mathbb{K} = \mathbb{C}$ schreiben wir $\beta(x, y) = |\beta(x, y)| \cdot \zeta$ mit $|\zeta| = 1$ (Polardarstellung komplexer Zahlen). Dann ist $\beta(\zeta^{-1}x, y) = \zeta^{-1}\beta(x, y) = |\beta(x, y)|$ reell, stimmt also mit $\gamma(\zeta^{-1}x, y)$ überein. Das führt zu

$$\begin{aligned} |\beta(x, y)| &= \gamma(\zeta^{-1}x, y) \leq \sqrt{\gamma(\zeta^{-1}x, \zeta^{-1}x)} \sqrt{\gamma(y, y)} = \sqrt{\beta(\zeta^{-1}x, \zeta^{-1}x)} \sqrt{\beta(y, y)} \\ &= \sqrt{\beta(x, x)} \sqrt{\beta(y, y)}, \end{aligned}$$

da $\beta(\zeta^{-1}x, \zeta^{-1}x) = \zeta^{-1}\bar{\zeta}^{-1}\beta(x, x) = |\zeta|^{-2}\beta(x, x) = \beta(x, x)$ ist. Durch Quadrieren erhalten wir die Cauchy-Schwarzsche Ungleichung auch im Komplexen und damit

$$|\langle x, y \rangle|^2 \leq \|x\|^2 \|y\|^2$$

für beliebige unitäre Räume über $\mathbb{R}$ oder $\mathbb{C}$. □

Definition 7.1.9. Sei V ein unitärer Raum. Der *Winkel* $\gamma \in [0, \pi[$ zwischen zwei nichtverschwindenden Vektoren v und w wird durch

$$\cos \gamma = \frac{\operatorname{Re}\langle v, w \rangle}{\|v\| \cdot \|w\|}$$

definiert. Nach der CAUCHY-SCHWARZschen Ungleichung wissen wir, dass $-1 \leq \frac{\operatorname{Re}\langle v, w \rangle}{\|v\| \cdot \|w\|} \leq 1$ gilt; es existiert also ein eindeutiges $\gamma \in [0, \pi[$ mit $\cos \gamma = \frac{\operatorname{Re}\langle v, w \rangle}{\|v\| \cdot \|w\|}$. Man beachte, dass das Winkelkonzept keinerlei Bezug auf die komplexe Skalarmultiplikation nimmt. Es wird lediglich das reelle Skalarprodukt $\operatorname{Re}\langle \cdot, \cdot \rangle$, d.h., der unterliegende euklidische Raum betrachtet. In diesem Sinne sind Winkel ein reelles Konzept.

Bemerkung 7.1.10. Wir fragen uns nun, wann für zwei Vektoren $x, y \neq 0$ sogar Gleichheit in der CS-Ungleichung gilt.

(a) Für $\mathbb{K} = \mathbb{R}$ erhalten wir aus der Gleichung

$$|\langle x, y \rangle|^2 = \langle x, x \rangle \langle y, y \rangle \tag{7.1}$$

und dem Beweis von Satz 7.1.8, dass für $\lambda := \langle x, y \rangle / \langle y, y \rangle$ sogar die Relation

$$0 = \|x - \lambda y\|^2,$$

also $x = \lambda y$ gilt. Umgekehrt gilt trivialerweise

$$|\langle \lambda y, y \rangle|^2 = \lambda^2 \langle y, y \rangle^2 = \langle \lambda y, \lambda y \rangle \langle y, y \rangle.$$

Für $0 \neq x, y \in V$ ist daher (7.1) äquivalent zu $x \in \mathbb{R}y$. Der Winkel zwischen x und y ist entweder 0 oder π .

Für $\mathbb{K} = \mathbb{R}$ liefert die CS-Ungleichung

$$\langle x, y \rangle \leq \|x\| \cdot \|y\|.$$

Gilt hier Gleichheit, so ist insbesondere $\langle x, y \rangle \geq 0$ und wir erhalten $x = \lambda y$ für ein $\lambda \geq 0$. In diesem Fall verschwindet der Winkel zwischen x und y .

(b) Für $\mathbb{K} = \mathbb{C}$ und $x, y \in V$ gelte

$$|\langle x, y \rangle|^2 = \langle x, x \rangle \langle y, y \rangle. \tag{7.2}$$

Wir wählen $\zeta \in \mathbb{C}$ mit $|\zeta| = 1$ wieder so, dass

$$|\langle x, y \rangle| = \zeta^{-1} \langle x, y \rangle = \langle \zeta^{-1}x, y \rangle = \operatorname{Re}\langle \zeta^{-1}x, y \rangle$$

gilt. Wenden wir nun (a) auf das reelle Skalarprodukt $\operatorname{Re}\langle x, y \rangle$ an, erhalten wir $x \in \zeta \mathbb{R}y$ für ein $\lambda > 0$, also $x \in \mathbb{C}y$. Umgekehrt zeigt eine direkte Rechnung, dass $x \in \mathbb{C}y$ zu Gleichheit in der CS-Ungleichung führt.

Gleichheit tritt in der CS-Ungleichung also genau dann auf, wenn x und y linear abhängig sind.

Beispiel 7.1.11. Da $\mathbb{C}^n$ mit dem Standardskalarprodukt ein unitärer Raum ist (Beispiel 7.1.3), erhalten wir durch Spezialisieren der allgemeinen Cauchy–Schwarzschen Ungleichung: Es gilt für alle Vektoren $z, w \in \mathbb{C}^n$:

$$|z_1 \overline{w_1} + z_2 \overline{w_2} + \cdots + z_n \overline{w_n}| \leq \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2} \sqrt{|w_1|^2 + |w_2|^2 + \cdots + |w_n|^2}.$$

Folgerung 7.1.12. [Die Dreiecksungleichung] Für Elemente x, y eines unitären Raums V gilt

$$\|x + y\| \leq \|x\| + \|y\|.$$

Beweis. Aus der CS-Ungleichung erhalten wir unmittelbar

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \end{aligned}$$

Wurzelziehen auf beiden Seiten ergibt die Dreiecksungleichung. $\square$

Folgerung 7.1.13. Ist V ein unitärer Raum, so definiert $\|\cdot\|$ eine Norm auf V , d.h.,

(N1) $(\forall x \in V) \|x\| \geq 0$ und $\|x\| = 0 \iff x = 0$.

(N2) $\|\lambda \cdot x\| = |\lambda| \cdot \|x\|$ für $x \in V$ und $\lambda \in \mathbb{K}$ (positive Homogenität).

(N3) $\|x + y\| \leq \|x\| + \|y\|$ für $x, y \in V$ (Subadditivität).

Beweis. (N1) folgt aus der positiven Definitheit des Skalarprodukts und (N3) ist die Dreiecksungleichung. (N2) folgt sofort aus

$$\|\lambda x\|^2 = \langle \lambda x, \lambda x \rangle = \lambda \overline{\lambda} \langle x, x \rangle = |\lambda|^2 \|x\|^2. \quad \square$$

Aufgaben zu Abschnitt 7.1

Aufgabe 7.1.1. Sei $\beta: V \times V \rightarrow \mathbb{C}$ eine Sesquilinearform auf dem komplexen Vektorraum V . Zeigen Sie, dass β die folgende *Polarisierungsidentität* erfüllt, die es erlaubt β aus seinen Werten auf Paaren der Form (v, v) zu rekonstruieren:

$$\beta(x, y) = \frac{1}{4} \sum_{k=0}^3 i^k \beta(x + i^k y, x + i^k y).$$

Aufgabe 7.1.2. Sei $\beta: V \times V \rightarrow \mathbb{R}$ eine Bilinearform auf dem reellen Vektorraum V . Zeigen Sie, dass β genau dann die *Polarisierungsidentität*

$$\beta(x, y) = \frac{1}{4} (\beta(x + y, x + y) - \beta(x - y, x - y))$$

erfüllt, wenn β symmetrisch ist.

Aufgabe 7.1.3. Sei V ein komplexer Vektorraum und $f: V \rightarrow \mathbb{R}$ eine reell lineare Abbildung. Zeigen Sie, dass durch

$$F: V \rightarrow \mathbb{C}, \quad F(v) := f(v) - if(iv)$$

eine komplex lineare Abbildung definiert wird.

Aufgabe 7.1.14. Zeige, dass in der Dreiecksungleichung für $y \neq 0$ Gleichheit genau dann gilt, wenn $x = \alpha y$ für eine reelle Zahl $\alpha \geq 0$ ist.

Aufgabe 7.1.15. Jeder euklidische (unitäre) Raum V wird durch

$$d(x, y) := \|x - y\|$$

zu einem metrischen Raum, d.h., für $x, y, z \in V$ gelten:

$$(M1) \quad d(x, y) = d(y, x).$$

$$(M2) \quad d(x, y) \geq 0 \text{ mit Gleichheit nur für } x = y.$$

$$(M3) \quad d(x, z) \leq d(x, y) + d(y, z).$$

Aufgabe 7.1.16. Für jedes Paar (p, q) von Punkten in einem euklidischen Raum V existiert genau ein Punkt $m = m_{p,q}$, genannt der *Mittelpunkt*, mit

$$d(p, m) = d(q, m) = \frac{1}{2}d(p, q).$$

Aufgabe 7.1.17. * Sei V ein euklidischer Raum und $\varphi: V \rightarrow V$ eine Abbildung mit $\varphi(0) = 0$ und

$$d(\varphi(x), \varphi(y)) = d(x, y),$$

d.h. φ ist isometrisch, bzw. erhält Abstände. Zeige: φ ist linear. Hinweis: Zeige zuerst, dass φ Mittelpunkte respektiert:

$$\varphi(m_{p,q}) = m_{\varphi(p), \varphi(q)},$$

dann $\varphi(2x) = 2\varphi(x)$ für alle $x \in \mathbb{R}$, dann die Additivität, und schließlich die Linearität.

Aufgabe 7.1.18. Seien $\varphi, \psi: V \rightarrow V$ zwei Endomorphismen des unitären Raums V . Zeigen Sie: Wenn

$$\langle \varphi(v), v \rangle = \langle \psi(v), v \rangle \quad \text{für alle } v \in V$$

gilt, dann ist $\varphi = \psi$. Hinweis: Aufgabe 7.1.1.

7.2 Orthonormalbasen

In diesem Abschnitt ist V wieder ein unitärer Raum über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 7.2.1. [Orthonormalbasen] Wir nennen ein System $v_1, v_2, \dots, v_m$ von Vektoren in V ein *Orthogonalsystem* (OS), wenn dessen Vektoren alle von 0 verschieden sind und wenn sie paarweise orthogonal sind, d.h. wenn gilt:

$$\langle v_i, v_j \rangle = 0, \quad \text{falls } i \neq j.$$

Sind zusätzlich alle v_i Einheitsvektoren, sprechen wir von einem *Orthonormalsystem* (ONS). Ein Orthonormalsystem wird durch die folgende Bedingung charakterisiert:

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

Ein Orthonormalsystem $v_1, v_2, \dots, v_n$, das eine Basis von V ist, heißt *Orthonormalbasis* (ONB).

Lemma 7.2.2. Jedes Orthogonalsystem $v_1, v_2, \dots, v_m$ und damit jedes Orthonormalsystem ist linear unabhängig.

Beweis. Zum Nachweis nehmen wir $\sum_{j=1}^m \lambda_j v_j = 0$ an. Dann folgt für jedes k :

$$0 = \left\langle \sum_{j=1}^m \lambda_j v_j, v_k \right\rangle = \sum_{j=1}^m \lambda_j \langle v_j, v_k \rangle = \lambda_k \langle v_k, v_k \rangle$$

Aus $\langle v_k, v_k \rangle \neq 0$ folgern wir $\lambda_k = 0$ für jedes k . □

Beispiel 7.2.3. (a) Ist $\dim V = n$, so ist ein Orthonormalsystem aus n Vektoren eine Orthonormalbasis.

(b) Die Standardbasis in $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ ist eine Orthonormalbasis bzgl. des Standardskalarprodukts.

(c) Im $\mathbb{R}^2$ bilden für jedes $\gamma \in \mathbb{R}$ die Vektoren

$$v_1 = \begin{pmatrix} \cos \gamma \\ \sin \gamma \end{pmatrix}, \quad v_2 = \begin{pmatrix} -\sin \gamma \\ \cos \gamma \end{pmatrix}$$

ebenfalls eine Orthonormalbasis.

Bemerkung 7.2.4. [Koordinaten bzgl. einer Orthonormalbasis] Sei $v_1, v_2, \dots, v_n$ eine Orthonormalbasis von V . Dann hat ein beliebiger Vektor $v \in V$ eine Darstellung als Linearkombination

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = \sum_{j=1}^n \lambda_j v_j$$

Bildung des Skalarproduktes mit v_i auf beiden Seiten ergibt

$$\langle v, v_i \rangle = \left\langle \sum_{j=1}^n \lambda_j v_j, v_i \right\rangle = \sum_{j=1}^n \lambda_j \langle v_j, v_i \rangle = \sum_{j=1}^n \lambda_j \delta_{ij} = \lambda_i.$$

Also ist die i -te Koordinate von v bzgl. der Orthonormalbasis $v_1, v_2, \dots, v_n$ gleich $\lambda_i = \langle v, v_i \rangle$ und die Darstellung wird zu

$$v = \langle v, v_1 \rangle v_1 + \langle v, v_2 \rangle v_2 + \dots + \langle v, v_n \rangle v_n.$$

Beispiel 7.2.5. Im euklidischen $\mathbb{R}^3$ haben die Vektoren

$$v_1 = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad v_2 = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \quad \text{und} \quad v_3 = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$$

Länge 1 und sind paarweise orthogonal, wie man leicht zeigt. Sie bilden also eine Orthonormalbasis von $\mathbb{R}^3$. Bzgl. dieser Basis hat der Vektor

$$v = \begin{pmatrix} -4 \\ 2 \\ 2 \end{pmatrix}$$

die Darstellung $v = \langle v_1, v \rangle v_1 + \langle v_2, v \rangle v_2 + \langle v_3, v \rangle v_3 = \frac{4}{3} v_1 - \frac{10}{3} v_2 - \frac{10}{3} v_3$.

7.2.1 Das Gram–Schmidt Verfahren

7.2.6. [Das GRAM–SCHMIDTsche Orthogonalisierungsverfahren] Wir betrachten mit $b_1, b_2, \dots, b_m$ ein System von m linear unabhängigen Vektoren in einem unitären $\mathbb{K}$ -Vektorraum V . Diese spannen einen m -dimensionalen linearen Teilraum U von V auf. Wie wir unten zeigen werden, ergibt die folgende Prozedur ein Orthogonalsystem $u_1, u_2, \dots, u_m$ und nach Normierung eine Orthonormalbasis $v_1, v_2, \dots, v_m$ von U . Beginnen wir insbesondere mit einer Basis von V , so

erhalten wir eine Orthonormalbasis von V :

$$\begin{aligned}
 u_1 &:= b_1, & v_1 &:= \frac{u_1}{\|u_1\|}, \\
 u_2 &:= b_2 - \langle b_2, v_1 \rangle v_1, & v_2 &:= \frac{u_2}{\|u_2\|}, \\
 u_3 &:= b_3 - \langle b_3, v_1 \rangle v_1 - \langle b_3, v_2 \rangle v_2, & v_3 &:= \frac{u_3}{\|u_3\|}, \\
 &\vdots & &\vdots \\
 u_m &:= b_m - \sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i, & v_m &:= \frac{u_m}{\|u_m\|},
 \end{aligned}$$

Für praktische Zwecke schreibt man das Schema oftmals in der folgenden Weise:

$$\begin{aligned}
 u_1 &:= b_1, \\
 u_2 &:= b_2 - \frac{\langle b_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1, \\
 u_3 &:= b_3 - \frac{\langle b_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle b_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2, \\
 &\vdots \\
 u_m &:= b_m - \sum_{i=1}^{m-1} \frac{\langle b_m, u_i \rangle}{\langle u_i, u_i \rangle} u_i.
 \end{aligned}$$

Wir zeigen nun das Folgende:

1. $u_m \neq 0$ (um weiter durch $\|u_m\|$ dividieren zu können und $v_m = \frac{u_m}{\|u_m\|}$ zu erhalten.)
2. $\text{span}(v_1, \dots, v_{m-1}, v_m) = \text{span}(b_1, \dots, b_{m-1}, b_m)$
3. $v_1, \dots, v_{m-1}, v_m$ ist ein Orthonormalsystem.

Beweis. Wir führen den Beweis durch Induktion nach m . Für $m = 1$ sind die Behauptungen offensichtlich. Wir nehmen nun an, die Behauptungen seien bewiesen für $m - 1$ Vektoren und wir beweisen sie für m :

1. Ist $u_m = 0$, so folgt aus der letzten Gleichung

$$b_m = \sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i \in \text{span}(v_1, \dots, v_{m-1}) = \text{span}(b_1, \dots, b_{m-1})$$

nach der Induktionsannahme. Dies widerspricht der Tatsache, dass die Vektoren $b_1, \dots, b_{m-1}, b_m$ linear unabhängig sind.

2. Wir wissen aus der Induktionsannahme, dass $\text{span}(v_1, \dots, v_{m-1}) = \text{span}(b_1, \dots, b_{m-1})$ ist; weiter gilt

$$v_m := \frac{u_m}{\|u_m\|} = \frac{1}{\|u_m\|} \left(b_m - \underbrace{\sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i}_{\in \text{span}(b_1, \dots, b_{m-1})} \right) \in \text{span}(b_1, \dots, b_{m-1}, b_m)$$

und

$$b_m = \|u_m\| v_m + \sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i \in \text{span}(v_1, \dots, v_{m-1}, v_m).$$

3. Nach der Induktionsannahme ist $v_1, \dots, v_{m-1}$ ein Orthonormalsystem. Aus der Definition von v_m folgt $\|v_m\| = 1$. Es bleibt zu zeigen, dass $v_i \perp v_m$ für alle $i = 1, \dots, m-1$ gilt:

$$\begin{aligned} \langle v_i, v_m \rangle &= \langle v_i, \frac{u_m}{\|u_m\|} \rangle = \frac{1}{\|u_m\|} \langle v_i, b_m - \sum_{j=1}^{m-1} \langle b_m, v_j \rangle v_j \rangle \\ &= \frac{1}{\|u_m\|} \left(\langle v_i, b_m \rangle - \sum_{j=1}^{m-1} \langle v_j, b_m \rangle \underbrace{\langle v_i, v_j \rangle}_{\delta_{ij}} \right) = \frac{1}{\|u_m\|} (\langle v_i, b_m \rangle - \langle v_i, b_m \rangle) = 0. \quad \square \end{aligned}$$

Bemerkung 7.2.7. Ist $b_1, \dots, b_m$ ein Erzeugendensystem eines euklidischen oder unitären Vektorraums V , das nicht linear unabhängig ist, so kann man eine Variante des Gram–Schmidtschen Verfahrens benutzen, um eine Orthonormalbasis zu konstruieren. Man hat es allerdings wie folgt zu modifizieren:

Sind $b_1, \dots, b_k$ linear unabhängig, so durchläuft man k Schritte des Gram–Schmidt Verfahrens und erhält ein Orthonormalsystem $v_1, \dots, v_k$.

Ist nun b_{k+1} linear abhängig von $b_1, \dots, b_k$, also die Menge $\{b_1, \dots, b_{k+1}\}$ linear abhängig, so ist b_{k+1} eine Linearkombination von $v_1, \dots, v_k$ und daher

$$b_{k+1} = \langle b_{k+1}, v_1 \rangle v_1 + \dots + \langle b_{k+1}, v_k \rangle v_k,$$

also $u_{k+1} = 0$. In diesem Fall dürfen wir b_{k+1} getrost weglassen und fahren fort mit dem nächsten Element b_{k+2} .

Als Algorithmus ergibt sich somit folgende Modifikation des Gram–Schmidt Verfahrens: Tritt bei der Durchführung des Verfahrens die Situation $u_j = 0$ auf, so lassen wir b_j weg und fahren fort mit b_{j+1} .

Das so modifizierte Verfahren eliminiert sukzessive alle “linear überflüssigen” Elemente des Erzeugendensystems $\{b_1, \dots, b_m\}$.

Da jeder endlichdimensionale lineare Teilraum U eines Vektorraumes eine Basis hat, können wir das GRAM–SCHMIDTSche Orthogonalisierungsverfahren anwenden und erhalten den folgenden Satz:

Satz 7.2.8. *Jeder endlichdimensionale lineare Teilraum eines unitären $\mathbb{K}$ -Vektorraumes hat eine Orthonormalbasis.*

Beispiel 7.2.9. Sei U der lineare Teilraum des euklidischen $\mathbb{R}^4$, der durch die homogene lineare Gleichung

$$x_1 + x_2 + x_3 + x_4 = 0$$

gegeben ist. Man zeigt leicht, dass die folgenden Vektoren Lösungen dieser Gleichung sind und dass sie linear unabhängig sind. Sie bilden also eine Basis von U

$$b_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 3 \\ 1 \\ -1 \\ -3 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

Wir benutzen nun das GRAM–SCHMIDTSche Orthogonalisierungsverfahren, um eine orthonormale

Basis von U zu finden:

$$u_1 = b_1, \quad \|u_1\| = 2, \quad v_1 = \frac{u_1}{\|u_1\|} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix},$$

$$u_2 = b_2 - \langle b_2, v_1 \rangle v_1 = b_2 - 2v_1 = \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix}, \quad \|u_2\| = 4, \quad v_2 = \frac{u_2}{\|u_2\|} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

$$u_3 = b_3 - \langle b_3, v_1 \rangle v_1 - \langle b_3, v_2 \rangle v_2 = b_3 - 2v_1 - 2v_2 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \quad \|u_3\| = 2,$$

$$v_3 = \frac{u_3}{\|u_3\|} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

Definition 7.2.10. Eine bijektive lineare Abbildung $\varphi: V \rightarrow W$ zwischen euklidischen (unitären) Räumen heißt *Isomorphismus unitärer Räume* oder *unitär*, wenn

$$\langle \varphi(x), \varphi(y) \rangle = \langle x, y \rangle$$

für alle $x, y \in V$ gilt.

Satz 7.2.11. Sei V ein unitärer $\mathbb{K}$ -Vektorraum der Dimension n und $B = (v_1, \dots, v_n)$ eine Orthonormalbasis von V (die immer existiert). Dann ist die zugehörige Koordinatenabbildung

$$\kappa_B: V \rightarrow \mathbb{K}^n, \quad v \mapsto [v]_B = \begin{pmatrix} \langle v, v_1 \rangle \\ \langle v, v_2 \rangle \\ \vdots \\ \langle v, v_n \rangle \end{pmatrix},$$

die jeden Vektor $v \in V$ auf seine Koordinaten bzgl. B abbildet, ein Isomorphismus von unitären Räumen, wobei $\mathbb{K}^n$ mit dem Standardskalarprodukt versehen ist.

Beweis. Wir wissen aus Satz 7.2.8, dass V eine Orthonormalbasis besitzt und wir wissen auch schon, dass für

$$v = \alpha_1 v_1 + \dots + \alpha_n v_n$$

die Relation $\langle v, v_j \rangle = \alpha_j$ gilt.

Wir zeigen jetzt, dass κ_B ein Isomorphismus unitärer Räume ist, d.h. für alle $v, w \in V$ gilt

$$\langle \kappa_B(v), \kappa_B(w) \rangle_{\mathbb{K}^n} = \langle v, w \rangle_V.$$

Wir berechnen hierzu das Skalarprodukt zweier Vektoren v und w , die durch die Koordinaten

$$v = \sum_{i=1}^n \alpha_i v_i, \quad w = \sum_{j=1}^n \beta_j v_j$$

bzgl. der Basis B dargestellt werden:

$$\begin{aligned} \langle v, w \rangle &= \left\langle \sum_{i=1}^n \alpha_i v_i, \sum_{j=1}^n \beta_j v_j \right\rangle = \sum_{i=1}^n \alpha_i \cdot \left\langle v_i, \sum_{j=1}^n \beta_j v_j \right\rangle \\ &= \sum_{i=1}^n \alpha_i \cdot \sum_{j=1}^n \overline{\beta_j} \langle v_i, v_j \rangle_V = \sum_{i,j=1}^n \alpha_i \overline{\beta_j} \delta_{ij} = \sum_{i=1}^n \alpha_i \overline{\beta_i} = \langle \kappa_B(v), \kappa_B(w) \rangle_{\mathbb{K}^n}. \end{aligned} \quad \square$$

7.2.2 Orthogonalprojektionen

Definition 7.2.12. [Orthogonale lineare Teilräume] Sei V ein euklidischer oder unitärer Vektorraum. Für jede Teilmenge $M \subseteq V$ definieren wir den *Orthogonalraum* von M durch

$$M^\perp := \{v \in V : (\forall a \in M) \langle a, v \rangle = 0\}.$$

Lemma 7.2.13. $M^\perp$ ist ein linearer Teilraum von V .

Beweis. (U1) $0 \in M^\perp$, da $\langle a, 0 \rangle = 0$ für alle $a \in M$.

(U2) Sei $v, w \in M^\perp$. Für alle $a \in V$ gilt dann $\langle a, v \rangle = 0$ und $\langle a, w \rangle = 0$. Daraus folgt

$$\langle a, v + w \rangle = \langle a, v \rangle + \langle a, w \rangle = 0$$

und damit $v + w \in M^\perp$.

(U3) Sei $v \in M^\perp$ und $\lambda \in \mathbb{K}$. Dann gilt für jedes $a \in M$ die Beziehung $\langle a, v \rangle = 0$. Also gilt $\langle a, \lambda v \rangle = \lambda \langle a, v \rangle = 0$ und damit $\lambda v \in M^\perp$. □

Bemerkung 7.2.14. [Spezialfall] Wähle $V = \mathbb{R}^n$ und ein festes $a \in \mathbb{R}^n$. Dann ist $\{a\}^\perp = \{v \in \mathbb{R}^n : \langle a, v \rangle = 0\}$. Für

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{und} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

kann die Gleichung $\langle a, v \rangle = 0$ geschrieben werden als

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0.$$

Also besteht $\{a\}^\perp$ aus den Lösungen v dieser homogenen linearen Gleichung. Das ist eine *Hyperebene* von $\mathbb{R}^n$, also ein linearer Unterraum der Dimension $n - 1$.

Analog dazu besteht für $a, b, c, \dots \in \mathbb{R}^n$ der lineare Teilraum $\{a, b, c, \dots\}^\perp$ aus den Lösungen v des Systems

$$\langle a, v \rangle = 0, \quad \langle b, v \rangle = 0, \quad \langle c, v \rangle = 0, \quad \dots$$

von homogenen linearen Gleichungen.

Lemma 7.2.15. Seien $a_1, a_2, \dots, a_m \in V$ und sei U der lineare Teilraum, der von diesen Vektoren aufgespannt wird. Dann gilt $\{a_1, a_2, \dots, a_m\}^\perp = U^\perp$.

Beweis. Offensichtlich ist $U^\perp \subseteq \{a_1, a_2, \dots, a_m\}^\perp$.

Für die umgekehrte Inklusion betrachten wir ein $v \in \{a_1, a_2, \dots, a_m\}^\perp$. Dann ist $\langle a_1, v \rangle = 0$, $\langle a_2, v \rangle = 0$, ..., $\langle a_m, v \rangle = 0$. Sei a ein Element von U . Da der Teilraum U durch $a_1, a_2, \dots, a_m$ aufgespannt wird, kann a als Linearkombination

$$a = \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_m a_m$$

dargestellt werden. Also folgt

$$\langle a, v \rangle = \langle \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_m a_m, v \rangle = \lambda_1 \langle a_1, v \rangle + \lambda_2 \langle a_2, v \rangle + \cdots + \lambda_m \langle a_m, v \rangle = 0.$$

Damit ergibt sich $\langle a, v \rangle = 0$ für jedes $a \in U$, also $v \in U^\perp$. □

Satz 7.2.16. [Orthogonalprojektion auf einen linearen Teilraum U] Sei U ein endlichdimensionaler linearer Teilraum eines euklidischen oder unitären Vektorraumes V und $v_1, \dots, v_m$ eine Orthonormalbasis von U . Definiere $\pi : V \rightarrow V$ durch

$$\pi(v) = \sum_{i=1}^m \langle v, v_i \rangle v_i$$

für alle $v \in V$.

Dann gelten die folgenden Eigenschaften:

- (i) π ist eine lineare Abbildung.
- (ii) $\pi(v) \in U$ für alle $v \in V$.
- (iii) $\pi(u) = u$ für alle $u \in U$.
- (iv) $\pi \circ \pi = \pi$.
- (v) $\text{im } \pi = U$ und $\ker \pi = U^\perp$.
- (vi) $v - \pi(v) \in U^\perp$ für alle $v \in V$.
- (vii) $\|v - \pi(v)\| \leq \|v - u\|$ für alle $u \in U$ und Gleichheit gilt nur für $u = \pi(v)$, d.h., $\pi(v)$ ist der eindeutig bestimmte Punkt von U mit kürzestem Abstand von v . In anderen Worten: $\pi(u)$ die beste Approximation von v durch ein Element aus U .

Die oben definierte Abbildung π heißt die *Orthogonalprojektion von V auf U* . Aus (vii) folgt insbesondere, dass die Abbildung π nicht von der Wahl der ONB in U abhängt.

Beweis. (i) Für jedes $w \in V$ ist die Abbildung

$$w^* : V \rightarrow \mathbb{K}, \quad v \mapsto \langle v, w \rangle$$

linear. Hieraus schliesst man leicht, dass auch $\pi = \sum_{i=1}^m v_i^* \cdot v_i$ linear ist (Nachweis!).

(ii) In der Tat ist $\pi(v)$ eine Linearkombination der Basisvektoren von U .

(iii) Ist $u = \sum_{j=1}^m \lambda_j v_j$ ein Element von U , so folgt

$$\pi(u) = \sum_{i=1}^m \left\langle \sum_{j=1}^m \lambda_j v_j, v_i \right\rangle v_i = \sum_{i=1}^m \sum_{j=1}^m \lambda_j \underbrace{\langle v_j, v_i \rangle}_{\delta_{ij}} v_i = \sum_{i=1}^m \lambda_i v_i = u.$$

(iv) Nach (ii) ist $\pi(v) \in U$ und damit $\pi(\pi(v)) = \pi(v)$ wegen (iii).

(v) Die Behauptung für das Bild folgt aus (ii) und (iii). Man beachte für den Kern:

$$\begin{aligned} \pi(v) = 0 &\iff \langle v, v_i \rangle = 0 \quad \text{für alle } i = 1, \dots, m \\ &\iff v \in \{v_1, \dots, v_m\}^\perp = U^\perp. \end{aligned} \quad (7.2.15)$$

(vi) Für $v \in V$ ergibt sich

$$\pi(v - \pi(v)) \underset{(i)}{=} \pi(v) - \pi(\pi(v)) \underset{(iv)}{=} \pi(v) - \pi(v) = 0$$

und damit $v - \pi(v) \in \ker \pi = U^\perp$.
(v)

(vii) Für jedes $u \in U$ gilt $\pi(v) - u \in U$, da U ein linearer Teilraum ist. Außerdem gilt $v - \pi(v) \in U^\perp$ wegen (vi). Also sind diese beiden Vektoren orthogonal zueinander. Damit folgt

$$\begin{aligned} \|v - u\|^2 &= \|(v - \pi(v)) + (\pi(v) - u)\|^2 \underset{\text{Pyth.}}{=} \|v - \pi(v)\|^2 + \|\pi(v) - u\|^2 \\ &\geq \|v - \pi(v)\|^2. \end{aligned}$$

Gilt Gleichheit, so zeigt die obige Rechnung, dass $\pi(v) - u = 0$ ist, also $u = \pi(v)$. $\square$

Satz 7.2.17. Für jeden endlichdimensionalen linearen Teilraum U von V gilt

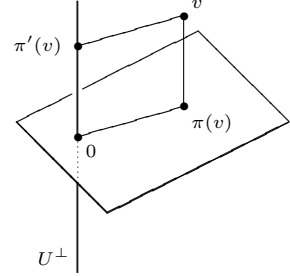
$$V = U \oplus U^\perp.$$

Insbesondere gilt für endlichdimensionales V : $\dim U^\perp = \dim V - \dim U$.

Beweis. (i) $U \cap U^\perp = \{0\}$, da für $v \in U$ und $v \in U^\perp$ die Beziehung $\langle v, v \rangle = 0$ gilt, also $v = 0$.
(ii) $U + U^\perp = V$, da für jedes $v \in V$

$$v = \pi(v) + (v - \pi(v)) \in U + U^\perp$$

gilt, wobei π die Orthogonalprojektion auf U ist. $\square$



Folgerung 7.2.18. Für jeden endlichdimensionalen Unterraum U von V gilt

$$(U^\perp)^\perp = U.$$

Beweis. Die Inklusion $U \subseteq (U^\perp)^\perp$ ist trivial. Um die andere Inklusion zu beweisen, erinnern wir uns an die Orthogonalprojektion $\pi: V \rightarrow U$. Für $v \in (U^\perp)^\perp$ ist dann einerseits $\pi(v) \in U \subseteq (U^\perp)^\perp$, also auch $v - \pi(v) \in (U^\perp)^\perp$ und andererseits $v - \pi(v) \in \ker \pi = U^\perp$, also

$$\|v - \pi(v)\|^2 = \langle v - \pi(v), v - \pi(v) \rangle = 0.$$

Also ist $v = \pi(v) \in U$.

Zweiter Beweis: Aus der orthogonalen Zerlegung $V = U \oplus U^\perp$ folgt daher umgekehrt

$$(U^\perp)^\perp = U + ((U^\perp)^\perp \cap U^\perp) = U. \quad \square$$

Bemerkung 7.2.19. Sei V ein endlichdimensionaler euklidischer Raum und $H \subseteq V$ eine Hyperebene. Dann ist $\dim H^\perp = \dim V - \dim H = 1$, also $H^\perp = \mathbb{R}n_H$ mit einem Einheitsvektor $n_H \in V$, den wir einen *Normalenvektor von H* nennen. Damit erhalten wir die Darstellung

$$H = (H^\perp)^\perp = (\mathbb{R}n_H)^\perp = n_H^\perp = \{v \in V: \langle v, n_H \rangle = 0\}.$$

Eine affine Hyperebene der Gestalt $p + H$ lässt sich daher auch beschreiben als

$$p + H = \{v \in V: \langle v, n_H \rangle = \langle p, n_H \rangle\}.$$

Also lassen sich alle affinen Hyperebenen $A \subseteq V$ in der Gestalt

$$A = \{v \in V: \langle v, n \rangle = d\}$$

mit $d \in \mathbb{R}$ und einem Einheitsvektor n darstellen, der senkrecht auf A (genauer auf dem zugehörigen Untervektorraum) steht.

Satz 7.2.20. (Darstellungssatz von Fréchet–Riesz) Ist V ein endlichdimensionaler unitärer Raum, so existiert zu jeder linearen Abbildung $\alpha: V \rightarrow \mathbb{K}$ genau ein Vektor $v_\alpha \in V$ mit

$$\alpha(x) = \langle x, v_\alpha \rangle \quad \text{für alle } x \in V. \quad (7.3)$$

Wir erhalten so eine antilineare Bijektion

$$\Gamma: V \rightarrow \text{Hom}(V, \mathbb{K}), \quad \Gamma(v)(x) := \langle x, v \rangle, \quad x \in V.$$

Beweis. Eindeutigkeit: Ist $\alpha(x) = \langle x, v \rangle = \langle x, w \rangle$ für alle $x \in V$, so ist $\langle x, v - w \rangle = 0$ für alle x , also $v - w \in V^\perp = \{0\}$.

Existenz: Für $\alpha = 0$ erfüllt $v = 0$ die Bedingung. Sei also $\alpha \neq 0$. Dann ist $H := \ker \alpha \subseteq V$ eine Hyperebene und daher $H^\perp = \mathbb{K}v_0$ für ein $v_0 \in V$ (Satz 7.2.17). Für $v_\alpha := \frac{\alpha(v_0)}{\|v_0\|^2}v_0$ gilt dann

$$\langle x, v_\alpha \rangle = \alpha(x) = 0 \quad \text{für} \quad x \in H \quad \text{und} \quad \langle v_0, v_\alpha \rangle = \alpha(v_0).$$

Aus $V = H \oplus \mathbb{K}v_0$ folgt daher (7.3).

Hiermit ist gezeigt, dass die Abbildung Γ surjektiv ist. Für $\lambda, \mu \in \mathbb{K}, v, w \in V$ ist

$$\Gamma(\lambda v + \mu w)(x) = \langle x, \lambda v + \mu w \rangle = \bar{\lambda} \langle x, v \rangle + \bar{\mu} \langle x, w \rangle = (\bar{\lambda} \Gamma(v) + \bar{\mu} \Gamma(w))(x).$$

Also ist Γ reell linear, aber antilinear für $\mathbb{K} = \mathbb{C}$. □

7.2.21. [Ein Hauch von Quantenmechanik] Das *Stern-Gerlach Experiment* funktioniert wie folgt: ein horizontaler Strahl von Partikeln (Silberatomen) wird durch ein Paar speziell präparierter Magnete gesendet. Das führt zu einer Aufteilung des Strahls in zwei Strahlen: einer wird nach oben abgelenkt (um etwa 30°) und der andere wird nach unten abgelenkt (um etwa -30°). Der Physiker sagt, die nach oben gehenden Partikel haben Spin $1/2$ und die nach unten gehenden haben Spin $-1/2$. Das magnetische Feld kann also die Partikel mit Spin $1/2$ von den Partikeln mit Spin $-1/2$ trennen.

Sendet man den Strahl mit Spin $1/2$ durch eine andere Anordnung gleich orientierter Magnete, so findet keine weitere Trennung mehr statt: der eintretende Strahl wird nach oben abgelenkt (um etwa 30°). Dasselbe gilt für den Strahl mit Spin $-1/2$.

Senden wir nun den Strahl mit Spin $1/2$ durch ein Paar von Magneten, die um 90° gedreht worden sind, so findet aber wiederum eine Aufspaltung in zwei Strahlen statt: einer wird nach links (um etwa 30°), der andere nach rechts (um etwa -30°) abgelenkt. Folgendes ist überraschend: wählen wir denjenigen der vier Strahlen, der durch die ersten Magnete nach oben abgelenkt wurde und durch das zweite Paar geschickt wurde, und senden ihn durch eine andere Anordnung von Magneten, die wieder wie die ersten Magneten orientiert sind, so erhalten wir wieder eine Aufspaltung in zwei Strahlen, der eine aufwärts abgelenkt, der andere nach unten. Es scheint so, als ob die dazwischenliegende Ablenkung, oder Polarisierung wie es die Physiker bezeichnen, den Spin nochmals verschoben hätte.

Wir stellen nun einige Überlegungen dazu an, wie dieses ungewöhnliche Verhalten mit dem unitären Vektorraum $\mathbb{C}^2$ modelliert werden kann: Der *Zustand* eines Partikels soll hierbei einem eindimensionalen Teilraum von $\mathbb{C}^2$ entsprechen, der durch einen Einheitsvektor $s \in \mathbb{C}^2$ dargestellt werden kann.

Die Ablenkung durch die anfängliche Menge von Magneten kann als sog. *Polarisierung* bzgl. der Zerlegung $\mathbb{C}^2 = U_1 \oplus U_2$ mit $U_1 = \mathbb{C}e_1$ und $U_2 = \mathbb{C}e_2$ verstanden werden. Das heißt, dass nach der Polarisierung der Zustand eines Partikels entweder durch U_1 oder U_2 repräsentiert sein wird, was mit Wahrscheinlichkeit $\|\pi_j(s)\|^2$ geschieht, wobei $\pi_j : \mathbb{C}^2 \rightarrow U_j$ die Projektion auf U_j ist. Nach der Polarisierung ist der Zustand des Partikels repräsentiert durch $s' := \pi_j(s)/\|\pi_j(s)\|$, wobei j die Nummer des Teilraums ist, zu dem die Polarisierung geführt hat. Man erinnere sich, dass $j = 1$ oder $j = 2$ i.A. beide möglich sind; ein Partikel j tritt mit der Wahrscheinlichkeit $\|\pi_j(s)\|^2$ auf. Der neue Zustand ist also entweder in U_1 oder in U_2 , so dass eine weitere Polarisierung nichts Neues mehr bringt. Ist nämlich $s' \in U_1$, so folgt $\|\pi_1(s')\|^2 = 1$ und $\|\pi_2(s')\|^2 = 0$, so dass wir mit Wahrscheinlichkeit 1 wieder in U_1 landen werden. Insbesondere ist der neue Zustand nach der zweiten Polarisierung wieder durch s' repräsentiert.

Wir bemerken, dass auch bei bekanntem Zustand das Ergebnis der Polarisierung nicht exakt bestimmt werden kann. Nur dessen Wahrscheinlichkeitsverteilung kann aus dem Ausgangszustand bestimmt werden. Wir bemerken auch, dass die Polarisierung den Zustand durch Projektion auf einen Teilraum verändert.

Wie kann die Ablenkung durch die um 90° gedrehten Magnete erklärt werden? Diese werden als eine Polarisierung bzgl. der Zerlegung $\mathbb{C}^2 = V_1 \oplus V_2$ angesehen, wobei $V_1 = \mathbb{C}(e_1 + e_2)$ und

$V_2 = \mathbb{C}(e_1 - e_2)$ gilt. Betrachten wir, was mit $s \in U_1$ geschieht, d.h. $s = e_1$ bzgl. dieser Polarisierung. Dies entspricht einer Sendung des nach oben abgelenkten Strahls durch die gedrehte Menge von Magneten, die wie betrachtet zu einer Aufspaltung in zwei Strahlen führt. Die Projektion von s auf V_1 ist gleich $\frac{1}{2}(e_1 + e_2)$, so dass die Wahrscheinlichkeit hierfür $(1/\sqrt{2})^2 = 1/2$ beträgt. Analogerweise ist die Wahrscheinlichkeit, dass s auf V_2 projiziert wird, gleich $1/2$, was die Aufspaltung in zwei Strahlen erklärt. Nach dieser Polarisierung ist der Zustand entweder repräsentiert durch $\mathbb{C}(e_1 + e_2)$ oder $\mathbb{C}(e_1 - e_2)$.

Quantenmechanische Experimente können durch unitäre Vektorräume modelliert werden; Zustände werden durch eindimensionale Unterräume repräsentiert, die durch Einheitsvektoren dargestellt werden (zwei Einheitsvektoren werden als äquivalent betrachtet, wenn einer aus dem anderen durch Multiplikation mit einer Zahl λ vom Betrag 1 gewonnen werden kann, d.h., wenn er den gleichen Unterraum aufspannt.)

Ein Experiment mit ℓ möglichen Ergebnissen wird durch eine orthogonale Zerlegung $V = U_1 \oplus \dots \oplus U_\ell$ beschrieben. Ist der Ausgangszustand durch den Einheitsvektor s gegeben und

$$s = s_1 + \dots + s_\ell$$

die Zerlegung von s in Komponenten $s_j \in U_j$, so folgt aus wiederholtem Anwenden des Satzes von Pythagoras die Beziehung

$$1 = \|s\|^2 = \|s_1\|^2 + \dots + \|s_\ell\|^2.$$

Die Wahrscheinlichkeit, nach dem Experiment den Endzustand j zu messen, ist nun $\|s_j\|^2$. Führt man eine Messung durch, ob sich das System im Zustand j befindet, so ist der neue Zustand repräsentiert durch den Einheitsvektor $\frac{1}{\|s_j\|} s_j$, wenn $s_j \neq 0$ gilt. Ist $s_j = 0$, so kann sich das System gar nicht im Zustand j befinden. Es gibt keine Möglichkeit, Information über den Zustand s zu erhalten, ohne s zu modifizieren.

Experimente mit unendlich vielen Ergebnissen, z.B. die Messung von Position oder Impuls eines Partikels in $\mathbb{R}^3$ können nicht mit endlichdimensionalen Vektorräumen beschrieben werden. Eine besondere Art von unendlichdimensionalen Vektorräumen wird benötigt: ein *Hilbertraum*, d.h., ein unitärer Raum V , der bzgl. der Metrik $d(x, y) := \|x - y\|$ vollständig ist.

Beispiel 7.2.22. [Der Hilbertraum ℓ^2] Sei $\ell^2 := \ell^2(\mathbb{N}, \mathbb{C})$ die Menge aller Folgen

$$x = (x_j)_{j=1,2,\dots} = (x_1, x_2, x_3, \dots)$$

von komplexen Zahlen mit

$$\sum_{j=1}^{\infty} |x_j|^2 < \infty.$$

Wir definieren für zwei Folgen $x = (x_j)$ und $y = (y_j)$ in ℓ^2

$$x + y := (x_j + y_j) \quad \text{und} \quad \lambda x = (\lambda x_j) \quad \text{für } \lambda \in \mathbb{C}.$$

Aufgabe 7.2.23. Zeige $\sum_{i=1}^{\infty} |x_j + y_j|^2 < \infty$, $\sum_{i=1}^{\infty} |\lambda x_j|^2 < \infty$, und dass $\sum_{i=1}^{\infty} x_j \overline{y_j}$ absolut konvergiert.

Ist also $x, y \in \ell^2$, so folgt $x + y \in \ell^2$ und $\lambda x \in \ell^2$. So wird ℓ^2 zu einem komplexen Vektorraum. Wir definieren für $x, y \in \ell^2$ ein Skalarprodukt durch

$$\langle x, y \rangle := \sum_{j=1}^{\infty} x_j \overline{y_j}$$

(Da die unendliche Reihe absolut konvergiert, beschreibt die Reihe eine komplexe Zahl.) Man zeigt leicht, dass $\langle \cdot, \cdot \rangle$ eine positiv definite hermitesche Form ist, so dass ℓ^2 zu einem komplexen unitären Raum wird.

Die Norm von $x = (x_j) \in \ell^2$ ist $\|x\| = \sqrt{\sum_{j=1}^{\infty} |x_j|^2}$. Die Vektoren

$$\begin{aligned} e_1 &= (1, 0, 0, 0, \dots) \\ e_2 &= (0, 1, 0, 0, \dots) \\ e_3 &= (0, 0, 1, 0, \dots) \\ &\vdots \end{aligned}$$

bilden ein unendliches Orthonormalsystem für ℓ^2 .

Beispiel 7.2.24. Sei $C([0, 1]) := C([0, 1], \mathbb{K})$ die Menge aller stetigen Funktionen $f : [0, 1] \rightarrow \mathbb{K}$. Bzgl. der gewöhnlichen Addition und der skalaren Multiplikation

$$(f + g)(x) := f(x) + g(x), \quad (\lambda f)(x) := \lambda \cdot f(x), \quad \lambda \in \mathbb{K}, x \in [0, 1],$$

wird $C([0, 1])$ zu einem $\mathbb{K}$ -Vektorraum. Wir definieren für $f, g \in C([0, 1])$ ein Skalarprodukt

$$\langle f, g \rangle := \int_0^1 f(x) \overline{g(x)} dx.$$

Aufgabe 7.2.25. Zeige, dass $\langle \cdot, \cdot \rangle$ eine positiv definite hermitesche Form definiert.

Ausgestattet mit diesem Skalarprodukt wird $C([0, 1])$ ein unitärer Raum. Die Norm einer Funktion f ist

$$\|f\|_2 = \sqrt{\int_0^1 |f(x)|^2 dx} = \left(\int_0^1 |f(x)|^2 dx \right)^{1/2}.$$

Beispiele von orthogonalen Systemen in $C([0, 1], \mathbb{K})$ (für $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$) erhält man aus den folgenden *Orthogonalitätsrelationen* für Sinus- und Kosinusfunktion:

$$\int_0^1 \sin(2\pi m x) \sin(2\pi n x) dx = \begin{cases} 0 & \text{für } n \neq m \\ \frac{1}{2} & \text{für } n = m > 0, \end{cases} \quad (7.4)$$

$$\int_0^1 \sin(2\pi m x) \cos(2\pi n x) dx = 0 \quad \text{für alle } n, m, \quad (7.5)$$

$$\int_0^1 \cos(2\pi m x) \cos(2\pi n x) dx = \begin{cases} 0 & \text{für } n \neq m \\ \frac{1}{2} & \text{für } n = m > 0 \\ 1 & \text{für } n = m = 0. \end{cases} \quad (7.6)$$

Betrachten wir die Funktionen $c_0, c_1, c_2, \dots, s_1, s_2, \dots$, die für $x \in [0, 1]$ durch

$$c_k(x) := \cos(2\pi k x), \quad s_k(x) := \sin(2\pi k x), \quad \text{für } k = 0, 1, 2, \dots$$

definiert sind, dann bilden nach den obigen Orthogonalitätsrelationen diese Funktionen ein Orthogonalsystem in $C([0, 1])$. Durch Normierung erhalten wir ein Orthonormalsystem von Funktionen.

Aufgabe 7.2.26. Beweisen Sie die Orthogonalitätsrelationen 7.4, 7.5 und 7.6.

Wir betrachten nun endlich viele von diesen Funktionen, z.B.

$$c_0, c_1, \dots, c_n, \quad s_1, \dots, s_n.$$

Diese bilden eine Orthogonalbasis eines linearen Teilraums $\mathcal{T}_n$ von $C([0, 1])$. Der Raum $\mathcal{T}_n$ besteht aus allen Linearkombinationen der Form

$$T = \alpha_0 c_0 + \sum_{k=1}^n (\alpha_k c_k + \beta_k s_k), \quad \text{d.h.}$$

$$T(x) = \alpha_0 + \sum_{k=1}^n \alpha_k \cos(2\pi kx) + \beta_k \sin(2\pi kx) .$$

Solche Funktionen heißen *trigonometrische Polynome* vom Grad $\leq n$.

Wir bezeichnen mit π_n die Orthogonalprojektion von $C([0, 1])$ auf den linearen Teilraum $\mathcal{T}_n$. Für jede stetige Funktion $f \in C([0, 1])$ ist die Projektion $\pi_n(f)$ ein trigonometrisches Polynom vom Grad $\leq n$; nach Satz 7.2.16(vii) ist $\pi_n(f)$ die beste Approximation von f durch ein trigonometrisches Polynom vom Grad $\leq n$ bzgl. der L^2 -Norm. Nach der allgemeinen Formel aus Satz 7.2.16 ist $\pi_n(f)$ gegeben durch

$$\pi_n(f) = \langle f, \tilde{c}_0 \rangle \tilde{c}_0 + \sum_{k=1}^n (\langle f, \tilde{c}_k \rangle \tilde{c}_k + \langle f, \tilde{s}_k \rangle \tilde{s}_k) = \frac{\langle f, c_0 \rangle}{\|c_0\|^2} + \dots,$$

wobei $\tilde{c}_0, \tilde{c}_1, \tilde{c}_2, \dots, \tilde{s}_1, \tilde{s}_2, \dots$ durch Normierung aus $c_0, c_1, c_2, \dots, s_1, s_2, \dots$ hervorgehen. Dies bedeutet explizit:

$$\pi_n(f)(x) = \frac{\alpha_0}{2} + \sum_{k=1}^n \alpha_k \cos(2\pi kx) + \beta_k \sin(2\pi kx)$$

mit

$$\alpha_k = 2 \int_0^1 f(x) \cos(2\pi kx) dx, \quad k = 0, 1, 2, \dots$$

$$\beta_k = 2 \int_0^1 f(x) \sin(2\pi kx) dx, \quad k = 1, 2, 3, \dots$$

Die Koeffizienten α_k and β_k heißen *Fourier-Koeffizienten* von f . Die unendliche Reihe

$$\frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \alpha_k \cos(2\pi kx) + \beta_k \sin(2\pi kx)$$

heißt *Fourierreihe* der Funktion f . Leider ist es nicht richtig, dass für jede stetige Funktion f die Fourierreihe von f für jedes x gegen $f(x)$ konvergiert. Ist allerdings die Funktion f zweimal stetig differenzierbar und $f(0) = f(1)$, dann konvergiert die Fourierreihe von f gleichmäßig gegen f . Das Teilgebiet der Mathematik, in welchem derartige Fragen untersucht werden, heißt *harmonische Analysis*.

Für $\mathbb{K} = \mathbb{C}$ erhalten wir ein weiteres Orthonormalsystem durch die Funktionen

$$e_n(x) := e^{2\pi i n x}, \quad n \in \mathbb{Z},$$

und entsprechende Fourierreihen

$$\sum_{k \in \mathbb{Z}} \gamma_k e_k(x), \quad \gamma_k = \langle f, e_k \rangle = \int_0^1 f(x) e^{-2\pi i k x} dx.$$

Mittels der Eulerschen Formel

$$e_k(x) = c_k(x) + i s_k(x), \quad e^{2\pi i k x} = \cos(2\pi kx) + i \sin(2\pi kx), \quad x \in [0, 1],$$

lassen sich beide Darstellungen ineinander umrechnen.

Aufgaben zu Abschnitt 7.2

Aufgabe 7.2.27. Zeige: Für eine $\mathbb{K}$ -lineare Abbildung $\varphi: V \rightarrow W$ zwischen unitären $\mathbb{K}$ -Vektorräumen sind äquivalent:

- (a) $\langle \varphi(x), \varphi(y) \rangle = \langle x, y \rangle$ für alle $x, y \in V$.
- (b) $\|\varphi(x)\| = \|x\|$ für alle $x \in V$.
- (c) $\|\varphi(x) - \varphi(y)\| = \|x - y\|$ für alle $x, y \in V$.

Hinweis: Die Polarisierungsidentität 7.1.6.

Aufgabe 7.2.28. Für Teilmengen eines unitären Raums V hat die Zuordnung $M \mapsto M^\perp$ folgende Eigenschaften:

1. $A \subseteq B \Rightarrow B^\perp \subseteq A^\perp$.
2. $A \subseteq B^\perp \iff B \subseteq A^\perp$.
3. $A \subseteq (A^\perp)^\perp$.
4. $A^\perp = ((A^\perp)^\perp)^\perp$.
5. Ist $\text{span}(A)$ endlichdimensional, so gilt $(A^\perp)^\perp = \text{span}(A)$.

Aufgabe 7.2.29. Sei $U \subseteq V$ ein endlichdimensionaler Unterraum des unitären Raums V und $\pi: V \rightarrow V$ die Orthogonalprojektion auf U (Satz 7.2.16). Zeige, dass $\pi' := \text{id}_V - \pi: V \rightarrow V$ die orthogonale Projektion auf $U^\perp$ ist. Vergleiche mit dem Gram–Schmidt-Verfahren.

Aufgabe 7.2.30. Wir betrachten den euklidischen Raum $V = \mathbb{R}^{(\mathbb{N})}$ (alle reellen Folgen, die nur endlich viele von 0 verschiedene Glieder haben). Wir wissen schon, dass die Folgen $e_k = (\delta_{kn})_{n \in \mathbb{N}}$ in diesem Raum eine Basis bilden. Zeigen Sie:

- (a) Die Abbildung $\Gamma: \text{Hom}(V, \mathbb{R}) \rightarrow \mathbb{R}^{\mathbb{N}}$, $\Gamma(\alpha) = (\alpha(e_n))_{n \in \mathbb{N}}$ ist ein Isomorphismus von Vektorräumen.
- (b) Der Satz von Fréchet–Riesz gilt nicht für V , d.h., es existiert ein $\alpha \in \text{Hom}(V, \mathbb{R})$, für das kein $w \in V$ existiert mit $\alpha(v) = \langle v, w \rangle$ für alle $v \in V$.

7.3 Orthogonale und unitäre Abbildungen

In diesem Abschnitt sei V ein unitärer $\mathbb{K}$ -Vektorraum. Wir wollen lineare Abbildungen $\varphi: V \rightarrow V$ betrachten, die das Skalarprodukt erhalten, so wie diese Abbildungen darstellende Matrizen.

Definition 7.3.1. [Orthogonale [unitäre] Abbildung] (a) Eine lineare Abbildung $\varphi: V \rightarrow V$ heißt *unitär*, wenn φ bijektiv ist und mit dem Skalarprodukt verträglich, d.h.,

$$\langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle \quad \text{gilt für alle } v, w \in V.$$

Für $\mathbb{K} = \mathbb{R}$ nennt man unitäre Abbildungen auch *orthogonal*.

- (b) Eine lineare Abbildung $\varphi: V \rightarrow V$ heißt *Isometrie*, wenn

$$\|\varphi(v)\| = \|v\| \quad \text{gilt für alle } v \in V.$$

Bemerkung 7.3.2. (a) Unitäre Abbildungen φ sind Isometrien und erhalten Orthogonalität, d.h.

$$\|\varphi(v)\| = \|v\|, \quad v \perp w \Rightarrow \varphi(v) \perp \varphi(w).$$

In der Tat gilt $\|\varphi(v)\| = \sqrt{\langle \varphi(v), \varphi(v) \rangle} = \sqrt{\langle v, v \rangle} = \|v\|$, also wird die Länge erhalten. Und aus $\langle v, w \rangle = 0$ folgt $\langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle = 0$. Damit wird die Orthogonalität erhalten.

Allgemeiner erhalten unitäre Abbildungen Winkel, da sich diese durch das Skalarprodukt ausdrücken lassen.

(b) Ist $\varphi: V \rightarrow V$ eine Isometrie, so folgt

$$\langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle \quad \text{für alle } v, w \in V$$

aus der Polarisierungsformel (Aufgabe 7.1.1), die man auf die hermitesche Form

$$(v, w) \mapsto \langle \varphi(v), \varphi(w) \rangle$$

anwendet.

Diese Beobachtung zeigt, dass unitäre Abbildungen genau die surjektiven Isometrien sind.

Beispiele 7.3.3. (orthogonaler Abbildungen)

- im euklidischen $\mathbb{R}^2$: Drehungen um den Ursprung und Spiegelungen an Geraden durch den Ursprung (wir werden das später beweisen);
- im euklidischen $\mathbb{R}^3$: Drehungen um eine Achse durch den Ursprung, Spiegelungen an Ebenen durch den Ursprung und Spiegelungen an Geraden durch den Ursprung.

Achtung: orthogonale Projektionen sind keine orthogonalen Abbildungen im obigen Sinn, da sie weder die Länge von Vektoren noch Orthogonalität erhalten.

Bemerkung 7.3.4. (a) Für endlichdimensionale unitäre Räume V folgt bekanntlich aus der Injektivität einer linearen Abbildung $\varphi: V \rightarrow V$ schon deren Surjektivität. Ist φ eine Isometrie, so haben Vektoren $0 \neq v$ ein nichtverschwindendes Bild, d.h. $\ker(\varphi) = \{0\}$. Also ist φ injektiv und somit auch surjektiv, d.h. unitär.

(b) Für unendlichdimensionale euklidische Vektorräume gibt es lineare Isometrien, die nicht surjektiv sind. Im Raum ℓ^2 (siehe Beispiel 7.2.22) ist der Shiftoperator

$$(x_1, x_2, x_3, \dots) \mapsto (0, x_1, x_2, x_3, \dots)$$

hierfür ein Beispiel.

Definition 7.3.5. [Die orthogonale und die unitäre Gruppe] Wir schreiben

$$U(V) := \{\varphi \in GL(V) : (\forall v, w \in V) \langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle\}$$

für die Menge aller unitären Abbildungen eines unitären Vektorraums V , die sogenannte *unitäre Gruppe von V* (vgl. Satz 7.3.6). Für $\mathbb{K} = \mathbb{R}$ heißt die Menge

$$O(V) := U(V) = \{\varphi \in GL(V) : (\forall v, w \in V) \langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle\}$$

der orthogonalen Abbildungen die *orthogonale Gruppe des euklidischen Raums V* .

Satz 7.3.6. Für einen unitären Raum V ist $U(V)$ eine Untergruppe von $GL(V)$. Für einen euklidischen Raum ist insbesondere $O(V)$ eine Untergruppe von $GL(V)$.

Beweis. (U0) $\text{id}_V \in U(V)$, da die identische Abbildung offensichtlich orthogonal ist.

(U1) $\varphi, \psi \in U(V) \Rightarrow \varphi \circ \psi \in U(V)$: Erhalten φ und ψ das Skalarprodukt, so folgt für alle $v, w \in V$:

$$\langle (\varphi \circ \psi)(v), (\varphi \circ \psi)(w) \rangle = \langle \varphi(\psi(v)), \varphi(\psi(w)) \rangle = \langle \psi(v), \psi(w) \rangle = \langle v, w \rangle.$$

(U2) $\varphi \in U(V) \Rightarrow \varphi^{-1} \in U(V)$: Eine orthogonale Abbildung φ ist definitionsgemäß bijektiv; sie hat also eine Inverse φ^{-1} und es gilt $v = \varphi(\varphi^{-1}(v))$ für alle $v \in V$. Also folgt

$$\langle v, w \rangle = \langle \varphi(\varphi^{-1}(v)), \varphi(\varphi^{-1}(w)) \rangle = \langle \varphi^{-1}(v), \varphi^{-1}(w) \rangle$$

für alle v, w . Damit ist φ^{-1} eine orthogonale Abbildung. □

Satz 7.3.7. [Orthogonalitätskriterium] Sei $v_1, v_2, \dots, v_n$ eine Orthonormalbasis der unitären $\mathbb{K}$ -Vektorraums V . Dann ist eine lineare Abbildung $\varphi: V \rightarrow V$ genau dann unitär, wenn $\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n)$ ebenfalls eine Orthonormalbasis von V ist.

Beweis. Da $v_1, v_2, \dots, v_n$ eine Orthonormalbasis ist, gilt $\langle v_i, v_j \rangle = \delta_{ij}$ für alle i, j . Ist φ unitär, so folgt aus der Definition $\langle \varphi(v_i), \varphi(v_j) \rangle = \langle v_i, v_j \rangle = \delta_{ij}$. Damit ist $\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n)$ ebenfalls eine Orthonormalbasis.

Umgekehrt sei $\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n)$ eine Orthonormalbasis, also

$$\langle \varphi(v_i), \varphi(v_j) \rangle = \delta_{ij}.$$

Wir zeigen, dass die Abbildung φ orthogonal [unitär] ist: Für beliebiges $v = \sum_{i=1}^n \alpha_i v_i$ und $w = \sum_{j=1}^n \beta_j v_j$ gilt nach Bemerkung 7.2.4:

$$\begin{aligned} \langle \varphi(v), \varphi(w) \rangle &= \langle \varphi(\sum_{i=1}^n \alpha_i v_i), \varphi(\sum_{j=1}^n \beta_j v_j) \rangle = \langle \sum_{i=1}^n \alpha_i \varphi(v_i), \sum_{j=1}^n \beta_j \varphi(v_j) \rangle \\ &= \sum_{i,j} \alpha_i \overline{\beta_j} \langle \varphi(v_i), \varphi(v_j) \rangle = \sum_{i,j} \alpha_i \overline{\beta_j} \delta_{ij} = \sum_{i=1}^n \alpha_i \overline{\beta_i} = \langle v, w \rangle. \end{aligned} \quad \square$$

Nachdem wir unitäre und orthogonale lineare Abbildungen im Abstrakten diskutiert haben, wenden wir uns nun den entsprechenden Begriffen für Matrizen zu. Hierzu definieren wir für $A \in M_n(\mathbb{C})$ den Begriff der adjungierten Matrix A^* , die für Matrizen mit reellen Einträgen gleich der Transponierten $A^\top$ ist und ähnliche Eigenschaften besitzt. Man beachte aber, dass dieser Begriff einer Adjungierten nichts mit dem Begriff der Adjungierten gemein hat, wie er für die Cramersche Regel definiert wurde.

Definition 7.3.8. [Die Adjungierte] Für $A = (a_{ij}) \in M_{m,n}(\mathbb{C}) = \mathbb{C}^{(m,n)}$ definieren wir die (komplex) konjugierte Matrix

$$\overline{A} = (\overline{a_{ij}})$$

durch Konjugation der einzelnen Matrixeinträge und die adjungierte Matrix durch

$$A^* := A^\dagger := \overline{A}^\top = \overline{A^\top} \in M_{n,m}(\mathbb{C}) = \mathbb{C}^{(n,m)}.$$

Beispiel 7.3.9.

$$A = \begin{pmatrix} 1+i & 1 \\ i & 2-i \end{pmatrix}, \quad \overline{A} = \begin{pmatrix} 1-i & 1 \\ -i & 2+i \end{pmatrix}, \quad A^* = \begin{pmatrix} 1-i & -i \\ 1 & 2+i \end{pmatrix}.$$

Bemerkung 7.3.10. Für $A, B \in M_n(\mathbb{C})$ und $\lambda \in \mathbb{C}$ gelten folgende Regeln:

$$\begin{aligned} (A+B)^* &= A^* + B^*, & (\lambda A)^* &= \overline{\lambda} A^* \quad (\text{für } \lambda \in \mathbb{C}), \\ (AB)^* &= B^* A^*, & A^{**} &= A. \end{aligned}$$

Die beiden ersten und die vierte Regel sind offensichtlich; wir berechnen zum Nachweis der dritten:

$$(AB)^* = \overline{AB}^\top = (\overline{A} \overline{B})^\top = \overline{B}^\top \overline{A}^\top = B^* A^*.$$

Hierbei verwenden wir $\overline{\overline{A}} = A$ und die bekannte Relation $(AB)^\top = B^\top A^\top$. Die Einträge a_{ij} von A sind genau dann reell, wenn $\overline{A} = A$ gilt. In diesem Fall ist $A^* = A^\top$.

Lemma 7.3.11. Für $A \in M_{m,n}(\mathbb{K})$ gilt $\langle Av, w \rangle = \langle v, A^* w \rangle$ für $v \in \mathbb{K}^n, w \in \mathbb{K}^m$.

Beweis. $\langle Av, w \rangle = (Av)^\top \overline{w} = v^\top A^\top \overline{w} = v^\top \overline{A^* w} = \langle v, A^* w \rangle.$ $\square$

Bemerkung 7.3.12. (1) Aus dem oben Gesagten

$$\langle w, Av \rangle = \overline{\langle Av, w \rangle} = \overline{\langle v, A^*w \rangle} = \langle A^*w, v \rangle$$

für $v \in \mathbb{K}^n$ und $w \in \mathbb{K}^m$.

(2) Ist $A \in M_{m,n}(\mathbb{R})$ eine *reelle* Matrix, so gilt $A^* = A^\top$ und damit:

$$\langle Av, w \rangle = \langle v, A^\top w \rangle, \quad \langle A^\top w, v \rangle = \langle w, Av \rangle \quad \text{für alle } v \in \mathbb{R}^n, w \in \mathbb{R}^m.$$

Für Anwendungen auf lineare Gleichungssysteme, liefert der folgende Satz ein wichtiges Lösbarkeitskriterium:

$$Ax = y$$

ist genau dann lösbar, wenn $y \perp \ker(A^*)$ ist. Dieses Kriterium ist besonders effektiv, da man oft Kerne leichter berechnen kann als Bilder.

Satz 7.3.13. Für $A \in M_{m,n}(\mathbb{K})$ gilt

$$\operatorname{im}(A) = \ker(A^*)^\perp, \quad \ker(A) = \operatorname{im}(A^*)^\perp, \quad \operatorname{im}(A^*) = \ker(A)^\perp \quad \text{und} \quad \ker(A^*) = \operatorname{im}(A)^\perp.$$

Beweis. Aus $\langle Av, w \rangle = \langle v, A^*w \rangle$ für $v \in \mathbb{K}^n, w \in \mathbb{K}^m$ folgt, dass $w \in \operatorname{im}(A)^\perp$ äquivalent ist zu $A^*w \in (\mathbb{K}^n)^\perp = \{0\}$, also $w \in \ker(A^*)$. Das ist die vierte Relation.

Die erste erhalten wir sofort aus

$$\operatorname{im}(A) = (\operatorname{im}(A)^\perp)^\perp = \ker(A^*)^\perp.$$

Die zweite und dritte Relation erhalten wir, indem wir die ersten beiden auf die Matrix A^* anwenden. $\square$

Definition 7.3.14. Wir nennen $Q \in M_n(\mathbb{C})$ eine *unitäre* Matrix, wenn

$$QQ^* = Q^*Q = \mathbf{1}_n$$

gilt, d.h. Q ist invertierbar mit $Q^{-1} = Q^*$. Entsprechend heißt $Q \in M_n(\mathbb{R})$ *orthogonal*, wenn

$$QQ^\top = Q^\top Q = \mathbf{1}_n$$

gilt, d.h. Q ist invertierbar mit $Q^{-1} = Q^\top$. Eine reelle Matrix Q ist also genau dann orthogonal, wenn sie, aufgefasst als komplexe Matrix, unitär ist.

Satz 7.3.15. [Charakterisierung unitärer Matrizen] Für $Q \in M_n(\mathbb{K})$, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ sind die folgenden Eigenschaften äquivalent:

- (i) Die lineare Abbildung $L_Q: \mathbb{K}^n \rightarrow \mathbb{K}^n, x \mapsto Qx$ ist unitär bzgl. des Standardskalarprodukts, d.h. es gilt:

$$\langle Qv, Qw \rangle = \langle v, w \rangle \quad \text{für alle } v, w.$$

- (ii) Die Spalten $s_1, \dots, s_n$ von Q bilden ein Orthonormalsystem, d.h.

$$\langle s_i, s_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \end{cases}$$

- (iii) $Q^*Q = \mathbf{1}_n$.

- (iv) Q ist invertierbar und es gilt $Q^{-1} = Q^*$, d.h. Q ist unitär.

- (v) Die Zeilen von Q bilden ein Orthonormalsystem.

- (vi) $QQ^* = \mathbf{1}_n$.

Beweis. (i) $\Leftrightarrow$ (ii): Betrachte die Standardbasis $e_1, e_2, \dots, e_n$ in $\mathbb{K}^n$, die eine Orthonormalbasis ist. Wir wissen aus Satz 7.3.7, dass L_Q genau dann unitär ist, wenn die Vektoren $L_Q(e_1) = Qe_1, L_Q(e_2) = Qe_2, \dots, L_Q(e_n) = Qe_n$, also die Spalten von Q , ebenfalls eine Orthonormalbasis bilden.

(ii) $\Leftrightarrow$ (iii) folgt aus $Q^*Q = (s_i^* s_j) = (\langle s_j, s_i \rangle)$. Es gilt also genau dann $Q^*Q = \mathbf{1}_n = (\delta_{ij})$, wenn die Spalten von Q ein Orthonormalsystem bilden.

(iii) $\Leftrightarrow$ (iv): Es ist klar, dass (iii) aus (iv) folgt. Ist $Q^*Q = \mathbf{1}_n$, so ist $\ker(Q) = \{0\}$, also Q invertierbar (nach der Dimensionsformel für lineare Abbildungen) und daher $Q^{-1} = Q^*$.

(v) $\Leftrightarrow$ (vi) zeigt man analog zu (ii) $\Leftrightarrow$ (iii).

(iv) $\Leftrightarrow$ (vi) zeigt man analog zu (iii) $\Leftrightarrow$ (iv). $\square$

Beispiel 7.3.16. Die Matrix

$$Q = \frac{1}{3} \begin{pmatrix} 1 & -2 & 2 \\ 2 & 2 & 1 \\ 2 & -1 & -2 \end{pmatrix} \in M_3(\mathbb{R})$$

ist orthogonal, da ihre Spalten eine Orthonormalbasis bilden. Die inverse Matrix ist gegeben durch

$$Q^{-1} = Q^T = \frac{1}{3} \begin{pmatrix} 1 & 2 & 2 \\ -2 & 2 & -1 \\ 2 & 1 & -2 \end{pmatrix}.$$

Folgerung 7.3.17. Ein Endomorphismus φ eines endlichdimensionalen unitären Vektorraumes V ist genau dann unitär, wenn seine Matrix Q bzgl. einer beliebigen Orthonormalbasis $v_1, v_2, \dots, v_n$ eine unitäre Matrix ist.

Beweis. Dies folgt aus dem Orthogonalitätskriterium Satz 7.3.7, der Charakterisierung unitärer Matrizen, und aus der Tatsache, dass die Spalten der Matrix Q von φ bzgl. der Basis $B = (v_1, \dots, v_n)$ die Koordinatenvektoren $[\varphi(v_1)]_B, \dots, [\varphi(v_n)]_B$ sind sowie der Relation

$$\langle x, y \rangle_V = \langle [x]_B, [y]_B \rangle_{\mathbb{K}^n}. \quad \square$$

Definition 7.3.18. [Die orthogonalen und unitären Matrixgruppen] Wir schreiben

$$O_n(\mathbb{R}) := \{Q \in GL_n(\mathbb{R}) : Q^T = Q^{-1}\}$$

für die Menge der orthogonalen reellen Matrizen und

$$U_n(\mathbb{C}) := \{Q \in GL_n(\mathbb{C}) : Q^* = Q^{-1}\}$$

für die Menge der unitären komplexen Matrizen. Beachte, dass die orthogonalen reellen Matrizen genau diejenigen unitären Matrizen sind, bei denen alle Einträge reell sind:

$$O_n(\mathbb{R}) = U_n(\mathbb{C}) \cap GL_n(\mathbb{R}),$$

aber

$$O_n(\mathbb{C}) := \{Q \in GL_n(\mathbb{C}) : Q^T = Q^{-1}\} \neq U_n(\mathbb{C})$$

gilt.

Wir werden sogleich zeigen, dass $O_n(\mathbb{R})$ und $U_n(\mathbb{C})$ Untergruppen von $GL_n(\mathbb{K})$ sind. Da

$$SL_n(\mathbb{K}) := \{g \in GL_n(\mathbb{K}) : \det g = 1\}$$

für jeden Körper eine Untergruppe von $GL_n(\mathbb{K})$ ist (Nachweis als Übung), erhalten wir weitere Untergruppen

$$SO_n(\mathbb{R}) := O_n(\mathbb{R}) \cap SL_n(\mathbb{R}) \quad \text{und} \quad SU_n(\mathbb{C}) := U_n(\mathbb{C}) \cap SL_n(\mathbb{C}).$$

Diese Untergruppen heißen die *spezielle orthogonale* und die *spezielle unitäre Gruppe*.

Lemma 7.3.19. $O_n(\mathbb{R})$ ist eine Untergruppe von $GL_n(\mathbb{R})$ und $U_n(\mathbb{C})$ ist eine Untergruppe von $GL_n(\mathbb{C})$.

Beweis. Man rechnet allgemein nach, dass $U_n(\mathbb{K}) \subseteq GL_n(\mathbb{K})$ eine Untergruppe ist:

(U0) Aus $\mathbf{1}^* = \mathbf{1}$ folgt sofort $\mathbf{1} \in U_n(\mathbb{K})$.

(U1) Aus $A^* = A^{-1}$ und $B^* = B^{-1}$ folgt

$$(AB)^* = B^* A^* = B^{-1} A^{-1} = (AB)^{-1},$$

also $AB \in U_n(\mathbb{K})$.

(U2) Aus $A^* = A^{-1}$ folgt $(A^{-1})^* = A^{**} = A = (A^{-1})^{-1}$, also $A^{-1} \in U_n(\mathbb{K})$.

Mit $O_n(\mathbb{R}) = U_n(\mathbb{R})$ ergibt sich damit die Behauptung im Reellen. $\square$

Satz 7.3.20. [Eigenschaften unitärer Matrizen Q] Für $Q \in U_n(\mathbb{K})$ gilt:

- (i) $|\det(Q)| = 1$. Für reelle orthogonale Matrizen bedeutet dies $\det(Q) = \pm 1$.
- (ii) Für jeden Eigenwert λ_i von Q gilt $|\lambda_i| = 1$.
(Die einzig möglichen reellen Eigenwerte sind also $\lambda_i = \pm 1$.)
- (iii) Sind v_1 und v_2 Eigenvektoren zweier verschiedener Eigenwerte λ_1 und λ_2 , dann sind v_1 und v_2 orthogonal zueinander.
- (iv) Ist U ein Q -invarianter linearer Teilraum von $\mathbb{K}^n$, d.h. gilt $Q \cdot U \subseteq U$, so ist $U^\perp$ ebenfalls Q -invariant.

Die entsprechenden Eigenschaften sind für orthogonale [unitäre] Transformationen φ eines euklidischen [unitären] Vektorraums gültig.

Beweis. (i) $\det(Q^*) = \det(\overline{Q}^\top) = \overline{\det(Q^\top)} = \overline{\det(Q)}$. Aus $Q^*Q = \mathbf{1}_n$ folgt daher

$$1 = \det(\mathbf{1}_n) = \det(Q^*Q) = \det(Q^*) \cdot \det(Q) = \overline{\det(Q)} \cdot \det(Q) = |\det(Q)|^2$$

und damit $|\det(Q)| = 1$.

(ii) Sei $\lambda \in \mathbb{C}$ ein Eigenwert von Q . Dann gibt es einen Eigenvektor $0 \neq v \in \mathbb{C}^n$ zum Eigenwert λ . Da Q unitär ist, erhalten wir

$$\|v\| = \|Qv\| = \|\lambda v\| = |\lambda| \|v\|,$$

also $|\lambda| = 1$.

(iii) Seien $\lambda_1 \neq \lambda_2$ zwei verschiedene Eigenwerte (in $\mathbb{C}$) von Q und seien v_1, v_2 zugehörige Eigenvektoren. Wir nehmen an, dass $\langle v_1, v_2 \rangle \neq 0$ ist. Dann gilt

$$\langle v_1, v_2 \rangle = \langle Qv_1, Qv_2 \rangle = \langle \lambda_1 v_1, \lambda_2 v_2 \rangle = \lambda_1 \overline{\lambda_2} \langle v_1, v_2 \rangle.$$

Wir folgern $\lambda_1 \overline{\lambda_2} = 1$, also $\lambda_1 = \overline{\lambda_2}^{-1}$. Aus $|\lambda_2| = 1$ folgt $\overline{\lambda_2}^{-1} = \lambda_2$ und damit ergibt sich $\lambda_2 = \lambda_1$, im Widerspruch zur Annahme.

(iv) Wir müssen zeigen, dass $v \in U^\perp$ die Gültigkeit von $Qv \in U^\perp$ nach sich zieht. Dazu sei $v \in U^\perp$, d.h. $\langle v, u \rangle = 0$ für alle $u \in U$. Da Q unitär ist, schließen wir $\langle Qv, Qu \rangle = 0$ für alle $u \in U$. Da Q unitär ist und da U nach Annahme Q -invariant ist, induziert Q eine unitäre Abbildung $L: U \rightarrow U, x \mapsto Qx$. Diese Abbildung ist injektiv und da U endlichdimensional ist auch surjektiv. Daher ist U auch invariant unter $Q^{-1} = Q^*$. Damit erhalten wir $\langle Qv, w \rangle = \langle v, Q^*w \rangle = 0$ für alle $w \in U$, also $Qv \in U^\perp$. $\square$

Satz 7.3.21. (Struktur orthogonaler (2×2) -Matrizen) Für $Q \in O_2(\mathbb{R})$ gilt:

- (a) Ist $\det Q = 1$, so beschreibt Q eine Drehung um den Ursprung um den Winkel $\gamma \in [0, 2\pi[$:
 (b) Ist $\det Q = -1$, so kann Q als eine Spiegelung an der x_1 -Achse mit anschließender Drehung aufgefasst werden:

$$Q = \begin{pmatrix} \cos \gamma & \sin \gamma \\ \sin \gamma & -\cos \gamma \end{pmatrix} = D(\gamma) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Dies ist eine Spiegelung an der Geraden g , die den Winkel $\frac{\gamma}{2}$ mit der x_1 -Achse einschließt.

Damit ist jede orthogonale Abbildung der euklidischen Ebene $\mathbb{R}^2$ entweder eine Drehung um den Ursprung 0 oder eine Spiegelungen an einer Geraden durch 0.

Beweis. Da die erste Spalte q_1 von Q ein Einheitsvektor ist, können wir sie für ein $\gamma \in [0, 2\pi[$ schreiben als

$$q_1 = \begin{pmatrix} \cos \gamma \\ \sin \gamma \end{pmatrix}.$$

Die zweite Spalte q_2 ist ebenfalls ein Einheitsvektor, der senkrecht auf q_1 ist, also von der Gestalt

$$q_2 = \begin{pmatrix} -\sin \gamma \\ \cos \gamma \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} \sin \gamma \\ -\cos \gamma \end{pmatrix}.$$

Die beiden Fälle unterscheiden sich darin, dass im ersten Fall $\det Q = \cos^2 \gamma + \sin^2 \gamma = 1$ ist und im zweiten Fall $\det Q = -\cos^2 \gamma - \sin^2 \gamma = -1$.

Dass Q im zweiten Fall eine Spiegelung an der Geraden g beschreibt, die den Winkel $\frac{\gamma}{2}$ mit der x_1 -Achse einschließt, folgt aus

$$Q \begin{pmatrix} \cos \frac{\gamma}{2} \\ \sin \frac{\gamma}{2} \end{pmatrix} = \begin{pmatrix} \cos \frac{\gamma}{2} \\ \sin \frac{\gamma}{2} \end{pmatrix}, \quad Q \begin{pmatrix} -\sin \frac{\gamma}{2} \\ \cos \frac{\gamma}{2} \end{pmatrix} = \begin{pmatrix} \sin \frac{\gamma}{2} \\ -\cos \frac{\gamma}{2} \end{pmatrix},$$

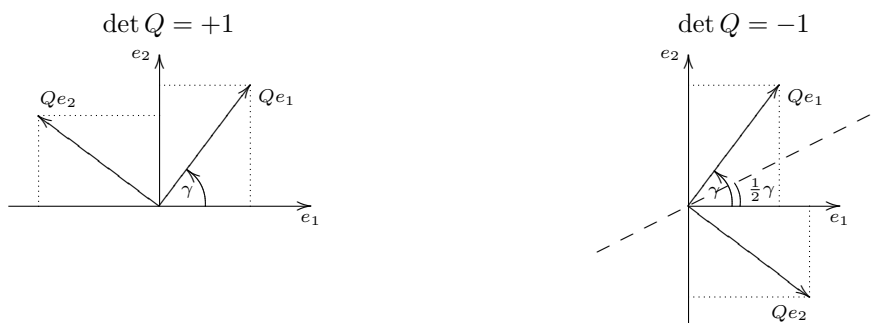
wie man leicht mittels der Relationen

$$\cos \gamma \cos \frac{\gamma}{2} + \sin \gamma \sin \frac{\gamma}{2} = \cos \left(\gamma - \frac{\gamma}{2} \right) = \cos \frac{\gamma}{2}$$

und

$$\sin \gamma \cos \frac{\gamma}{2} - \cos \gamma \sin \frac{\gamma}{2} = \sin \left(\gamma - \frac{\gamma}{2} \right) = \sin \frac{\gamma}{2}$$

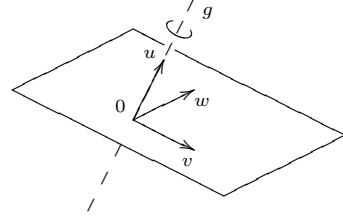
nachrechnet. □



Satz 7.3.22. (Orthogonale Abbildungen des $\mathbb{R}^3$) Sei $Q \in O_3(\mathbb{R})$ und $L_Q x := Qx$ die zugehörige orthogonale Abbildung des $\mathbb{R}^3$.

- (a) Ist $\det Q = 1$, dann ist L_Q eine Drehung um eine Achse g durch den Ursprung 0. Die Achse g geht durch einen Eigenvektor u von Q zum Eigenwert 1. Für den Winkel γ der Drehung gilt

$$\cos \gamma = \frac{1}{2}(\operatorname{tr}(Q) - 1).$$



- (b) Ist $\det Q = -1$, dann ist L_Q eine Spiegelung an einer Ebene U gefolgt von einer Drehung um die Gerade g , die orthogonal zu U ist. Die Gerade g wird erzeugt durch einen Eigenvektor u zum Eigenwert -1 .

Wählt man eine Orthogonalbasis u, v, w , wie in der Zeichnung angedeutet, so sieht die Matrix Q' von L_Q bzgl. dieser neuen Basis wie folgt aus:

$$(a) \det Q = 1: \quad Q' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{pmatrix}$$

$$(b) \det Q = -1: \quad Q' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{pmatrix}$$

Im zweiten Fall liegt eine Spiegelung vor, wenn $\gamma = 0$ ist. Es handelt sich um eine Spiegelung gefolgt von einer Drehung, wenn $\gamma \neq 0$ gilt.

Beweis. Wir bemerken zunächst, dass entweder $\det Q \in \{1, -1\}$ aus Satz 7.3.20(i) folgt.

Da Q eine reelle Matrix ist, hat das charakteristische Polynom von Q reelle Koeffizienten. Da dessen Grad 3 und demnach ungerade ist, besitzt es nach dem Zwischenwertsatz eine reelle Wurzel λ_1 . Aus Satz 7.3.20(ii) folgt $\lambda_1 \in \{\pm 1\}$.

(a) $\det Q = 1$: Wir behaupten: 1 ist Eigenwert von Q . Seien $\lambda_1, \lambda_2, \lambda_3$ die komplexen Eigenwerte von Q , wobei λ_1 ein reeller Eigenwert sei. Nach dem oben Gesagten gilt $\lambda_1 = 1$ oder $\lambda_1 = -1$. Ist λ_2 nicht reell, so ist $\lambda_3 = \bar{\lambda}_2$ ebenfalls nicht reell. In diesem Fall gilt $\det Q = 1 = \lambda_1 \lambda_2 \lambda_3 = \lambda_1 |\lambda_2|^2 = \lambda_1$, also $\lambda_1 = 1$. Ist λ_2 reell, so auch λ_3 . Wegen $\lambda_j \in \{\pm 1\}$ und $\det Q = \lambda_1 \lambda_2 \lambda_3 = 1$ muss einer der Eigenwerte 1 sein. Wir wählen nun einen normierten Eigenvektor u zum Eigenwert 1.

(b) $\det Q = -1$: Wenden wir (a) auf die Matrix $-Q$ mit $\det(-Q) = (-1)^3 \det(Q) = 1$ an, so sehen wir, dass -1 ein Eigenwert von Q ist. Die Behauptung folgt nun aus

$$-D(\gamma) = -\begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} = D(\gamma + \pi). \quad \square$$

7.3.1 Die QR-Zerlegung einer Matrix

In diesem Abschnitt lernen wir eine wichtige Zerlegung von Matrizen kennen, durch die das Lösen des zugehörigen linearen Gleichungssystems sehr effektiv möglich ist.

Satz 7.3.23. [Die QR-Zerlegung einer Matrix B] Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $B \in M_{n,m}(\mathbb{K})$ mit $\operatorname{rank}(B) = m$, d.h. die Spalten

$$b_1, b_2, \dots, b_m \in \mathbb{K}^n$$

von B sind linear unabhängig. Dann hat B eine Zerlegung

$$B = QR,$$

wobei Q eine $n \times m$ -Matrix ist, deren Spalten $v_1, v_2, \dots, v_m$ ein Orthonormalsystem bilden und R eine obere $m \times m$ -Dreiecksmatrix mit positiven Diagonaleinträgen ist.

Beweis. Wegen $\text{rank}(B) = m$ sind die Spalten $b_1, \dots, b_m$ von B linear unabhängig. Zur Berechnung dieser Zerlegung wenden wir das GRAM-SCHMIDTsche Orthogonalisierungsverfahren auf $b_1, b_2, \dots, b_m$ an. Man erhält ein Orthonormalsystem $v_1, v_2, \dots, v_m$. Sei Q die Matrix, deren Spalten die Vektoren $v_1, \dots, v_m$ bilden.

Mit GRAM-SCHMIDT erhalten wir aus $u_i = \|u_i\|v_i$ die Relationen:

$$\begin{aligned} b_1 &= \|u_1\| \cdot v_1, \\ b_2 &= \langle b_2, v_1 \rangle v_1 + \|u_2\| v_2, \\ &\vdots \\ b_m &= \langle b_m, v_1 \rangle v_1 + \langle b_m, v_2 \rangle v_2 + \dots + \|u_m\| v_m. \end{aligned}$$

Wir definieren nun die Matrix $R = (r_{ij})$ durch

$$R = \begin{pmatrix} \|u_1\| & \langle b_2, v_1 \rangle & \dots & \langle b_m, v_1 \rangle \\ 0 & \|u_2\| & \dots & \langle b_m, v_2 \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & & \vdots \\ 0 & \dots & 0 & \|u_m\| \end{pmatrix}.$$

Um die Identität $B = Q \cdot R$ einzusehen, beachten wir, dass die k -te Spalte der Matrix QR die Form

$$\sum_{j=1}^m r_{jk} v_j = \sum_{j=1}^k r_{jk} v_j = \langle b_k, v_1 \rangle v_1 + \dots + \langle b_k, v_{k-1} \rangle v_{k-1} + \|u_k\| v_k = b_k.$$

Also ist $B = QR$. □

Beispiel 7.3.24. Wir betrachten die Matrix B , wobei die Spalten b_1, b_2, b_3 die Vektoren aus Beispiel 7.2.9 sind:

$$B = \begin{pmatrix} 1 & 3 & 3 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -3 & -1 \end{pmatrix}$$

Wir berechnen nun

$\langle b_j, v_i \rangle$	b_1	b_2	b_3
v_1	2	2	2
v_2	0	4	2
v_3	0	0	2

Aus den Ergebnissen des GRAM-SCHMIDTschen Algorithmus in 7.2.9 erhalten wir

$$\begin{aligned} \begin{pmatrix} 1 & 3 & 3 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -3 & -1 \end{pmatrix} &= \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 2 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 2 \end{pmatrix} \\ B &= \underbrace{\hspace{10em}}_Q \cdot \underbrace{\hspace{10em}}_R \end{aligned}$$

Sei B eine invertierbare $n \times n$ -Matrix mit reellen [komplexen] Koeffizienten. Dann sind die n Spalten von B linear unabhängig und wir können die QR -Zerlegung von B wie in Satz 7.3.23 konstruieren und erhalten

$$B = QR$$

Dabei ist Q eine unitäre Matrix und R eine obere Dreiecksmatrix mit positiven reellen Diagonaleinträgen.

Bemerkung 7.3.25. Anwendung auf das Lösen linearer Gleichungssysteme

$$Bx = b$$

im Falle, dass B eine invertierbare $n \times n$ -Matrix mit reellen [komplexen] Einträgen ist: Ersetzt man die Matrix B durch ihre QR -Zerlegung, so erhält man

$$QRx = b$$

und wegen $Q^{-1} = Q^*$ können wir das System in der Form

$$Rx = Q^*b$$

schreiben. Das letzte Gleichungssystem kann leicht gelöst werden, da R obere Dreiecksgestalt besitzt.

Aufgaben zu Abschnitt 7.3

Aufgabe 7.3.26. Sei V ein euklidischer Vektorraum. Zeige: Ist $v \in V$ ein Einheitsvektor, so wird durch

$$s_v: V \rightarrow V, \quad w \mapsto w - 2\langle w, v \rangle v$$

eine orthogonale Abbildung definiert. Diese Abbildung heißt *orthogonale Spiegelung an der Hyperebene orthogonal zu v* . Warum? Welche Punkte werden durch s_v fest gelassen?

Aufgabe 7.3.27. Sei V ein endlichdimensionaler euklidischer Vektorraum. Zeige: Jede orthogonale Abbildung ist ein Produkt von Spiegelungen an Hyperebenen, d.h., zu jedem $\varphi \in O(V)$ existieren $v_1, \dots, v_n \in V$ mit $\varphi = s_{v_1} s_{v_2} \cdots s_{v_n}$. Hinweis: Induktion nach $\dim V$: Hat φ einen Eigenvektor v zum Eigenwert 1 (Fixvektor), so betrachte man $\varphi|_{V^\perp}$. Ist dies nicht der Fall, so wähle man $0 \neq v \in V$ und suche eine Spiegelung s_{v_1} mit $s_{v_1}(\varphi(v)) = v$; dann hat die neue Abbildung $s_{v_1} \circ \varphi$ einen Fixvektor $\neq 0$.

Aufgabe 7.3.28. Zeige, dass die Zerlegung einer invertierbaren Matrix $B \in GL_n(\mathbb{C})$ in ein Produkt aus einer unitären Matrix Q und einer oberen Dreiecksmatrix R mit positiven reellen Diagonaleinträgen eindeutig ist.

Aufgabe 7.3.29. Zeigen Sie: Auf dem Raum $M_n(\mathbb{K})$ wird durch

$$\langle A, B \rangle := \operatorname{tr}(AB^*) = \sum_{j,k=1}^n a_{jk} \overline{b_{jk}}$$

eine positiv definite hermitesche Form definiert. Die zugehörige Norm

$$\|A\|_2 := \operatorname{tr}(AA^*)^{1/2} = \left(\sum_{j,k=1}^n |a_{jk}|^2 \right)^{1/2}$$

heißt *Hilbert-Schmidt Norm* der Matrix. Es gilt

$$\|A^*\|_2 = \|A\| \quad \text{sowie} \quad \langle A^*, B^* \rangle = \langle B, A \rangle \quad \text{für} \quad A, B \in M_n(\mathbb{K}).$$

7.4 Normale Matrizen

In diesem Abschnitt werden wir die komplexen Matrizen A beschreiben, die mittels einer unitären Matrix Q diagonalisiert werden können. Wir wollen also die komplexen Matrizen A charakterisieren, für die eine unitäre Matrix Q existiert, so dass

$$A' := Q^{-1}AQ \text{ eine Diagonalmatrix ist,}$$

d.h., es existiert eine Orthonormalbasis aus Eigenvektoren.

In diesem Kapitel seien $A, B, \dots$ immer $n \times n$ -Matrizen mit komplexen oder reellen Einträgen und $\mathbb{R}^n$ und $\mathbb{C}^n$ seien immer mit dem Standardskalarprodukt versehen.

Bemerkung 7.4.1. Für eine komplexe $n \times n$ -Matrix A definieren wir:

$$\begin{aligned} A \text{ ist symmetrisch} &: \iff A = A^\top \\ A \text{ ist schiefsymmetrisch} &: \iff A = -A^\top \\ A \text{ ist hermitesch} &: \iff A = A^* \\ A \text{ ist schiefhermitesch} &: \iff A = -A^* \end{aligned}$$

Hermitesche Matrizen heißen auch *selbstadjungiert*.

Bemerkung 7.4.2. Man beachte, dass die reellen (schief-)hermiteschen Matrizen genau die (schief-)symmetrischen sind. Einige Beispiele für obige Klassen:

symmetrisch	schiefsymmetrisch	hermitesch	schiefhermitesch
$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 2 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -i & 1-i \\ i & 2 & 2+i \\ 1+i & 2-i & -2 \end{pmatrix}$	$\begin{pmatrix} i & i & 1-i \\ i & 2i & 2+i \\ -1-i & -2+i & -2i \end{pmatrix}$

Man beachte weiterhin, dass jede Matrix A dargestellt werden kann als

- a) Summe einer reellen und einer rein imaginären Matrix:

$$A = \frac{1}{2}(A + \overline{A}) + \frac{1}{2}(A - \overline{A}).$$

- b) Summe einer symmetrischen und einer schiefsymmetrischen Matrix:

$$A = \frac{1}{2}(A + A^\top) + \frac{1}{2}(A - A^\top).$$

- c) Summe einer hermiteschen und einer schiefhermiteschen Matrix:

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*).$$

Wir kommen nun zum zentralen Begriff dieses Kapitels:

Definition 7.4.3. [Normale Matrizen] Eine Matrix $A \in M_n(\mathbb{C})$ heißt *normal*, wenn sie mit ihrer Adjungierten vertauscht, d.h.

$$\boxed{AA^* = A^*A.}$$

Lemma 7.4.4. Eine Matrix $A \in M_n(\mathbb{C})$ ist genau dann normal, wenn $\|Az\| = \|A^*z\|$ für alle $z \in \mathbb{C}^n$ gilt.

Beweis. Für $z \in \mathbb{C}^n$ ist

$$\|Az\|^2 = \langle Az, Az \rangle = \langle A^*Az, z \rangle \quad \text{und} \quad \|A^*z\|^2 = \langle A^*z, A^*z \rangle = \langle AA^*z, z \rangle.$$

Ist A normal, so folgt daraus sofort $\|Az\| = \|A^*z\|$. Ist dies umgekehrt für alle $z \in \mathbb{C}^n$ der Fall, so erhalten wir aus der Polarisierungsformel (Aufgabe 7.1.1)

$$\langle A^*Az, w \rangle = \langle AA^*z, w \rangle \quad \text{für alle} \quad z, w \in \mathbb{C}^n$$

und damit erhalten wir $A^*A = AA^*$, wenn wir Basisvektoren für z, w einsetzen. $\square$

Beispiele 7.4.5. (a) $A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$ ist nicht normal, da $A^*A = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} = AA^*$ gilt.

(b) Diagonalmatrizen sind normal: Ist A diagonal, so auch A^* , und Diagonalmatrizen vertauschen miteinander.

(c) Unitäre und reelle orthogonale Matrizen Q sind normal: Ist Q unitär, so folgt $Q^*Q = \mathbf{1}_n = QQ^*$ aus der Definition.

(d) Komplexe hermitesche und schiefhermitesche Matrizen sind normal:

Ist $A^* = \varepsilon A$ mit $\varepsilon \in \{\pm 1\}$, so ist $A^*A = \varepsilon AA = \varepsilon A^2 = A \cdot (\varepsilon A) = AA^*$.

(e) Reelle symmetrische und reelle schiefsymmetrische Matrizen sind normal: Das folgt sofort aus (d).

Diese Beispiele zeigen, dass es viele Typen normaler Matrizen gibt.

Lemma 7.4.6. *Ist A normal, so gilt:*

(a) $A - \lambda \mathbf{1}$ ist normal für jedes $\lambda \in \mathbb{C}$,

(b) Q^*AQ ist normal für jede unitäre Matrix Q .

Beweis. (a) folgt aus folgender Rechnung:

$$\begin{aligned} (A - \lambda \mathbf{1})(A - \lambda \mathbf{1})^* &= (A - \lambda \mathbf{1})(A^* - \bar{\lambda} \mathbf{1}) = AA^* - \lambda A^* - \bar{\lambda} A + \lambda \bar{\lambda} \mathbf{1}, \\ (A - \lambda \mathbf{1})^*(A - \lambda \mathbf{1}) &= (A^* - \bar{\lambda} \mathbf{1})(A - \lambda \mathbf{1}) = A^*A - \bar{\lambda} A - \lambda A^* + \bar{\lambda} \lambda \mathbf{1}. \end{aligned}$$

Beide Ausdrücke sind gleich, da $AA^* = A^*A$ gilt (A ist normal).

(b) folgt aus

$$\begin{aligned} (Q^*AQ)(Q^*AQ)^* &= (Q^*AQ)(\underbrace{Q^*A^*Q^{**}}_1) = Q^*AA^*Q, \\ (Q^*AQ)^*(Q^*AQ) &= (Q^*A^*Q^{**})(Q^*AQ) = Q^*A^* \underbrace{QQ^*}_1 AQ = Q^*A^*AQ. \end{aligned}$$

Wegen $AA^* = A^*A$ sind die Matrizen Q^*AA^*Q und Q^*A^*AQ gleich. $\square$

Lemma 7.4.7. *Sei A eine normale Matrix. Ist $\lambda \in \mathbb{C}$ ein Eigenwert von A und $v \in \mathbb{C}^n$ ein Eigenvektor von A bzgl. λ , dann ist $\bar{\lambda}$ ein Eigenwert von A^* und v ist ein Eigenvektor von A^* bzgl. $\bar{\lambda}$.*

Beweis. Da A normal ist, ist auch $A - \lambda \mathbf{1}$ normal (Lemma 7.4.6(i)) und wir erhalten mit Lemma 7.4.4 für jeden Eigenvektor v zum Eigenwert λ

$$0 = \|(A - \lambda \mathbf{1})v\| = \|(A - \lambda \mathbf{1})^*v\| = \|(A^* - \bar{\lambda} \mathbf{1})v\|,$$

also $A^*v = \bar{\lambda}v$. $\square$

Satz 7.4.8. *Ist A eine normale Matrix, so sind die Eigenvektoren von A zu verschiedenen Eigenwerten orthogonal.*

Beweis. Wir nehmen an, v und w seien Eigenvektoren zu verschiedenen Eigenwerten λ und μ , d.h. es gilt $Av = \lambda v$ und $Aw = \mu w$. Dann folgt

$$\lambda \langle v, w \rangle = \langle \lambda v, w \rangle = \langle Av, w \rangle \stackrel{(7.3.11)}{=} \langle v, A^* w \rangle \stackrel{(7.4.7)}{=} \langle v, \bar{\mu} w \rangle = \mu \langle v, w \rangle.$$

Aus $\mu \neq \lambda$ ergibt sich, dass $\langle v, w \rangle$ verschwindet. Also sind v und w orthogonal. $\square$

Lemma 7.4.9. *Ist A eine normale Matrix und $v \in \mathbb{C}^n$ ein Eigenvektor von A , dann ist der $(n-1)$ -dimensionale lineare Teilraum $v^\perp$ invariant unter A .*

Beweis. Sei $\lambda \in \mathbb{C}$ der Eigenwert, für den v ein Eigenvektor ist. Sei $w \in v^\perp$, d.h. $\langle v, w \rangle = 0$. Wir wollen $\langle v, Aw \rangle = 0$, d.h. $Aw \in v^\perp$ zeigen:

$$\langle v, Aw \rangle \stackrel{7.3.11}{=} \langle A^* v, w \rangle \stackrel{7.4.7}{=} \langle \bar{\lambda} v, w \rangle = \bar{\lambda} \langle v, w \rangle = 0. \quad \square$$

7.4.1 Diagonalisierungssätze für normale Matrizen

Nun können wir unsere wichtigsten Sätze angeben.

Satz 7.4.10. [Hauptsatz über normale Matrizen] *Für $A \in M_n(\mathbb{C})$ sind die folgenden Eigenschaften äquivalent:*

- (i) A ist normal.
- (ii) Es gibt in $\mathbb{C}^n$ eine Orthonormalbasis von Eigenvektoren von A .
- (iii) Es gibt eine unitäre Matrix Q , so dass $A' = Q^{-1}AQ$ eine Diagonalmatrix ist.

Beweis. (i) $\Rightarrow$ (ii): Wir konstruieren induktiv eine Orthonormalbasis aus Eigenvektoren von A . Wir haben in Lemma 5.5.1 bereits gesehen, dass A einen Eigenwert λ_1 besitzt. Sei v_1 ein zugehöriger Eigenvektor. Wir normieren v_1 , so dass $\|v_1\| = 1$ ist.

Ist $n = 1$, so sind wir fertig. Ist dies nicht der Fall, so nehmen wir an, wir hätten bereits ein Orthonormalsystem $(v_1, \dots, v_k)$ von Eigenvektoren von A konstruiert. Nach Lemma 7.4.9 ist der $(n-k)$ -dimensionale Unterraum

$$V_k := \{v_1, \dots, v_k\}^\perp = v_1^\perp \cap \dots \cap v_k^\perp$$

invariant unter A . Ist $k = n$, so sind wir fertig. Ist $k < n$, so hat die lineare Abbildung

$$V_k \rightarrow V_k, v \mapsto Av$$

einen normierten Eigenvektor v_{k+1} . Dann bilden $(v_1, \dots, v_{k+1})$ ein Orthonormalsystem. Dieses Verfahren bricht ab, wenn wir $k = n$ erreicht haben. Wir erhalten also schließlich ein Orthonormalsystem $v_1, \dots, v_n$ von Eigenvektoren von A .

(ii) $\Rightarrow$ (iii). Sei $B' = (v_1, v_2, \dots, v_n)$ eine Orthonormalbasis von $\mathbb{C}^n$ aus Eigenvektoren von A . Sei Q die Matrix, deren Spalten die Vektoren $v_1, \dots, v_n$ sind. Dann ist Q unitär (Satz 7.3.15) und $A' = Q^{-1}AQ$ ist eine Diagonalmatrix. Die Einträge auf der Diagonalen von A' sind die zugehörigen Eigenwerte, siehe Bemerkung 7.4.6.

(iii) $\Rightarrow$ (i). Die Diagonalmatrix A' ist normal. Gilt $A' = Q^{-1}AQ$ für eine unitäre Matrix Q , so folgt $QA'Q^* = QQ^{-1}AQQ^{-1} = A$. Wir folgern aus Lemma 7.4.6(b), dass A ebenfalls normal ist. $\square$

Wir wenden nun diesen Satz auf die speziellen Klassen normaler Matrizen an, die in Beispiel 7.4.3 betrachtet wurden. Wir beginnen mit hermiteschen Matrizen:

Satz 7.4.11. [Hauptsatz über hermitesche Matrizen] *Für $A \in M_n(\mathbb{C})$ sind die folgenden Eigenschaften äquivalent:*

(i) A ist hermitesch.

(ii) A ist normal und alle Eigenwerte sind reell.

(iii) Es gibt eine unitäre Matrix Q , so dass $A' = Q^{-1}AQ$ eine reelle Diagonalmatrix ist.

Beweis. (i) $\Rightarrow$ (ii): Es ist klar, dass jede hermitesche Matrix normal ist, denn $A^*A = A^2 = AA^*$. Ist $v \in \mathbb{C}^n$ ein Eigenvektor von A und $Av = \lambda v$, so folgt aus Lemma 7.4.7 auch $Av = A^*v = \bar{\lambda}v$, also $\lambda = \bar{\lambda}$, da $v \neq 0$ ist. Also ist $\lambda \in \mathbb{R}$.

(ii) $\Rightarrow$ (iii) Aus dem Hauptsatz über normale Matrizen folgt die Existenz einer unitären Matrix Q , so dass $A' := Q^*AQ$ eine Diagonalmatrix ist. Die Einträge λ_i auf der Diagonalen von A' sind die Eigenwerte von A , also reell.

(iii) $\Rightarrow$ (i): Die reelle Diagonalmatrix A' ist trivialerweise hermitesch. Aus $A' = Q^*AQ$ folgt nun $A = QA'Q^*$ und damit

$$A^* = (QA'Q^*)^* = Q^{**}A'^*Q^* = QA'Q^* = A,$$

d.h. A ist hermitesch. □

Satz 7.4.12. [Hauptsatz über reelle symmetrische Matrizen] Für $A \in M_n(\mathbb{R})$ sind die folgenden Eigenschaften äquivalent:

(i) A ist symmetrisch.

(ii) Es gibt in $\mathbb{R}^n$ eine Orthonormalbasis von Eigenvektoren von A .

(iii) Es gibt eine orthogonale Matrix $Q \in O_n(\mathbb{R})$, so dass $A' := Q^{-1}AQ$ eine Diagonalmatrix ist.

Beweis. (i) $\Rightarrow$ (ii). Da A insbesondere eine normale komplexe Matrix ist, existiert eine Orthonormalbasis von $\mathbb{C}^n$ aus Eigenvektoren von A . Insbesondere ist A als komplexe Matrix diagonalisierbar. Da A hermitesch ist, sind nach Satz 7.4.11 alle Eigenwerte von A reell, so dass die Diagonalisierbarkeit von A über $\mathbb{R}$ aus Satz 5.4.10 folgt.

Wählen wir zu jedem Eigenwert eine Orthonormalbasis des zugehörigen Eigenraums, so folgt aus der Orthogonalität der Eigenräume (Satz 7.4.8) die Existenz einer Orthonormalbasis von $\mathbb{R}^n$, die aus Eigenvektoren von A besteht.

(ii) $\Rightarrow$ (iii): Genau wie in Satz 7.4.10 für normale Matrizen.

(iii) $\Rightarrow$ (i): Angenommen, Q ist eine orthogonale Matrix und $A' = Q^T A Q$ ist eine Diagonalmatrix, dann gilt $A = QA'Q^T$ und damit $A^T = (QA'Q^T)^T = Q^{TT}A'^TQ^T = QA'Q^T = A$, d.h. A ist symmetrisch. □

Bemerkung 7.4.13. Reelle symmetrische Matrizen sind normal, aber eine komplexe symmetrische Matrix muss nicht normal sein:

$$A = \begin{pmatrix} i & 1 \\ 1 & 0 \end{pmatrix}, \quad A^* = \begin{pmatrix} -i & 1 \\ 1 & 0 \end{pmatrix}; \quad AA^* = \begin{pmatrix} 2 & i \\ -i & 1 \end{pmatrix} \neq \begin{pmatrix} 2 & -i \\ i & 1 \end{pmatrix} = A^*A.$$

Ist eine komplexe symmetrische Matrix normal, so müssen deren Eigenwerte nicht reell sein. Betrachte z.B. die Matrix $A = \begin{pmatrix} i & 0 \\ 0 & 0 \end{pmatrix}$, welche normal und symmetrisch ist. Ihre Eigenwerte sind $\lambda_1 = i$ und $\lambda_2 = 0$.

Beispiel 7.4.14. Die Matrix

$$A = \frac{1}{9} \begin{pmatrix} 2 & 2 & 10 \\ 2 & 11 & -8 \\ 10 & -8 & 5 \end{pmatrix}$$

ist reell und symmetrisch. Also hat diese Matrix im $\mathbb{R}^3$ eine Orthonormalbasis aus Eigenvektoren v_1, v_2, v_3 . Wir wollen eine solche Orthonormalbasis aus Eigenvektoren explizit berechnen:

Charakteristische Gleichung:

$$\det(A - \lambda \mathbf{1}) = \begin{vmatrix} \frac{2}{9} - \lambda & \frac{2}{9} & \frac{10}{9} \\ \frac{2}{9} & \frac{11}{9} - \lambda & -\frac{8}{9} \\ \frac{10}{9} & -\frac{8}{9} & \frac{5}{9} - \lambda \end{vmatrix} = -\lambda^3 + 2\lambda^2 + \lambda - 2 = -(\lambda - 1)(\lambda - 2)(\lambda + 1).$$

Eigenwerte: $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = -1$.

Eigenvektoren: $w_1 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, w_2 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$ und $w_3 = \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}$.

Da es drei verschiedene Eigenwerte gibt, stehen die zugehörigen Eigenvektoren paarweise orthogonal aufeinander (Satz 7.4.8). Also ergibt Normieren der w_1, w_2, w_3 eine Orthonormalbasis des $\mathbb{R}^3$:

$$v_1 = \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, v_2 = \frac{1}{3} \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, v_3 = \frac{1}{3} \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}.$$

Für die orthogonale Matrix

$$Q := \frac{1}{3} \begin{pmatrix} 2 & 1 & -2 \\ 2 & -2 & 1 \\ 1 & 2 & 2 \end{pmatrix}.$$

ist dann $Q^{-1}AQ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. D.h. wir können A mittels einer orthogonalen Matrix Q diagonalisieren.

Satz 7.4.15. [Hauptsatz über unitäre Matrizen] Für $A \in M_n(\mathbb{C})$ sind die folgenden Eigenschaften äquivalent:

- (i) A ist unitär.
- (ii) A ist normal und alle Eigenwerte λ haben Betrag 1.
- (iii) Es gibt eine unitäre Matrix Q , so dass $A' = Q^{-1}AQ$ eine Diagonalmatrix ist, deren Einträge alle Betrag 1 haben.

Beweis. (i) $\Rightarrow$ (ii): Aus $A^*A = A^{-1}A = \mathbf{1} = AA^{-1} = AA^*$ folgt, dass jede unitäre Matrix normal ist. Aus Satz 7.3.20 wissen wir, dass $|\lambda| = 1$ für alle Eigenwerte von A gilt.

(ii) $\Rightarrow$ (iii) Aus dem Hauptsatz über normale Matrizen folgt die Existenz einer unitären Matrix Q , so dass $A' := Q^*AQ$ eine Diagonalmatrix ist. Die Einträge λ_i auf der Diagonalen von A' sind die Eigenwerte von A , haben also Betrag 1.

(iii) $\Rightarrow$ (i): Die Diagonalmatrix A' ist unitär, da für komplexe Zahlen λ vom Betrag 1 die Relation $\bar{\lambda} = \lambda^{-1}$ gilt. Aus $A' = Q^*AQ$ folgt nun $A = QA'Q^*$ und damit

$$A^* = (QA'Q^*)^* = Q^{**}A'^*Q^* = Q(A')^{-1}Q^{-1} = A^{-1},$$

d.h. A ist unitär. □

Ist A eine unitäre Matrix, so gilt $|\lambda_j| = 1$ für alle Eigenwerte. Also ist λ_j eine komplexe Zahl der Form

$$\lambda_j = e^{i\gamma_j} = \cos \gamma_j + i \sin \gamma_j.$$

Deshalb gilt

$$Q^{-1}AQ = A' = \begin{pmatrix} e^{i\gamma_1} & & & 0 \\ & e^{i\gamma_2} & & \\ & & \ddots & \\ 0 & & & e^{i\gamma_n} \end{pmatrix}.$$

7.4.2 Normalform orthogonaler Matrizen

In diesem Abschnitt dehnen wir die Klassifikation der orthogonalen $(n \times n)$ -Matrizen, die wir bereits für $n = 2, 3$ durchgeführt haben (Sätze 7.3.21 und 7.3.22) auf den allgemeinen Fall aus.

Satz 7.4.16. (Normalform reeller orthogonaler Matrizen) *Sei A eine (reelle) orthogonale $n \times n$ -Matrix. Dann existiert eine orthogonale Matrix $Q \in O_n(\mathbb{R})$ mit*

$$Q^{-1}AQ = \begin{pmatrix} \overbrace{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}^p & \overbrace{\begin{pmatrix} -1 & & 0 \\ & \ddots & \\ 0 & & -1 \end{pmatrix}}^q & \overbrace{\begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}}^{2r} \\ 0 & & \begin{matrix} \boxed{D_1} \\ \ddots \\ \boxed{D_r} \end{matrix} \end{pmatrix}.$$

Hierbei ist:

- p die Vielfachheit des Eigenwertes $+1$ von A ;
- q die Vielfachheit des Eigenwertes -1 von A ;
- r die Anzahl von Paaren komplex konjugierter Eigenwerte $e^{\pm i\gamma_j}$ von A mit $\gamma_j \in]0, \pi[$.

Jeder Block $D_j \in M_2(\mathbb{R})$ ist eine Drehmatrix der Form

$$D_j = \begin{pmatrix} \cos \gamma_j & -\sin \gamma_j \\ \sin \gamma_j & \cos \gamma_j \end{pmatrix}$$

Beweis. Eine orthogonale Matrix ist ein Spezialfall einer unitären Matrix. Wir gehen deshalb zunächst vor wie für unitäre Matrizen in Satz 7.4.15:

(a) Wir bestimmen die komplexen Eigenwerte von A : Sei p die Vielfachheit des Eigenwertes 1 und q die Vielfachheit des Eigenwertes -1 . Die anderen Eigenwerte haben die Form $e^{\pm i\gamma} = \cos \gamma \pm i \sin \gamma$ mit $0 < \gamma < \pi$. Da A eine reelle Matrix ist, hat das charakteristische Polynom $\chi_A(X)$ reelle Koeffizienten. Ist $\lambda = a + ib$ eine komplexe Wurzel von $\chi_A(X)$, dann ist $\bar{\lambda} = a - ib$ ebenso eine Wurzel von $\chi_A(X)$, da $\chi_A(\bar{\lambda}) = \overline{\chi_A(\lambda)} = 0$ ist. Also ist für jeden Eigenwert $e^{i\gamma_j} = \cos \gamma_j + i \sin \gamma_j$ von A die komplex konjugierte Zahl $e^{-i\gamma_j} = \cos \gamma_j - i \sin \gamma_j$ ebenfalls ein Eigenwert von A . Wir ordnen diese Eigenwerte nach Paaren komplex konjugierter Zahlen. Die Eigenwerte ordnen wir wie folgt nämlich:

$$\underbrace{1, \dots, 1}_p, \underbrace{-1, \dots, -1}_q, \underbrace{e^{i\gamma_1}, e^{-i\gamma_1}, \dots, e^{i\gamma_r}, e^{-i\gamma_r}}_{2r}.$$

entsprechend ihrer Vielfachheit.

(b) Da A insbesondere unitär ist, existiert in $\mathbb{C}^n$ eine Orthonormalbasis von Eigenvektoren zu diesen Eigenwerten:

$$v_1, \dots, v_p, v_{p+1}, \dots, v_{p+q}, w_1, \bar{w}_1, \dots, w_r, \bar{w}_r$$

Man beachte, dass die Eigenvektoren $v_1, \dots, v_p, v_{p+1}, \dots, v_{p+q}$ zu den reellen Eigenwerten ± 1 reell gewählt werden können, denn die algebraische Vielfachheit der Eigenwerte ± 1 ist über $\mathbb{R}$ und

$\mathbb{C}$ gleich. Man kann sogar für die komplex konjugierten Eigenwerte die Eigenvektoren komplex konjugiert wählen: Aus $Aw = \lambda w$ und $\bar{A} = A$ folgt $A\bar{w} = \bar{A}\bar{w} = \bar{A}w = \bar{\lambda}w = \bar{\lambda}\bar{w}$.

Wir schreiben $w_j = x_j - iy_j$ mit $x_j, y_j \in \mathbb{R}^n$. Aus $w_j \perp \bar{w}_j = x_j + iy_j$ folgt dann

$$0 = \langle x_j + iy_j, x_j - iy_j \rangle = \|x_j\|^2 - \|y_j\|^2 + 2i\langle y_j, x_j \rangle.$$

Durch Vergleich von Real- und Imaginärteil erhalten wir also

$$x_j \perp y_j \quad \text{und} \quad \|x_j\| = \|y_j\|.$$

Aus $1 = \|w_j\|^2 = \|x_j\|^2 + \|y_j\|^2$ folgt daher $\|x_j\| = \|y_j\| = \frac{1}{\sqrt{2}}$. Mit $u_j := \sqrt{2}x_j$ und $u'_j := \sqrt{2}y_j$ erhalten wir daher eine ONB

$$v_1, \dots, v_p, v_{p+1}, \dots, v_{p+q}, u_1, u'_1, \dots, u_r, u'_r \quad (7.7)$$

von $\mathbb{R}^n$. Aus Bemerkung 5.5.9 wissen wir, dass die von A auf $\mathbb{R}u_j + \mathbb{R}u'_j$ induzierte lineare Abbildung bzgl. der Basis (u_j, u'_j) zur Matrix

$$D_j = \begin{pmatrix} \cos \gamma_j & -\sin \gamma_j \\ \sin \gamma_j & \cos \gamma_j \end{pmatrix},$$

gehört. Die Behauptung folgt also, wenn $Q \in O_n(\mathbb{R})$ die orthogonale Matrix ist, deren Spalten die ONB (7.7) ist. $\square$

Aufgaben zu Abschnitt 7.4

Aufgabe 7.4.17. Seien $A, B \in M_n(\mathbb{C})$ normal sowie

$$AB = BA \quad \text{und} \quad AB^* = B^*A.$$

Zeigen Sie, dass $A + B$ normal ist.

Aufgabe 7.4.18. Sei $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \in M_n(\mathbb{K})$ ein Blockmatrix. Dann ist

$$A^* = \begin{pmatrix} A_{11}^* & A_{21}^* \\ A_{12}^* & A_{22}^* \end{pmatrix}.$$

Aufgabe 7.4.19. Sei $A \in M_n(\mathbb{K})$ normal mit der Blockstruktur

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad A_{11} \in M_p(\mathbb{K}), A_{12} \in M_{p,q}(\mathbb{K}), A_{22} \in M_q(\mathbb{K}), \quad p + q = n.$$

Zeigen Sie: $A_{12} = 0$. Folgern Sie daraus, daß für jeden A -invarianten Unterraum $U \subseteq \mathbb{K}^n$ auch sein orthogonales Komplement $U^\perp$ invariant ist.

Hinweis: Aus der Normalität leite man die Relation $A_{11}A_{11}^* + A_{12}A_{12}^* = A_{11}^*A_{11}$ ab und schließe daraus

$$\|A_{11}\|_2^2 + \|A_{12}\|_2^2 = \|A_{11}^*\|_2^2 = \|A_{11}\|_2^2$$

für die Hilbert-Schmidt Norm $\|X\|_2^2 = \text{tr}(XX^*)$ der Matrizen (vgl. Aufgabe 7.3.29).

7.5 Die Adjungierte einer Abbildung unitärer Räume

Jede Operation auf Matrizen korrespondiert zu einer Operation auf linearen Abbildungen. Wir kennen bereits die Adjungierte A^* einer Matrix A . Was ist die Adjungierte einer linearen Abbildung?

Im Folgenden seien V, W unitäre $\mathbb{K}$ -Vektorräume und $\varphi : V \rightarrow W$ eine lineare Abbildung.

Definition 7.5.1. [Die Adjungierte einer linearen Abbildung] Eine Abbildung $\varphi^* : W \rightarrow V$ heißt *adjungiert zu* $\varphi : V \rightarrow W$, wenn

$$\langle \varphi^*(w), v \rangle_V = \langle w, \varphi(v) \rangle_W \quad \text{für alle } v \in V, w \in W$$

gilt.

Wie wir in Lemma 7.3.11 gesehen haben, ist für $A \in M_{m,n}(\mathbb{K})$ die Adjungierte der linearen Abbildung $L_A : \mathbb{K}^n \rightarrow \mathbb{K}^m, x \mapsto Ax$, die lineare Abbildung $L_{A^*} : \mathbb{K}^m \rightarrow \mathbb{K}^n$.

Es ist i.A. nicht offensichtlich, dass eine lineare Abbildung $\varphi : V \rightarrow W$ eine Adjungierte besitzt. Besitzt sie eine Adjungierte, so ist nicht offensichtlich, ob die Adjungierte linear und eindeutig ist:

Satz 7.5.2. *Hat eine lineare Abbildung $\varphi : V \rightarrow W$ eine Adjungierte $\varphi^* : W \rightarrow V$, dann ist diese Adjungierte φ^* linear und durch φ eindeutig bestimmt.*

Beweis. Eindeutigkeit: Angenommen, φ^* und φ' sind zwei Adjungierte von φ , d.h. es gilt

$$\langle \varphi^*(w), v \rangle = \langle w, \varphi(v) \rangle \quad \text{und} \quad \langle \varphi'(w), v \rangle = \langle w, \varphi(v) \rangle$$

für alle $v \in V, w \in W$. Dann folgern wir $\langle \varphi^*(w) - \varphi'(w), v \rangle = \langle \varphi^*(w), v \rangle - \langle \varphi'(w), v \rangle = 0$, d.h.

$$\varphi^*(w) - \varphi'(w) \in V^\perp = \{0\}.$$

Damit ist $\varphi' = \varphi^*$.

Linearität: Wir müssen $\varphi^*(u+w) = \varphi^*(u) + \varphi^*(w)$ und $\varphi^*(\lambda w) = \lambda \varphi^*(w)$ für $u, w \in W$ und $\lambda \in \mathbb{K}$ zeigen. Es gilt für alle $v \in V, u, w \in W$

$$\begin{aligned} \langle \varphi^*(u+w), v \rangle &= \langle u+w, \varphi(v) \rangle = \langle u, \varphi(v) \rangle + \langle w, \varphi(v) \rangle = \langle \varphi^*(u), v \rangle + \langle \varphi^*(w), v \rangle \\ &= \langle \varphi^*(u) + \varphi^*(w), v \rangle \end{aligned}$$

Da dies für alle $v \in V$ gilt, erhalten wir $\varphi^*(u+w) = \varphi^*(u) + \varphi^*(w)$. Analog zeigt man $\varphi^*(\lambda w) = \lambda \varphi^*(w)$. $\square$

Auf unendlichdimensionalen Vektorräumen V hat nicht jede lineare Abbildung eine Adjungierte. Auf endlichdimensionalen Räumen existiert die Adjungierte immer:

Satz 7.5.3. [Existenz der Adjungierten] *Sind V, W endlichdimensionale unitäre $\mathbb{K}$ -Vektorräume, so hat jede lineare Abbildung $\varphi : V \rightarrow W$ eine Adjungierte $\varphi^* : W \rightarrow V$.*

Ist $B = (b_1, \dots, b_n)$ eine ONB von V und $C = (c_1, \dots, c_m)$ eine ONB von W und $A = [\varphi]_B^C$ die Matrix von φ bzgl. B und C , so ist A^ die Matrix $[\varphi^*]_C^B$ von φ^* bzgl. C, B .*

Beweis. Aus dem Darstellungssatz von Fréchet–Riesz 7.2.20 folgt für jedes $v \in V$ die Existenz genau eines $\varphi^*(v) \in W$ mit

$$\langle w, \varphi^*(v) \rangle = \langle \varphi(w), v \rangle \quad \text{für alle } w \in W,$$

denn die Abbildung $W \rightarrow \mathbb{K}, w \mapsto \langle \varphi(w), v \rangle$ ist linear. Aus

$$\langle \varphi^*(v), w \rangle = \overline{\langle \varphi(w), v \rangle} = \langle v, \varphi(w) \rangle$$

folgt daher die Existenz von φ^* .

Ist $B = (b_1, \dots, b_n)$ eine ONB von V und $C = (c_1, \dots, c_m)$ eine ONB von W sowie $A = (a_{ij}) = [\varphi]_B^C$ die zugehörige Matrix von φ , so ist der Eintrag von $[\varphi^*]_C^B$ an der Stelle (i, j) gegeben durch

$$\langle \varphi^*(c_j), b_i \rangle = \langle c_j, \varphi(b_i) \rangle = \overline{a_{ji}}$$

und daher $[\varphi^*]_C^B = A^*$. $\square$

Definition 7.5.4. [Normale Endomorphismen] Ein Endomorphismus φ des endlichdimensionalen unitären Raums V heißt

(a) *normal*, wenn $\varphi \circ \varphi^* = \varphi^* \circ \varphi$.

(b) *selbstadjungiert*, wenn $\varphi^* = \varphi$.

Satz 7.5.5. Für einen Endomorphismus φ eines endlichdimensionalen unitären Vektorraumes sind die folgenden Eigenschaften äquivalent:

(i) φ ist normal

(ii) Es gibt in V eine Orthonormalbasis von Eigenvektoren für φ

(iii) Die Matrix von φ ist diagonal bzgl. einer Orthonormalbasis B .

Beweis. Ist V endlichdimensional, so ist φ nach Satz 7.5.3 genau dann normal, wenn die Matrix $[\varphi]_B$ bzgl. einer Orthonormalbasis B von V normal ist, denn es ist

$$[\varphi\varphi^*]_B = [\varphi]_B[\varphi^*]_B = [\varphi]_B[\varphi]_B^* \quad \text{und} \quad [\varphi^*\varphi]_B = [\varphi^*]_B[\varphi]_B = [\varphi]_B^*[\varphi]_B.$$

Die Behauptung folgt damit aus Satz 7.4.10. $\square$

Bemerkung 7.5.6. (a) Ist V endlichdimensional, so ist φ nach Satz 7.5.3 genau dann selbstadjungiert, wenn die Matrix A von φ bzgl. irgendeiner Orthonormalbasis hermitesch ist.

(b) Aus Satz 7.4.11 folgt, dass alle Eigenwerte eines selbstadjungierten Endomorphismus reell sind.

(c) Wir folgern aus Satz 7.4.12, dass ein selbstadjungierter Endomorphismus eines endlichdimensionalen euklidischen Vektorraumes eine Orthonormalbasis von Eigenvektoren besitzt und alle Eigenwerte reell sind.

Aufgaben zu Abschnitt 7.5

Aufgabe 7.5.7. Sei V ein endlichdimensionaler unitärer Raum und $\varphi \in \text{End}(V)$ normal. Zeigen Sie: Ist $U \subseteq V$ ein φ -invarianter Unterraum, so ist auch sein orthogonales Komplement $U^\perp$ invariant unter φ . Hinweis: Aufgabe 7.4.19.

Aufgabe 7.5.8. (Simultane Diagonalisierbarkeit) Sei V ein endlichdimensionaler unitärer Raum und $\varphi, \psi \in \text{End}(V)$ normal. Zeigen Sie: Genau dann existiert in V eine ONB aus simultanen Eigenvektoren für φ und ψ , wenn φ mit ψ vertauscht, d.h. $\varphi\psi = \psi\varphi$. Hinweis: Lemma 6.2.5. Können Sie diese Aussage auf eine beliebige Teilmenge $\mathcal{S} \subseteq \text{End}(V)$ normaler Endomorphismen verallgemeinern?

7.6 Singulärwertzerlegung

In diesem Abschnitt sei V ein endlichdimensionaler unitärer Raum.

Definition 7.6.1. Für $x, y \in V$ schreiben wir $P_{x,y}$ für den Rang-1-Operator, der durch $P_{x,y}(v) = \langle v, y \rangle x$ gegeben ist, und setzen $P_x := P_{x,x}$. Für das Rechnen mit solchen Operatoren gelten

$$P_{x,y}^* = P_{y,x} \quad \text{und} \quad P_{x,y}P_{z,w} = \langle z, y \rangle P_{x,w}, \quad (7.8)$$

wie man leicht verifiziert (Übung!).

Bemerkung 7.6.2. (a) Ist $\|x\| = 1$, so erhalten wir mit (7.8) sofort $P_x = P_x^2 = P_x^*$. Weiter ist $P_x(x) = x$ und $P_x(y) = 0$ für $y \perp x$. Also ist P_x die Orthogonalprojektion auf $\mathbb{K}x$.

(b) Sind $x \perp y$, so folgt $P_x P_y = P_y P_x = 0$ aus (7.8).

(c) Ist $x_1, \dots, x_n$ eine Orthonormalbasis des Unterraums U , so ist

$$P = \sum_{j=1}^n P_{x_j}, \quad P(y) = \sum_{j=1}^n \langle y, x_j \rangle x_j$$

die Orthogonalprojektion auf U (Satz 7.2.16).

Ist $\varphi \in \text{End}(V)$ normal und $\dim V < \infty$, so haben wir in Satz 7.5.5 gesehen, dass in V eine ONB $e_1, \dots, e_n$ aus Eigenvektoren von φ existiert. Sind $\lambda_1, \dots, \lambda_n$ die zugehörigen Eigenwerte, so erhalten wir die sogenannte *Spektraldarstellung* von φ

$$\varphi = \sum_{j=1}^n \lambda_j P_{e_j}, \quad \varphi(v) = \sum_{j=1}^n \lambda_j \langle v, e_j \rangle e_j, \quad v \in V,$$

denn wegen $P_{e_j} e_k = \delta_{jk} e_k$ haben beide Seiten auf den Basisvektoren e_j die gleichen Werte.

Fassen wir jeweils gleiche Eigenwerte zusammen, so erhalten wir die folgende Spektraldarstellung

$$\varphi = \sum_{k=1}^m \lambda_k P_k,$$

wobei P_k die Orthogonalprojektion auf den Eigenraum zum Eigenwerte λ_k ist. In der Tat haben beide Seiten jeweils auf den Basisvektoren e_j die gleichen Werte.

Satz 7.6.3. (Singulärwertzerlegung) *Sei $\varphi: V \rightarrow W$ eine lineare Abbildung der unitären Räume U und V . Dann existieren Orthonormalbasen $(g_1, \dots, g_m)$ von W und $(f_1, \dots, f_n)$ von V sowie reelle Zahlen $\sigma_j \geq 0$, $j = 1, \dots, k := \min(n, m)$ mit*

$$\varphi = \sum_{j=1}^k \sigma_j P_{g_j, f_j} \quad \text{d.h.} \quad \varphi(v) = \sum_{j=1}^k \sigma_j \langle v, f_j \rangle g_j \quad \text{für } v \in V.$$

Setzen wir $\sigma_j := 0$ für $k < j \leq \max(n, m)$, so sind $f_1, \dots, f_n$ Eigenvektoren von $\varphi^* \varphi$ zu den Eigenwerten σ_j^2 und $g_1, \dots, g_m$ Eigenvektoren von $\varphi \varphi^*$ zu den Eigenwerten σ_j^2 .

Beweis. Sei $\psi := \varphi^* \varphi$. Da ψ hermitesch ist, existiert nach Bemerkung 7.5.6 eine orthonormale Basis $(f_1, \dots, f_n)$ von V aus Eigenvektoren zu den Eigenwerten $\lambda_1, \dots, \lambda_n$. Weiter ist

$$\lambda_j = \langle \psi(f_j), f_j \rangle = \langle \varphi^* \varphi(f_j), f_j \rangle = \langle \varphi(f_j), \varphi(f_j) \rangle = \|\varphi(f_j)\|^2 \geq 0.$$

Für $j = 1, \dots, n$ setzen wir $\sigma_j := \sqrt{\lambda_j}$.

Für $\sigma_j > 0$, d.h. $\varphi(f_j) \neq 0$, setzen wir $g_j := \sigma_j^{-1} \varphi(f_j)$. Für zwei solche Vektoren ist dann

$$\langle g_j, g_k \rangle = \sigma_j^{-1} \sigma_k^{-1} \langle \varphi(f_j), \varphi(f_k) \rangle = \sigma_j^{-1} \sigma_k^{-1} \langle \varphi^* \varphi(f_j), f_k \rangle = \sigma_j^{-1} \sigma_k^{-1} \lambda_j \langle f_j, f_k \rangle = \delta_{jk} \sigma_j^{-1} \sigma_k^{-1} \lambda_j = \delta_{jk}.$$

Also ist $\{g_j : \sigma_j > 0\}$ ein Orthonormalsystem, das wir nun zu einer Orthonormalbasis $g_1, \dots, g_m$ von W ergänzen.

Ist $v = \sum_j \langle v, f_j \rangle f_j$ ein beliebiges Element von V , so ist

$$\varphi(v) = \sum_j \langle v, f_j \rangle \varphi(f_j) = \sum_j \sigma_j \langle v, f_j \rangle g_j.$$

Wir erhalten somit die Darstellung $\varphi = \sum_j \sigma_j P_{g_j, f_j}$ und somit

$$\varphi \varphi^* = \sum_{i,j} \sigma_i \sigma_j P_{g_i, f_i} P_{g_j, f_j}^* = \sum_{i,j} \sigma_i \sigma_j P_{g_i, f_i} P_{f_j, g_j} = \sum_{i,j} \sigma_i \sigma_j \delta_{ij} P_{g_i} = \sum_j \sigma_j^2 P_{g_j}.$$

Also sind $g_1, \dots, g_m$ Eigenvektoren von $\varphi \varphi^*$ zu den Eigenwerten σ_j^2 . □

Definition 7.6.4. Eine Matrix $D \in M_{m,n}(\mathbb{K})$ heißt *verallgemeinerte Diagonalmatrix*, wenn $d_{ij} = 0$ für $i \neq j$ gilt. Ist $k = \min(n, m)$, so schreiben wir dann

$$D = \text{diag}(d_{11}, \dots, d_{kk}),$$

um verallgemeinerte Diagonalmatrizen zu spezifizieren.

Satz 7.6.3 lässt sich wie folgt für Matrizen formulieren:

Folgerung 7.6.5. (Singulärwertzerlegung für Matrizen) Für jede Matrix $A \in M_{m,n}(\mathbb{K})$ und $k = \min(n, m)$ existieren unitäre Matrizen $U \in U_m(\mathbb{K})$ und $V \in U_n(\mathbb{K})$, so dass

$$\Sigma := U^{-1}AV = \text{diag}(\sigma_1, \dots, \sigma_k)$$

eine verallgemeinerte Diagonalmatrix mit nichtnegativen Einträgen ist.

Die Zahlen $\sigma_1, \dots, \sigma_k$ heißen *Singulärwerte* der Matrix A .

Beweis. Wir betrachten die lineare Abbildung $L_A: \mathbb{K}^n \rightarrow \mathbb{K}^m$. Aus Satz 7.6.3 folgt nun die Existenz von Orthonormalbasen $g_1, \dots, g_m$ von $\mathbb{K}^m$, $f_1, \dots, f_n$ von $\mathbb{K}^n$ und nichtnegative Zahlen $\sigma_j \geq 0$, so dass

$$L_A x = Ax = \sum_{j=1}^n \sigma_j \langle x, f_j \rangle g_j, \quad x \in \mathbb{K}^n.$$

Ist $U \in U_m(\mathbb{K})$ die unitäre Matrix deren Spalten die Vektoren $Ue_j = g_j$ sind und $V \in U_n(\mathbb{K})$ die unitäre Matrix deren Spalten die Vektoren $Ve_j = f_j$ sind, so erhalten wir

$$Ax = \sum_{j=1}^n \sigma_j \langle x, Ve_j \rangle Ue_j = U \sum_{j=1}^n \sigma_j \langle V^{-1}x, e_j \rangle e_j, \quad x \in \mathbb{K}^n,$$

also

$$U^{-1}AVx = \sum_{j=1}^n \sigma_j \langle x, e_j \rangle e_j = \Sigma x$$

für $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$. □

Bemerkung 7.6.6. (a) Sei $m \leq n$ (also $k = m$). Dann ist

$$\text{diag}(\sigma_1^2, \dots, \sigma_m^2) = \Sigma \Sigma^* = U^{-1}AVV^*A^*(U^{-1})^* = U^{-1}AA^*U$$

unitär ähnlich zu der hermiteschen $(m \times m)$ -Matrix AA^* . Die Zahlen $\sigma_1^2, \dots, \sigma_m^2$ sind also die Eigenwerte von AA^* .

(b) Sei $n \leq m$ (also $k = n$). Dann ist

$$\text{diag}(\sigma_1^2, \dots, \sigma_n^2) = \Sigma^* \Sigma = V^*A^*(U^{-1})^*U^{-1}AV = V^{-1}A^*AV$$

unitär ähnlich zu der hermiteschen $(n \times n)$ -Matrix A^*A . Die Zahlen $\sigma_1^2, \dots, \sigma_n^2$ sind also die Eigenwerte von A^*A .

Bemerkung 7.6.7. Zur Berechnung der Singulärwertzerlegung einer Matrix A kann man entsprechend der Konstruktion im Beweis von Satz 7.6.3 vorgehen: Man bestimmt zuerst eine Orthonormalbasis $f_1, \dots, f_n$ von Eigenvektoren der Matrix A^*A . Wegen $\text{rank}(A^*A) \leq \text{rank}(A) \leq k$, sind höchstens k der zugehörigen Eigenwerte λ_j von 0 verschieden.

Wir ordnen sie der Größe nach:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0 = \lambda_{k+1} = \dots = \lambda_n$$

(falls $n \geq m$ ist). Für $\lambda_j > 0$ setzen wir dann

$$g_j := \sqrt{\lambda_j}^{-1} A f_j$$

und ergänzen diese Vektoren zu einer ONB $g_1, \dots, g_m$ von $\mathbb{K}^m$.

Kapitel 8

Quotientenraum und Dualraum

8.1 Der Quotientenraum

In der Mathematik treten oft Situationen auf, in denen Objekte in Bezug auf eine spezifische Eigenschaft nicht eindeutig spezifiziert werden können. Dies führt dazu, dass man statt einzelner Elemente die Klassen einer geeigneten Äquivalenzrelation betrachten muss. Trägt die ursprünglich betrachtete Menge zusätzliche Struktur, ist sie zum Beispiel eine Gruppe, ein Ring oder ein Vektorraum, so stellt sich insbesondere die Frage, ob die Menge der Äquivalenzklassen diese Struktur erbt. In diesem Abschnitt diskutieren wir diese Frage für Vektorräume. In weiterführenden Algebra-Vorlesungen werden ähnliche Konstruktionen bzw. Resultate immer wieder auftreten.

Um ein Gefühl für die auftretenden Strukturen zu bekommen, betrachten wir zuerst ein paar Beispiele.

Beispiel 8.1.1. (Lösungen linearer Gleichungssysteme) Wir haben im ersten Teil der Vorlesung gesehen, dass die Lösungsmenge eines linearen Gleichungssystems

$$Ax = b, \quad b \in \mathbb{K}^m, A \in M_{m,n}(\mathbb{K})$$

entweder leer ist oder ein affiner Unterraum der Gestalt

$$\bar{x} + \ker A,$$

wobei $\bar{x} \in \mathbb{K}^n$ eine spezielle Lösung ist.

In diesem Kontext stellt sich also die Frage, inwiefern man diese Lösungsmengen (für variable $b \in \mathbb{K}^m$) selbst wieder als Elemente eines Vektorraums auffassen kann.

Beispiel 8.1.2. (Datenreduktion) Wir stellen uns vor, dass die Elemente $x \in \mathbb{K}^n$ Informationen über gewisse Objekte kodieren. Man könnte sich x z.B. als einen Eintrag in einer Datenbank vorstellen. Jeder Vektor x enthält die Komponenten $x_1, \dots, x_n$ und in diesem Sinne n -verschiedene Daten über ein Objekt.

Oft ist man allerdings gar nicht an der vollständigen Information interessiert, sondern nur an gewissen Daten. Wir denken uns daher, dass wir nur an $x_1, \dots, x_m$ interessiert sind und $x_{m+1}, \dots, x_n$ für unsere Zwecke irrelevant sind.

Für jeden Vektor $\bar{x} \in \mathbb{K}^m$ liefert jedes Element der Menge

$$\left\{ \begin{pmatrix} x \\ x' \end{pmatrix} \in \mathbb{K}^n : x = \bar{x} \in \mathbb{K}^m, x' \in \mathbb{K}^{n-m} \right\}$$

die gleichen Daten. Auch hier handelt es sich wieder um einen affinen Unterraum

$$\begin{pmatrix} \bar{x} \\ 0 \end{pmatrix} + U, \quad U := \left\{ \begin{pmatrix} 0 \\ x' \end{pmatrix} \in \mathbb{K}^n : x' \in \mathbb{K}^{n-m} \right\}.$$

In diesem Fall ist klar, wie man die auftretenden affinen Unterräume als Punkte eines Vektorraums auffassen kann, denn sie stehen in bijektiver Korrespondenz zu den Elementen $\bar{x} \in \mathbb{K}^m$.

Definition 8.1.3. Sei U ein linearer Unterraum des Vektorraums V . Wir haben im ersten Teil der Linearen Algebra bereits gesehen, dass durch

$$x \sim_U y \quad :\Leftrightarrow \quad x - y \in U$$

auf V eine Äquivalenzrelation definiert wird, deren Äquivalenzklassen genau die affinen Unterräume der Gestalt

$$[x] = \{y \in V : y \sim_U x\} = x + U, \quad x \in V$$

sind.¹

Wir schreiben $V/U := V/\sim_U = \{x + U : x \in V\}$ für die Menge der Äquivalenzklassen. Um auf dieser Menge die Struktur eines Vektorraums zu definieren, beachten wir, dass für $x + U = x' + U$ und $y + U = y' + U$ und $\lambda \in \mathbb{K}$ gilt:

$$x + y + U = x' + y' + U \quad \text{und} \quad \lambda x + U = \lambda x' + U,$$

denn

$$x + y - (x' + y') = (x - x') + (y - y') \in U + U \subseteq U \quad \text{und} \quad \lambda x - \lambda x' = \lambda(x - x') \in \lambda U \subseteq U.$$

Wir erhalten daher eine wohldefinierte Addition

$$+ : V/U \times V/U \rightarrow V/U, \quad (x + U) + (y + U) := x + y + U$$

und eine wohldefinierte Skalarmultiplikation

$$\cdot : \mathbb{K} \times V/U \rightarrow V/U, \quad \lambda \cdot (x + U) := \lambda x + U.$$

Bemerkung 8.1.4. Die Addition der affinen Unterräume kann man auch als elementweise Addition von Teilmengen von V auffassen, denn

$$\{x + u : u \in U\} + \{y + u' : u' \in U\} = \{x + u + y + u' : u, u' \in U\} = \{x + y + u : u \in U\} = x + y + U.$$

Bei der Skalarmultiplikation ist allerdings Vorsicht geboten: Für $0 \neq \lambda$ ist zwar

$$\{\lambda(x + u) : u \in U\} = \{\lambda x + \lambda u : u \in U\} = \{\lambda x + u : u \in U\} = \lambda x + U,$$

aber $\{0(x + u) : u \in U\} = \{0\}$ besteht nur aus der 0, ist also von der Teilmenge $U \subseteq V$ i.a. verschieden.

Satz 8.1.5. Addition und Skalarmultiplikation definieren auf V/U die Struktur eines $\mathbb{K}$ -Vektorraums mit Nullelement $0 + U = U$.

Die Quotientenabbildung

$$q_U : V \rightarrow V/U, \quad x \mapsto x + U$$

ist eine surjektive lineare Abbildung mit $\ker(q_U) = U$.

Ist V endlichdimensional, so gilt insbesondere

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

¹Der Vollständigkeit halber wiederholen wir das Argument: Zur Verifikation der Eigenschaften einer Äquivalenzrelation haben wir zunächst zu zeigen, dass die reflexiv ist: $x \sim_U x$ ist wegen $x - x = 0 \in U$ trivial. Die Symmetrie folgt aus $y - x = -(x - y) \in U$ falls $x \sim_U y$. Schließlich erhalten wir die Transitivität wie folgt: $x \sim_U y, y \sim_U z$ führt zu $x - z = (x - y) + (y - z) \in U + U \subseteq U$.

Beweis. Zuerst beachten wir, dass q_U eine surjektive Abbildung ist, die mit Addition und Skalarmultiplikation verträglich ist, d.h.

$$q_U(x + y) = x + y + U = q_U(x) + q_U(y) \quad \text{und} \quad q_U(\lambda x) = \lambda q_U(x), \quad x, y \in V, \lambda \in \mathbb{K}. \quad (8.1)$$

Das ergibt sich unmittelbar aus den Definitionen.

Nun lassen sich leicht die Axiome für einen Vektorraum verifizieren, da man jedes Element von V/U als Bild eines Elements aus V schreiben kann. Zum Beispiel ergibt sich die Assoziativität der Addition aus

$$\begin{aligned} (q_U(x) + q_U(y)) + q_U(z) &= q_U(x + y) + q_U(z) = q_U((x + y) + z) = q_U(x + (y + z)) \\ &= q_U(x) + q_U(y + z) = q_U(x) + (q_U(y) + q_U(z)). \end{aligned}$$

Ebenso verifiziert man z.B. die Distributivgesetze:

$$\begin{aligned} \lambda(q_U(x) + q_U(y)) &= \lambda q_U(x + y) = q_U(\lambda(x + y)) = q_U(\lambda x + \lambda y) = q_U(\lambda x) + q_U(\lambda y) \\ &= \lambda q_U(x) + \lambda q_U(y). \end{aligned}$$

Die Verifikation der anderen Axiome ist eine leichte Übung.

Nachdem wir wissen, dass V/U ein $\mathbb{K}$ -Vektorraum ist, bedeutet (8.1), dass $q_U: V \rightarrow V/U$ eine lineare Abbildung ist. Aus der Konstruktion folgt unmittelbar, dass

$$\ker q_U = \{x \in V : x + U = U\} = U$$

ist.

Ist V endlichdimensional, liefert die Dimensionsformel für lineare Abbildungen sofort

$$\dim(V/U) = \dim(\operatorname{im}(q_U)) = \dim V - \dim(\ker q_U) = \dim V - \dim U. \quad \square$$

Definition 8.1.6. Ist $U \subseteq V$ ein linearer Unterraum, so definieren wir die *Kodimension von U* durch

$$\operatorname{codim}(U) := \dim(V/U).$$

Ist $\operatorname{codim}(U) = 1$, so nennen wir U eine *Hyperebene*.

Beispiel 8.1.7. (a) Es kann passieren, dass zwar V unendlichdimensional ist, aber V/U endlichdimensional. Ein einfaches Beispiel ist der Raum $V = \mathbb{K}[X]$ der formalen Polynome und der Unterraum

$$U = (X^d) = X^d \mathbb{K}[X] = \operatorname{span}\{X^m : m \geq d\}.$$

In der Tat ist

$$\mathbb{K}[X] = U \oplus \operatorname{span}\{X^j : 0 \leq j < d\},$$

so dass wir aus Bemerkung 8.1.11 die Beziehung $\operatorname{codim}(U) = \dim(\mathbb{K}[X]/U) = d$ erhalten. Für $d = 1$ erhalten wir eine Hyperebene.

Beispiel 8.1.8. Für $V = \mathbb{K}^2$ und $U = \mathbb{K} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ist die zugehörige Äquivalenzrelation gegeben durch

$$x \sim_U y \Leftrightarrow (\exists \lambda \in \mathbb{K}) x - y = \begin{pmatrix} \lambda \\ -\lambda \end{pmatrix} \Leftrightarrow (\exists \lambda \in \mathbb{K}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 + \lambda \\ y_2 - \lambda \end{pmatrix} \Leftrightarrow x_1 + x_2 = y_1 + y_2.$$

Wegen $\dim U = 1$ und $\dim V = 2$ ist $\dim V/U = 1$. Das Element $\begin{pmatrix} 1 \\ 0 \end{pmatrix} + U$ ist also eine Basis von V/U .

Wir sehen auch, dass die lineare Abbildung

$$\varphi: V \rightarrow \mathbb{K}, \quad x \mapsto x_1 + x_2$$

die Eigenschaft besitzt, dass $x + U = y + U$ äquivalent ist zu $\varphi(x) = \varphi(y)$. Im diesem Sinne verhält sie sich vollkommen analog zur Quotientenabbildung q_U . Diesen Sachverhalt wollen wir jetzt besser verstehen.

Der folgende Satz drückt, zusammen mit seinen zahlreichen Verwandten, einen ganz zentralen Sachverhalt aus, nämlich die Verträglichkeit algebraischer Strukturen mit dem Übergang zu Mengen von Äquivalenzklassen.

Der Homomorphiesatz für lineare Abbildungen

Satz 8.1.9. *Sei $\varphi: V \rightarrow W$ eine lineare Abbildung und $U \subseteq V$ ein linearer Unterraum. Dann existiert genau dann eine lineare Abbildung $\bar{\varphi}: V/U \rightarrow W$ mit $\bar{\varphi} \circ q_U = \varphi$, wenn $U \subseteq \ker \varphi$ gilt. Ist sogar $\ker(\varphi) = U$, so ist $\bar{\varphi}$ injektiv, liefert also einen linearen Isomorphismus*

$$V/\ker \varphi \rightarrow \operatorname{im}(\varphi).$$

In Diagrammen stellen wir uns das wie folgt vor:

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ \downarrow q_U & & \uparrow \\ V/U & \xrightarrow{\bar{\varphi}} & \operatorname{im}(\varphi). \end{array}$$

Beweis. Existiert eine lineare Abbildung $\bar{\varphi}: V/U \rightarrow W$ mit $\bar{\varphi} \circ q_U = \varphi$, so ist

$$\ker \varphi \supseteq \ker q_U = U.$$

Gilt umgekehrt $U \subseteq \ker \varphi$, so folgt aus $x + U = y + U$ die Beziehung $\varphi(x) = \varphi(y)$. Daher wird durch

$$\bar{\varphi}(x + U) := \varphi(x)$$

eine Abbildung $\bar{\varphi}: V/U \rightarrow W$ definiert. Wir müssen nur noch verifizieren, dass $\bar{\varphi}$ linear ist. Das ergibt sich aus folgenden Rechnungen für $x, y \in V$ und $\lambda \in \mathbb{K}$:

$$\bar{\varphi}((x + U) + (y + U)) = \bar{\varphi}(x + y + U) = \varphi(x + y) = \varphi(x) + \varphi(y) = \bar{\varphi}(x + U) + \bar{\varphi}(y + U)$$

sowie

$$\bar{\varphi}(\lambda(x + U)) = \bar{\varphi}(\lambda x + U) = \varphi(\lambda x) = \lambda \varphi(x) = \lambda \bar{\varphi}(x + U).$$

Für $x \in V$ ist $\bar{\varphi}(x + U) = \varphi(x) = 0$ äquivalent zu $x \in \ker \varphi$. Daher ist

$$\ker(\bar{\varphi}) = q_U(\ker \varphi) = \ker \varphi / U.$$

Stimmt $\ker \varphi$ mit U überein, so sehen wir, dass $\bar{\varphi}$ injektiv ist. □

Bemerkung 8.1.10. (Kanonische Faktorisierung) Ist $\varphi: V \rightarrow W$ gegeben, so setzen wir $U := \ker \varphi$ und erhalten mit dem Homomorphiesatz eine injektive lineare Abbildung $\bar{\varphi}: V/U \rightarrow W$ mit $\bar{\varphi} \circ q_U = \varphi$. Wir haben die Abbildung φ damit als Komposition einer injektiven und einer surjektiven Abbildung geschrieben. Jede lineare Abbildung φ lässt sich also als Komposition einer injektiven und einer surjektiven Abbildung darstellen.

Bemerkung 8.1.11. (Quotienten und Komplemente) In vielen Situationen liegt für den Unterraum $U \subseteq V$ ein Komplement vor, d.h. $V = U \oplus W$ für einen Unterraum W . Schränken wir die Quotientenabbildung $q_U: V \rightarrow V/U$ auf W ein, so erhalten wir wegen $W \cap U = W \cap \ker q_U = \ker(q_U|_W) = \{0\}$ eine injektive Abbildung

$$q_U|_W: W \rightarrow V/U.$$

Andererseits folgt aus $V = U + W$, dass jede Nebenklasse $x + U$ einen Repräsentanten $x' \in W$ besitzt, also $x + U = x' + U$ für ein $x' \in W$. Daher ist $q_U|_W: W \rightarrow V/U$ auch surjektiv, mithin bijektiv. Wir erhalten daher einen linearen Isomorphismus $W \cong V/U$.

Mit dem Begriff der Kodimension können wir insbesondere die Dimensionsformel für lineare Abbildungen etwas verallgemeinern. Im folgenden Satz müssen wir nicht voraussetzen, dass V endlichdimensional ist. Für $\dim V = \infty$ enthält die bekannte Dimensionsformel

$$\dim V = \dim(\operatorname{im}(\varphi)) + \dim(\ker(\varphi))$$

keine Information, da beide Seiten unendlich sind.

Satz 8.1.12. *Sei $\varphi: V \rightarrow W$ eine lineare Abbildung, für die $\operatorname{im}(\varphi) \subseteq W$ endlichdimensional ist. Dann ist*

$$\operatorname{codim}(\ker \varphi) = \dim(\operatorname{im}(\varphi)).$$

Beweis. Aus dem Homomorphiesatz wissen wir, dass eine bijektive lineare Abbildung $\bar{\varphi}: V/\ker \varphi \rightarrow \operatorname{im} \varphi$ existiert. Da diese Abbildung ein linearer Isomorphismus ist, sind die Räume $V/\ker \varphi$ und $\operatorname{im} \varphi$ isomorph, haben also die gleiche Dimension. $\square$

Beispiel 8.1.13. (Unbestimmte Integrale) Sei $[a, b] \subseteq \mathbb{R}$ ein Intervall und $C([a, b])$ der Raum der stetigen Funktionen $f: [a, b] \rightarrow \mathbb{R}$. In der Analysis lernt man, dass jede stetige Funktion f eine Stammfunktion $F: [a, b] \rightarrow \mathbb{R}$ besitzt, d.h. F ist differenzierbar mit $F' = f$. Allerdings sind solche Stammfunktionen nicht eindeutig bestimmt. Für jede Wahl von $x_0 \in [a, b]$ erhält man z.B. eine durch das Riemann-Integral

$$F_{x_0}(x) := \int_{x_0}^x f(t) dt$$

(Hauptsatz der Differential- und Integralrechnung). Ist F eine Stammfunktion von f , so kann man mit ihr Integrale berechnen:

$$\int_a^b f(t) dt = F(b) - F(a).$$

Dies führt zur etwas unklaren Notation

$$F(x) = \int f(x) dx \quad \text{oder} \quad F(x) = \int f(x) dx + C$$

der *unbestimmten Integrale*.

Die Problematik dieser Notation lässt sich dadurch beheben, dass man die Stammfunktion F als Element des Quotientenraums $C([a, b])/\mathbb{R}1$ auffasst, aus dem man die konstanten Funktionen faktorisiert hat. Als Element dieses Quotientenraums sind also Ausdrücke der Gestalt $\int f(x) dx$ wieder sinnvoll, aber nicht als Funktionen $[a, b] \rightarrow \mathbb{R}$. Die Wahl einer konkreten Stammfunktion F von f entspricht also der Wahl eines Elements des zugehörigen eindimensionalen affinen Unterraums $F + \mathbb{R}1 \subseteq C([a, b])$.

Beispiel 8.1.14. (Riemann integrierbare Funktionen) In der Analysis 1 lernt man den Raum $\mathcal{R}$ der auf einem Intervall $[a, b] \subseteq \mathbb{R}$ Riemann integrierbaren Funktionen $f: [a, b] \rightarrow \mathbb{R}$ kennen. Auf diesem Raum möchte man durch

$$\|f\|_1 := \int_a^b |f(x)| dx$$

eine Norm definieren, muss aber feststellen, dass man hierdurch lediglich eine Halbnorm erhält. D.h. $\|\cdot\|_1$ genügt zwar den Relationen

$$\|\lambda f\|_1 = |\lambda| \cdot \|f\|_1 \quad \text{und} \quad \|f + g\|_1 \leq \|f\|_1 + \|g\|_1, \quad (8.2)$$

aber es gibt Funktionen $0 \neq f$ mit $\|f\|_1 = 0$.

In dieser Situation kann die Konstruktion eines geeigneten Quotientenraums Abhilfe schaffen. Man betrachtet hierzu die Teilmenge

$$U := \{f \in \mathcal{R} : \|f\|_1 = 0\}$$

und erhält aus (8.2) sofort, dass $U \subseteq \mathcal{R}$ ein linearer Unterraum ist. Der Raum U besteht in einem gewissen Sinn aus den “irrelevanten” Funktionen und man möchte nicht zwischen Funktionen unterscheiden, die sich nur um ein Element aus U unterscheiden. Typische Beispiele von Elementen aus U sind Funktionen, die nur an endlich vielen Stellen einen Wert $\neq 0$ annehmen. Die eigentlichen Objekte der Riemannschen Integrationstheorie sind also die Elemente $\bar{f}$ des Quotientenraums $\mathcal{R}/U$. Auf diesem Raum erhält man durch

$$\|\bar{f}\|_1 := \|f\|_1, \quad f \in \mathcal{R},$$

eine wohldefinierte Norm und damit einen für die Analysis natürlicheren Kontext.

Aufgaben zu Abschnitt 8.1

Aufgabe 8.1.1. Sei $U \subseteq V$ ein linearer Unterraum. Zeigen Sie:

- (a) Ist $\sigma: V/U \rightarrow V$ eine lineare Abbildung mit $q_U \circ \sigma = \text{id}_{V/U}$, so ist $\text{im}(\sigma)$ ein Komplement von U , d.h. $V = U \oplus \text{im}(\sigma)$.
- (b) Ist $V = U \oplus W$ für einen Unterraum $W \subseteq V$, so existiert genau eine lineare Abbildung $\sigma: V/U \rightarrow V$ mit $q_U \circ \sigma = \text{id}_{V/U}$ und $\text{im}(\sigma) \subseteq W$.

Aufgabe 8.1.2. Sei $0 \neq f \in \mathbb{K}[X]$ ein Polynom und $(f) = f \cdot \mathbb{K}[X]$ das von f erzeugte Hauptideal. Wir betrachten den Quotientenraum $\mathcal{A} := \mathbb{K}[X]/(f)$ und schreiben $\bar{p} := p + (f)$ für seine Elemente. Zeigen Sie:

- (a) $\dim(\mathcal{A}) = \text{grad}(f)$. Insbesondere ist $\mathcal{A}$ endlichdimensional. Hinweis: Ist $d = \text{grad}(f)$, so ist $\bar{1}, \bar{X}, \dots, \bar{X}^{d-1}$ eine Basis von $\mathcal{A}$.
- (b) Durch $\bar{p} \cdot \bar{q} := \overline{pq}$ erhalten wir auf $\mathcal{A}$ eine wohldefinierte bilineare Multiplikation. Diese Multiplikation ist kommutativ, assoziativ und $\bar{1}$ (das Bild des konstanten Polynoms) ist ein Einselement.
- (c) $\bar{p} = 0$ ist äquivalent dazu, dass f ein Teiler von p ist.
- (d) $\bar{p} \in \mathcal{A}$ ist genau dann invertierbar, wenn p teilerfremd zu f ist, d.h. $\text{ggT}(p, f) = 1$. Für $qp + hf = 1$ ist dann $\bar{p}^{-1} = \bar{q}$.
- (e) $\mathcal{A}$ ist genau dann ein Körper, wenn f *prim* ist, also keine nicht konstanten Teiler besitzt.
- (f) Ist $f(X) = a_0 + a_1X + \dots + a_dX^d$, so genügt das Element $\bar{X} \in \mathcal{A}$ genügt der Relation

$$0 = f(\bar{X}) = a_0 + a_1\bar{X} + \dots + a_d\bar{X}^d.$$

- (g) Ist $\mathbb{K} = \mathbb{R}$ und $f(X) = X^2 + aX + b$ ohne reelle Nullstellen, so ist $\mathcal{A}$ ein Körper, der zu $\mathbb{C}$ isomorph ist. Hinweis: Betrachte eine reell lineare Abbildung $\varphi: \mathcal{A} \rightarrow \mathbb{C}$, die $\bar{X}$ auf eine komplexe Nullstelle von f abbildet.

Aufgabe 8.1.3. Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und V ein $\mathbb{K}$ -Vektorraum. Eine *Halbnorm* auf V ist eine Funktion

$$p: V \rightarrow [0, \infty[$$

mit den Eigenschaften

$$(HN1) \quad p(\lambda x) = |\lambda|p(x) \text{ für } x \in V, \lambda \in \mathbb{K}.$$

$$(HN2) \quad p(x + y) \leq p(x) + p(y) \text{ für } x, y \in V.$$

Zeigen Sie: Für jede Halbnorm p auf V ist $U_p := \{x \in V: p(x) = 0\}$ ein linearer Unterraum und durch $\|\bar{x}\| := p(x)$ wird auf dem Quotientenraum V/U_p eine Norm definiert.

Aufgabe 8.1.4. Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und V ein $\mathbb{K}$ -Vektorraum sowie $\beta: V \times V \rightarrow \mathbb{K}$ eine positiv semidefinite hermitesche Form. Zeigen Sie:

- (a) $p(v) := \sqrt{\beta(v, v)}$ definiert eine Halbnorm auf V . Hinweis: Cauchy–Schwarzsche Ungleichung.
- (b) $p(v) = 0$ ist äquivalent zu $\beta(v, w) = 0$ für alle $w \in V$.
- (c) Auf dem Quotientenraum V/U_p wird durch

$$\langle \bar{x}, \bar{y} \rangle := \beta(x, y), \quad x, y \in V$$

eine wohldefinierte positiv definite hermitesche Form definiert. Hierdurch wird V/U_p zu einem unitären $\mathbb{K}$ -Vektorraum.

- (d) Ist $\mathcal{R}$ der Raum der Riemann-integriblen Funktionen auf dem Intervall $[a, b]$, so ist

$$\beta(f, g) := \int_a^b f(x)g(x) dx$$

eine positiv semidefinite symmetrische Bilinearform auf $\mathcal{R}$ mit

$$\int_a^b |f(x)| dx = 0 \iff \int_a^b f(x)^2 dx = 0.$$

Aufgabe 8.1.5. Ist $\varphi: V \rightarrow W$ eine surjektive lineare Abbildung und $U \subseteq W$ ein Unterraum endlicher Kodimension, so ist

$$\text{codim}(\varphi^{-1}(U)) = \text{codim}(U).$$

Aufgabe 8.1.6. Ist $U \subseteq V$ ein linearer Unterraum der Kodimension k , so existieren k Hyperebenen $H_1, \dots, H_k$ mit $U = H_1 \cap \dots \cap H_k$. Hinweis: $V/U \cong \mathbb{K}^k$ und Aufgabe 8.1.5.

Aufgabe 8.1.7. Sind $U_1, U_2 \subseteq V$ lineare Unterräume, so ist

$$(U_1 + U_2)/U_2 \cong U_1/(U_1 \cap U_2).$$

Aufgabe 8.1.8. Sei V ein endlichdimensionaler euklidischer Vektorraum und $U \subseteq V$ ein linearer Unterraum. Wir definieren den Abstand zweier affiner Unterräume $x + U$ und $y + U'$ durch

$$\text{dist}(x + U, y + U') := \inf\{\|v - w\| : v \in x + U, w \in y + U'\}.$$

Zeigen Sie

$$\text{dist}(x + U, y + U') = \|p(x - y)\|,$$

wobei $p: V \rightarrow (U + U')^\perp$ die Orthogonalprojektion mit $\ker p = U + U'$ ist. Hinweis: Welche Minimaleigenschaft besitzt die Orthogonalprojektion auf einen Unterraum?

8.2 Der Dualraum

In diesem Abschnitt betrachten wir den Raum $V^* := \text{Hom}(V, \mathbb{K})$ der linearen Abbildungen von einem Vektorraum V in der Grundkörper etwas genauer. Hierbei wird sich zeigen, dass es oft zweckmäßig ist, für viele Problemstellungen der linearen Algebra eine “duale Perspektive” zu entwickeln, da man so zusätzliche Einsichten erhält.

8.2.1 Lineare Funktionale und duale Basen

Definition 8.2.1. Sei V ein $\mathbb{K}$ -Vektorraum. Eine lineare Abbildung $\varphi: V \rightarrow \mathbb{K}$ heißt *Linearform* oder *lineares Funktional*. Der Vektorraum $V^* := \text{Hom}(V, \mathbb{K})$ aller Linearformen heißt *Dualraum* von V .

Bemerkung 8.2.2. (a) Für $\varphi \in V^*$ ist $\varphi(V) = \text{im}(\varphi)$ ein linearer Unterraum von $\mathbb{K}$, also entweder $\{0\}$ oder $\mathbb{K}$. Ist $\varphi \neq 0$, so ist $\varphi(V) = \mathbb{K}$ und $\ker \varphi$ daher eine Hyperebene in V .

(b) Ist $H \subseteq V$ eine Hyperebene, so ist $\dim(V/H) = 1$, also $V/H \cong \mathbb{K}$. Ist $\varphi: V/H \rightarrow \mathbb{K}$ ein linearer Isomorphismus und $q_H: V \rightarrow V/H$ die Quotientenabbildung, so erhalten wir eine Linearform $\alpha := \varphi \circ q_H: V \rightarrow \mathbb{K}$ mit $\ker \alpha = H$. Insbesondere ist jeder Hyperbene Kern einer Linearform.

Wir schauen uns nun ein paar Beispiele von Dualräumen an.

Beispiel 8.2.3. (a) Auf $V = \mathbb{K}^n$ hat jede Linearform $\varphi \in V^*$ die Gestalt

$$\varphi(x) = \sum_{j=1}^n a_j x_j \quad \text{mit} \quad a_j = \varphi(e_j).$$

Dann ist der Zeilenvektor $a := (a_1, \dots, a_n) \in M_{1,n}(\mathbb{K})$ die Matrix der linearen Abbildung φ bzgl. der Standardbasis.

Wir können φ also darstellen als Produkt des Typs Zeile mal Spalte:

$$\varphi(x) = a \cdot x.$$

In diesem Sinn können wir den Dualraum $(\mathbb{K}^n)^*$ mit dem Raum der Zeilenvektoren identifizieren, aber das ist nur ein Bild, das wir uns von seinen Elementen machen.

Der Koordinatenprojektion

$$\pi_j: \mathbb{K}^n \rightarrow \mathbb{K}, \quad x \mapsto x_j = e_j^\top x$$

entspricht hierbei der Zeilenvektor $e_j^\top$.

(b) Auf dem Raum $V = C([a, b], \mathbb{R})$, der auf dem Intervall $[a, b] \subseteq \mathbb{R}$ stetigen Funktionen ist das Riemann-Integral

$$I: C([a, b], \mathbb{R}) \rightarrow \mathbb{R}, \quad I(f) := \int_a^b f(x) dx$$

ein lineares Funktional, denn es gilt

$$I(f_1 + f_2) = I(f_1) + I(f_2) \quad \text{und} \quad I(\lambda f) = \lambda I(f) \quad \text{für} \quad f_1, f_2, f \in V, \lambda \in \mathbb{K}.$$

Weitere Funktionale auf diesem Raum erhalten wir durch Auswertung in Punkten des Intervalls:

$$\delta_p(f) := f(p) \quad \text{für} \quad p \in [a, b].$$

Man kann sich das Integral I als eine Art “kontinuierliche Summe” dieser Funktionale vorstellen.

Definition 8.2.4. (Duale Basis) Ist V ein Vektorraum und $(v_j)_{j \in J}$ eine Basis (endlich oder unendlich), so enthält der Dualraum V^* lineare Funktionale v_j^* , die eindeutig charakterisiert sind durch

$$v_j^*(v_i) = \delta_{ij}.$$

Auf einem Vektor $v = \sum_{k \in J} x_k v_k$ haben sie den Wert

$$v_j^*(v) = \sum_{k \in J} x_k v_j^*(v_k) = \sum_{k \in J} x_k \delta_{jk} = x_j.$$

Hieraus ergibt sich die Basisdarstellung

$$v = \sum_{j \in J} v_j^*(v) v_j.$$

Das Funktional v_j^* liefert zu v also genau den Koeffizienten x_j von v bzgl. der Darstellung von v als Linearkombination der Basiselemente.

Ist J endlich, so heißt $\{v_j^*: j \in J\}$ die *duale Basis von V^** (vgl. Lemma 8.2.5).

Ist J unendlich, so ist die Menge $(v_j^*)_{j \in J}$ zwar linear unabhängig (siehe Beweis von Lemma 8.2.5), aber keine Basis von V^* . In der Tat, wissen wir, dass es eine Linearform $\alpha: V \rightarrow \mathbb{K}$ gibt, die auf allen Basiswerten v_j den Wert 1 annimmt: $\alpha(\sum_j x_j v_j) := \sum_j x_j$. Diese Linearform ist aber keine Linearkombination der v_j^* .

Lemma 8.2.5. *Ist $v_1, \dots, v_n$ eine Basis von V , so ist $v_1^*, \dots, v_n^*$ eine Basis von V^* . Für $\alpha \in V^*$ ist die Basisdarstellung gegeben durch*

$$\alpha = \sum_{j=1}^n \alpha(v_j) v_j^*.$$

Beweis. Aus der Zuordnung von Matrizen zu linearen Abbildungen kennen wir schon die Dimension von V^* :

$$\dim V^* = \dim \operatorname{Hom}(V, \mathbb{K}) = \dim M_{1,n}(\mathbb{K}) = n.$$

Es reicht daher einzusehen, dass die n Elemente v_j^* den Raum V^* erzeugen. Hierzu verifizieren wir die angegebene Basisdarstellung. Für $\alpha \in V^*$ erhalten wir für $k = 1, \dots, n$ die Relation

$$\left(\sum_{j=1}^n \alpha(v_j) v_j^* \right) (v_k) = \sum_{j=1}^n \alpha(v_j) v_j^*(v_k) = \sum_{j=1}^n \alpha(v_j) \delta_{jk} = \alpha(v_k).$$

Da die Funktionale α und $\sum_{j=1}^n \alpha(v_j) v_j^*$ auf allen Basiselementen den gleichen Wert annehmen, sind sie gleich. $\square$

Beispiel 8.2.6. Im Raum $\mathbb{R}[X]$ der Polynomfunktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ bilden die Monome $v_n = X^n$, $n \in \mathbb{N}_0$, eine Basis. Für die dualen Funktionale ist dann $v_n^*(X^m) = \delta_{nm}$, also

$$v_n^*(a_0 + a_1 X + \dots + a_k X^k) = a_n.$$

Wir haben dann

$$v_n^*(f) = \frac{1}{n!} f^{(n)}(0),$$

wobei $f^{(n)}$ die n -Ableitung des Polynoms f ist. Die Rekonstruktion des Polynoms f vom Grad d aus diesen Zahlen nimmt dann die Gestalt

$$f(x) = \sum_{n=0}^d v_n^*(f) x^n = \sum_{n=0}^d \frac{f^{(n)}(0)}{n!} x^n,$$

was genau der Taylorentwicklung von f im Nullpunkt entspricht.

Ist $x_0 \in \mathbb{R}$ beliebig, so bilden die Polynome $(X - x_0)^n$, $n \in \mathbb{N}_0$, ebenfalls eine Basis von $\mathbb{R}[X]$ und wir erhalten entsprechend die Formel für die Taylorentwicklung in x_0 :

$$f(x) = \sum_{n=0}^d \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n.$$

Beispiel 8.2.7. Sei V ein unitärer $\mathbb{K}$ -Vektorraum für $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Dann erhalten wir durch $\alpha_v(w) := \langle w, v \rangle$ ein lineares Funktional auf V , denn das Skalarprodukt ist linear im ersten Argument. Der Darstellungssatz von Fréchet–Riesz 7.2.20 besagt, dass für $\dim V < \infty$ jedes linear Funktional auf V diese Gestalt hat. Wir erhalten so eine antilineare Bijektion

$$\Gamma: V \rightarrow V^*, \quad \Gamma(v) = \alpha_v.$$

Ist $v_1, \dots, v_n \in V$ eine ONB, so erhalten wir für die Basisdarstellung eines Elements v die Form

$$v = \sum_{j=1}^n v_j^*(v) v_j = \sum_{j=1}^n \langle v, v_j \rangle v_j.$$

Insbesondere ist in diesem Fall

$$v_j^*(v) = \langle v, v_j \rangle.$$

8.2.2 Die duale Abbildung

Definition 8.2.8. (Die adjungierte/duale Abbildung) Ist $\varphi: V \rightarrow W$ eine lineare Abbildung, so erhalten wir durch

$$\varphi^*: W^* \rightarrow V^*, \quad \varphi^*(\alpha) := \alpha \circ \varphi$$

eine lineare Abbildung (Nachweis als Übung). Diese lineare Abbildung heißt die zu φ *adjungierte* oder zu φ *duale* Abbildung.

Bemerkung 8.2.9. Aus der Definition ergeben sich unmittelbar folgende Regeln für die duale Abbildung

$$\text{id}_V^* = \text{id}_{V^*} \quad \text{und} \quad (\varphi \circ \psi)^* = \psi^* \circ \varphi^*$$

für die Komposition zweier linearer Abbildungen $\psi: U \rightarrow V$ und $\varphi: V \rightarrow W$. In der Tat ist

$$(\varphi \circ \psi)^*(\alpha) = \alpha \circ \varphi \circ \psi = (\alpha \circ \varphi) \circ \psi = \psi^*(\alpha \circ \varphi) = (\psi^* \circ \varphi^*)(\alpha).$$

Es ist instruktiv zu sehen, wie die duale Abbildung auf der Ebene der Matrizen aussieht.

Satz 8.2.10. Sei $B := (b_1, \dots, b_n)$ eine Basis des $\mathbb{K}$ -Vektorraums V , $B^* \subseteq V^*$ die duale Basis sowie $C := (c_1, \dots, c_m)$ eine Basis des $\mathbb{K}$ -Vektorraums W und C^* die duale Basis von W^* . Ist $\varphi: V \rightarrow W$ eine lineare Abbildung und $[\varphi]_B^C \in M_{m,n}(\mathbb{K})$ die Matrix von φ bzgl. B und C , so ist

$$[\varphi^*]_{C^*}^{B^*} = ([\varphi]_B^C)^\top \in M_{n,m}(\mathbb{K})$$

die transponierte Matrix.

Vereinfacht gesagt gilt also: Die Matrix der adjungierten Abbildung erhält man durch Transponieren.

Beweis. Sei $A = (a_{ij}) = [\varphi]_B^C$. Für $c_j^* \in W^*$ erhalten wir

$$\varphi^*(c_j^*)(b_i) = (c_j^* \circ \varphi)(b_i) = c_j^*(\varphi(b_i)) = c_j^*\left(\sum_{k=1}^n a_{ki} c_k\right) = \sum_{k=1}^n a_{ki} \delta_{jk} = a_{ji}.$$

Also ist

$$\varphi^*(c_j^*) = \sum_{i=1}^m a_{ji} b_i^*$$

und somit $[\varphi^*]_{C^*}^{B^*} = A^\top$. □

8.2.3 Der Bidualraum

Definition 8.2.11. Sei V ein $\mathbb{K}$ -Vektorraum. Jedes Element $v \in V$ definiert durch Auswertung in v ein lineares Funktional $V^* \rightarrow \mathbb{K}, \alpha \mapsto \alpha(v)$. Wir erhalten so eine lineare Abbildung

$$\eta_V: V \rightarrow (V^*)^*.$$

(Nachweis der Linearität als Übung!)

Satz 8.2.12. Ist $\dim V < \infty$, so ist $\eta_V: V \rightarrow (V^*)^*$ ein Isomorphismus von $\mathbb{K}$ -Vektorräumen.

Beweis. Wir zeigen zuerst, dass η_V injektiv ist. Sei dazu $0 \neq v \in V$. Dann existiert eine Basis $v_1 = v, v_2, \dots, v_n$ in V . Für die duale Basis $v_1^*, \dots, v_n^*$ ist dann

$$\eta(v)(v_1^*) = v_1^*(v) = v_1^*(v_1) = 1.$$

Also ist $\eta(v) \neq 0$.

Wir wissen schon, dass $\dim(V^*)^* = \dim V^* = \dim V = n$ ist. Daher ist die lineare Abbildung η_V auch surjektiv, mithin bijektiv. $\square$

Bemerkung 8.2.13. Ist V endlichdimensional und $n = \dim V$, so ist auch $\dim V^* = n$, also $V \cong \mathbb{K}^n \cong V^*$. Es gibt also auch einen linearen Isomorphismus $\zeta: V \rightarrow V^*$. Der wesentliche Unterschied zu dem Isomorphismus $\eta_V: V \rightarrow (V^*)^*$ ist, dass es keinen *natürlichen* Isomorphismus $\zeta_V: V \rightarrow V^*$. Das ist ein Begriff aus der Kategorientheorie.

Der Isomorphismus η_V ist “natürlich”, weil für jede lineare Abbildung $\varphi: V \rightarrow W$ die Beziehung $\eta_W \circ \varphi = \varphi^{**} \circ \eta_V$ gilt, also das Diagramm

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ \downarrow \eta_V & & \downarrow \eta_W \\ V^{**} & \xrightarrow{\varphi^{**}} & W^{**} \end{array}$$

kommutiert.

Gäbe es für jeden endlichdimensionalen Vektorraum V einen in diesem Sinne natürlichen Isomorphismus $\zeta_V: V \rightarrow V^*$, so würden wir für jede lineare Abbildung $\varphi: V \rightarrow W$ die Kommutativität des Diagramms

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ \downarrow \zeta_V & & \downarrow \zeta_W \\ V^* & \xleftarrow{\varphi^*} & W^* \end{array}$$

erwarten. Man beachte die Richtung des unteren Pfeils. Für $\varphi = 0$ würde sich damit aber sofort $\zeta_V = \varphi^* \circ \eta_W \circ \varphi = 0$ ergeben, im Widerspruch dazu, dass ζ_V ein Isomorphismus sein soll.

8.2.4 Annulatoren

Definition 8.2.14. Sei V ein $\mathbb{K}$ -Vektorraum. Für eine Teilmenge $A \subseteq V$ definieren wir den *Annulator in V^**

$$A^\perp := \{\alpha \in V^* : \alpha(A) = \{0\}\} = \{\alpha \in V^* : (\forall v \in A) \alpha(v) = 0\} \subseteq V^*$$

und für eine Teilmenge $A \subseteq V^*$ definieren wir analog

$$A^\perp := \{v \in V : (\forall \beta \in A) \beta(v) = 0\} \subseteq V.$$

Lemma 8.2.15. Für Teilmengen $A \subseteq V$ oder $A \subseteq V^*$ gilt:

- (i) $A^\perp$ ist ein linearer Unterraum.
- (ii) $A_1 \subseteq A_2 \Rightarrow A_2^\perp \subseteq A_1^\perp$.
- (iii) $A_1 \subseteq A_2^\perp \Leftrightarrow A_2 \subseteq A_1^\perp$ für $A_1 \subseteq V$ und $A_2 \subseteq V^*$.
- (iv) $A \subseteq A^{\perp\perp}$ und $A^\perp = A^{\perp\perp\perp}$.
- (v) $A^\perp = \text{span}(A)^\perp$.
- (vi) Ist V endlichdimensional, so haben wir zusätzlich:
 - (a) $\text{codim}(A^\perp) = \dim(\text{span}(A))$.
 - (b) $A^{\perp\perp} = \text{span}(A)$.

Beweis. (i) Dass $A^\perp$ ein linearer Unterraum ist, folgt für $A \subseteq V$ aus

$$(\lambda_1 \alpha_1 + \lambda_2 \alpha_2)(a) = \lambda_1 \alpha_1(a) + \lambda_2 \alpha_2(a) = 0 \quad \text{für} \quad a \in A, \alpha_1, \alpha_2 \in A^\perp, \lambda_1, \lambda_2 \in \mathbb{K},$$

und für $A \subseteq V^*$ aus

$$\alpha(\lambda_1 a_1 + \lambda_2 a_2) = \lambda_1 \alpha(a_1) + \lambda_2 \alpha(a_2) = 0 \quad \text{für} \quad a_1, a_2 \in A^\perp, \alpha \in A, \lambda_1, \lambda_2 \in \mathbb{K}.$$

(ii) folgt unmittelbar aus der Definition.

(iii) $A_1 \subseteq A_2^\perp$ bedeutet, dass $\alpha(v) = 0$ für alle $\alpha \in A_2$ und $v \in A_1$ gilt. Dies ist aber auch äquivalent zu $A_2 \subseteq A_1^\perp$.

(iv) Aus (iii) und $A^\perp \subseteq A^\perp$ folgt $A \subseteq A^{\perp\perp}$. Hieraus erhalten wir mit (ii) $A^{\perp\perp\perp} \subseteq A^\perp$. Mit (iii) ergibt sich aber auch die umgekehrte Inklusion $A^\perp \subseteq A^{\perp\perp\perp}$ aus $A^{\perp\perp} = A^{\perp\perp}$. Also gilt hier Gleichheit.

(v) Da $A^{\perp\perp} \supseteq A$ ein linearer Unterraum ist, gilt auch $\text{span}(A) \subseteq A^{\perp\perp}$, also $A^\perp \subseteq \text{span}(A)^\perp$. Die umgekehrte Inklusion folgt unmittelbar aus (ii), so dass wir die Gleichheit erhalten.

(vi)(a) Da $V \cong (V^*)^*$ ist (Satz 8.2.12), reicht es die Behauptung für den Fall $A \subseteq V$ zu zeigen. Der Fall $A \subseteq V^*$ folgt dann durch Anwendung dieses Falles auf den Raum V^* anstelle von V .

Sei $A \subseteq V$. Wir wählen eine Basis $b_1, \dots, b_k$ von A , die wir durch $b_{k+1}, \dots, b_n$ zu einer Basis von V ergänzen. Dann ist $b_1^*, \dots, b_n^*$ eine Basis von V^* und eine Linearkombination $\alpha = \sum_{j=1}^n \alpha_j b_j^*$ verschwindet genau dann auf A , wenn für $j = 1, \dots, k$ die Beziehung

$$0 = \alpha(b_j) = \alpha_j$$

gilt. Also ist $A^\perp = \text{span}\{b_{k+1}^*, \dots, b_n^*\}$ und somit $\text{codim}(A^\perp) = k = \dim(A)$.

(b) Wegen (v) dürfen wir $A = \text{span}(A)$ annehmen, d.h., dass A ein linearer Unterraum ist. Wegen $A \subseteq A^{\perp\perp}$ reicht es dann zu zeigen, dass die Dimensionen dieser Unterräume gleich sind. Dies ergibt sich mit (a) für $A \subseteq V$ aus

$$\dim(A) = \text{codim}(A^\perp) = \dim V^* - \dim(A^\perp) = \dim V - \text{codim}(A^{\perp\perp}) = \dim(A^{\perp\perp})$$

und vollkommen analog für $A \subseteq V^*$. □

Satz 8.2.16. Für die duale Abbildung $\varphi^*: W^* \rightarrow V^*$ von $\varphi: V \rightarrow W$ ist

$$\ker(\varphi^*) = \text{im}(\varphi)^\perp. \quad (8.3)$$

Insbesondere ist φ^* injektiv, wenn φ surjektiv ist.

Ist $\dim V < \infty$, so gilt auch

$$\text{im}(\varphi) = \ker(\varphi^*)^\perp. \quad (8.4)$$

Beweis. Für $\alpha \in W^*$ ist die Relation $0 = \varphi^*(\alpha) = \alpha \circ \varphi$ äquivalent zu $\alpha|_{\varphi(V)} = 0$. Hieraus folgt (8.3). Ist V endlichdimensional, so ergibt sich damit wegen Lemma 8.2.15(vi)

$$\text{im}(\varphi) = (\text{im}(\varphi)^\perp)^\perp = \ker(\varphi^*)^\perp. \quad \square$$

Bemerkung 8.2.17. Mittels der dualen Abbildungen erhalten wir einen sehr natürlichen begrifflichen Beweis des Satzes, dass für jede Matrix Zeilen- und Spaltenrang übereinstimmen.

Sei hierzu $\varphi: V \rightarrow W$ eine lineare Abbildung und $\varphi^*: W^* \rightarrow V^*$ die duale Abbildung. Dann ist

$$\text{rank}(\varphi^*) = \dim(\text{im}(\varphi^*)) = \text{codim}(\ker \varphi^*) = \text{codim}(\text{im}(\varphi)^\perp) = \dim \text{im}(\varphi) = \text{rank}(\varphi).$$

Also haben die linearen Abbildungen φ und φ^* den gleichen Rang.

Aus Satz 8.2.10 ergibt sich damit für jede Matrix $A \in M_{m,n}(\mathbb{K})$ die Relation

$$\text{rank}(A) = \text{rank}(A^\top)$$

bzw. "Zeilenrang = Spaltenrang".

Satz 8.2.18. Ist $U \subseteq V$ ein linearer Unterraum und $q_U: V \rightarrow V/U$ die Quotientenabbildung, so ist $q_U^*: (V/U)^* \rightarrow V^*$ eine injektive Abbildung mit Bild $U^\perp$. Insbesondere haben wir

$$(V/U)^* \cong U^\perp.$$

Hat U endliche Kodimension, so ergibt sich hieraus insbesondere

$$\text{codim}(U) = \dim(U^\perp).$$

Beweis. Da q_U surjektiv ist, folgt aus Satz 8.2.16, dass q_U^* injektiv ist. Weiter folgt aus dem Homomorphiesatz 8.1.9, dass zu einem linearen Funktional $\varphi: V \rightarrow \mathbb{K}$ genau dann ein lineares Funktional $\bar{\varphi}: V/U \rightarrow \mathbb{K}$ existiert mit $\varphi = \bar{\varphi} \circ q_U = q_U^*(\bar{\varphi})$, wenn $U \subseteq \ker \varphi$ gilt, d.h., $\varphi|_U = 0$ ist. Hieraus folgt $\text{im}(q_U^*) = U^\perp$ und damit $(V/U)^* \cong U^\perp$.

Hat U endliche Kodimension, so erhalten wir damit

$$\text{codim } U = \dim(V/U) = \dim((V/U)^*) = \dim(U^\perp). \quad \square$$

Aufgaben zu Abschnitt 8.2

Aufgabe 8.2.1. Sei $(v_j)_{j \in J}$ eine Basis des $\mathbb{K}$ -Vektorraums V . Wir schreiben $\mathbb{K}^J$ für den Raum aller Funktionen $J \rightarrow \mathbb{K}$ und $\mathbb{K}^{(J)} \subseteq \mathbb{K}^J$ für den Teilraum, der aus denjenigen Funktionen besteht, die nur an endlich vielen Stellen von 0 verschieden sind. Zeigen Sie:

(a) Die Abbildung

$$\Gamma: \mathbb{K}^{(J)} \rightarrow V, \quad \Gamma(f) := \sum_{j \in J} f(j)v_j$$

ist ein Isomorphismus von $\mathbb{K}$ -Vektorräumen.

(b) Die Abbildung

$$\Gamma^*: V^* \rightarrow \mathbb{K}^J, \quad \Gamma^*(\alpha)(j) := \alpha(v_j)$$

ist ein Isomorphismus von $\mathbb{K}$ -Vektorräumen. Für $V = \mathbb{K}^{(J)}$ erhalten wir also insbesondere $(\mathbb{K}^{(J)})^* \cong \mathbb{K}^J$. Schließen Sie hieraus, dass V^* nicht von den Elementen v_j^* , $j \in J$, erzeugt wird wenn J unendlich ist.

Aufgabe 8.2.2. (Zum Unterschied von $\text{Hom}(\mathbb{K}, V)$ und $\text{Hom}(V, \mathbb{K})$) Sei V ein $\mathbb{K}$ -Vektorraum. Dann ist

$$\text{ev}_1: \text{Hom}(\mathbb{K}, V) \rightarrow V, \quad \gamma \mapsto \gamma(1)$$

ein linearer Isomorphismus.

Aufgabe 8.2.3. Sei V ein endlichdimensionaler $\mathbb{K}$ -Vektorraum und $S \subseteq V^*$ eine Teilmenge. Wir sagen S trennt die Punkte von V , wenn zu jedem $0 \neq v \in V$ ein $\alpha \in S$ existiert mit $\alpha(v) \neq 0$. Zeigen Sie: Für $S \subseteq V^*$ sind äquivalent:

(a) S trennt die Punkte von V .

(b) $S^\perp = \{0\}$.

(c) $\text{span } S = V^*$.

Kapitel 9

Bilinearformen und Quadriken

In diesem Kapitel werden wir die Diagonalisierungssätze aus Kapitel 7 aus einer neuen Perspektive wiedersehen, indem wir sie auf quadratische Formen anwenden. Zuerst diskutieren wir Bilinearformen $V \times V \rightarrow \mathbb{K}$ über einem allgemeinen Körper und werden sehen, dass Bilinearformen sich, ähnlich wie lineare Abbildung, durch Matrizen darstellen lassen. Allerdings haben diese Matrizen hierbei eine vollkommen andere Interpretation. Symmetrische Bilinearformen führen uns dann auf die quadratischen Formen und damit zusammenhängende geometrische Fragen.

9.1 Bilinearformen und Sesquilinearformen

Sei V ein Vektorraum über einem beliebigen Körper $\mathbb{K}$. In diesem Abschnitt studieren wir Bilinearformen auf V , d.h. bilineare Abbildungen $F: V \times V \rightarrow \mathbb{K}$. Damit wir damit auch komplexe Skalarprodukte erfassen, tun wir das in einem etwas allgemeineren Rahmen, was auf den Begriff der α -Sesquilinearform führt, wobei $\alpha \in \text{Aut}(\mathbb{K})$ ein involutiver Automorphismus des Körpers $\mathbb{K}$ ist.

9.1.1 Die Matrix einer Sesquilinearform

In diesem Abschnitt ist $\mathbb{K}$ ein allgemeiner Körper sofern nichts anderes gesagt wird und $\alpha: \mathbb{K} \rightarrow \mathbb{K}$ ist ein *Automorphismus von $\mathbb{K}$* , d.h., eine bijektive Abbildung, die mit Addition und Skalarmultiplikation verträglich ist:

$$\alpha(x + y) = \alpha(x) + \alpha(y) \quad \text{und} \quad \alpha(xy) = \alpha(x)\alpha(y) \quad \text{für} \quad x, y \in \mathbb{K}.$$

Darüber hinaus nehmen wir an, dass α involutiv ist, d.h., $\alpha^2 = \text{id}_{\mathbb{K}}$ und wir schreiben wie im Falle $\mathbb{K} = \mathbb{C}$:

$$\bar{x} = \alpha(x) \quad \text{für} \quad x \in \mathbb{K}.$$

Beispiel 9.1.1. (a) Für jeden Körper $\mathbb{K}$ ist $\alpha = \text{id}_{\mathbb{K}}$ ein involutiver Automorphismus.

(b) Für $\mathbb{K} = \mathbb{C}$ kennen wir außer der Identität noch den Automorphismus $\alpha(z) = \bar{z}$, der durch die komplexe Konjugation $x + iy = x - iy$ für $x, y \in \mathbb{R}$ definiert ist. Man kann zeigen, dass

$$\text{Aut}_{\mathbb{R}}(\mathbb{C}) = \{\varphi \in \text{Aut}(\mathbb{C}) : \varphi|_{\mathbb{R}} = \text{id}_{\mathbb{R}}\} = \{1, \alpha\}$$

ist. Allerdings gibt es viele weitere Automorphismen, die den Unterkörper $\mathbb{R}$ nicht punktweise festlassen.

(c) Der Körper $\mathbb{R}$ besitzt nur den trivialen Automorphismus $\beta = \text{id}_{\mathbb{R}}$. Hier ist eine Beweisskizze: Jeder Automorphismus β lässt das Einselement 1 fest, denn $\beta(x) = \beta(1x) = \beta(1)\beta(x)$ gilt für $x \neq 0$. Hieraus folgt durch Induktion leicht $\beta(n) = n$ für alle $n \in \mathbb{N}$ und damit auch für $n \in \mathbb{Z}$. Für $m \in \mathbb{N}$ und $n \in \mathbb{Z}$ erhalten wir damit $\beta(\frac{n}{m}) = \frac{\beta(n)}{\beta(m)}$, also $\beta(x) = x$ für $x \in \mathbb{Q}$.

Ist $x \geq 0$, so existiert ein $y \in \mathbb{R}$ mit $x = y^2$. Also ist auch $\beta(x) = \beta(y)^2 \geq 0$. Damit sehen wir, dass aus $x \leq y$ auch $\beta(x) \leq \beta(y)$ folgt, d.h. β ist mit der Ordnung verträglich. Ist nun $x \in \mathbb{R}$ beliebig und $q < x$, so ist auch $q = \beta(q) < \beta(x)$ und für $q > x$ folgt ebenso $q > \beta(x)$. Damit ist

$$x = \inf\{q \in \mathbb{Q} : x < q\} = \inf\{q \in \mathbb{Q} : \beta(x) < q\} = \beta(x)$$

für alle $x \in \mathbb{R}$.

(d) Ist $\mathbb{K} = \mathbb{R}(X)$ der Körper der rationalen reellwertigen Funktionen, so ist $\alpha(f(X)) := f(-X)$ ein involutiver Automorphismus.

Definition 9.1.2. Sei V ein $\mathbb{K}$ -Vektorraum und $\alpha \in \text{Aut}(\mathbb{K})$ involutiv.

(a) Eine Abbildung $F: V \times V \rightarrow \mathbb{K}$ heißt α -sesquilinear oder einfach *sesquilinear* bzw. *Sesquilinearform*, wenn für $v, v', w, w' \in V$ und $\lambda \in \mathbb{K}$ gilt

$$F(v + v', w) = F(v, w) + F(v', w), \quad F(\lambda v, w) = \lambda F(v, w). \quad (\text{B1})$$

$$F(v, w + w') = F(v, w) + F(v, w'), \quad F(v, \lambda w) = \bar{\lambda} F(v, w). \quad (\text{B2})$$

Ist $\alpha = \text{id}_{\mathbb{R}}$, so bedeutet das gerade, dass F bilinear ist.

Für $\mathbb{K} = \mathbb{C}$ und $\alpha(z) = \bar{z}$ kennen wir bereits Sesquilinearformen.

(b) Eine sequilinare Abbildung $F: V \times V \rightarrow \mathbb{K}$ heißt *hermitesch*, wenn

$$\overline{F(x, y)} = F(y, x) \quad \text{für} \quad x, y \in V$$

gilt. Für $\alpha = \text{id}$ bedeutet diese Bedingung, dass F *symmetrisch* ist, d.h.,

$$F(x, y) = F(y, x) \quad \text{für} \quad x, y \in V.$$

Eine sequilinare Abbildung $F: V \times V \rightarrow \mathbb{K}$ heißt *schiefhermitesch*, wenn

$$\overline{F(x, y)} = -F(y, x) \quad \text{für} \quad x, y \in V$$

gilt. Für $\alpha = \text{id}$ bedeutet diese Bedingung, dass F *schiefsymmetrisch* ist, d.h.,

$$F(x, y) = -F(y, x) \quad \text{für} \quad x, y \in V.$$

Für $\alpha = \text{id}$ nennen wir F *alternierend*, wenn

$$F(x, x) = 0 \quad \text{für} \quad x \in V$$

gilt. Wir haben schon im Kapitel über Determinanten gesehen, dass jede alternierende Form auch schiefsymmetrisch ist. Die Umkehrung aber nur dann gilt, wenn $2 := 1 + 1 \neq 0$ in $\mathbb{K}$ gilt.

(c) Eine sequilinare Abbildung $F: V \times V \rightarrow \mathbb{K}$ heißt *orthosymmetrisch*, wenn aus $F(x, y) = 0$ auch $F(y, x) = 0$ folgt. Man sieht sofort ein, dass symmetrische, schiefsymmetrische, hermitesche und schiefhermitesche Formen orthosymmetrisch sind.

Beispiele 9.1.3. (a) Das Standardskalarprodukt

$$F: \mathbb{K}^n \times \mathbb{K}^n \rightarrow \mathbb{K}, \quad F(x, y) := \sum_{j=1}^n x_j \bar{y}_j$$

ist hermitesch: Die α -Sesquilinearität ist klar. Darüber hinaus haben wir wegen $\alpha^2 = \text{id}$ bzw. $\bar{\bar{x}} = x$:

$$\overline{F(x, y)} = \sum_{j=1}^n \overline{x_j \bar{y}_j} = \sum_{j=1}^n y_j \bar{x}_j = F(y, x).$$

Für $\alpha = \text{id}$ erhalten wir insbesondere auf $\mathbb{K}^n$ die Standardbilinearform

$$F(x, y) = \sum_{j=1}^n x_j y_j.$$

(b) Definiere für $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ in $\mathbb{K}^2$

$$F(x, y) = x_1 \overline{y_1} - x_1 \overline{y_2} + 2x_2 \overline{y_1} - x_2 \overline{y_2}$$

Unter Benutzung der Matrix $A = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$ können wir dies in folgender Form schreiben:

$$F(x, y) = (Ax)^\top \overline{y} = x^\top A^\top \overline{y} = (x_1 \ x_2) \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \overline{y_1} \\ \overline{y_2} \end{pmatrix}.$$

(c) Sei allgemeiner $A = (a_{ij}) \in M_n(\mathbb{K})$. Für $x, y \in \mathbb{K}^n$ sei

$$F_A(x, y) = x^\top A^\top \overline{y} = \sum_{i,j=1}^n a_{ji} x_i \overline{y_j}.$$

Dann ist F_A eine Sesquilinearform auf $\mathbb{K}^n$.

Wir werden sehen, dass jede Bilinearform auf $\mathbb{K}^n$ durch eine Matrix A dieser Weise dargestellt werden kann:

Definition 9.1.4. [Matrix einer Sesquilinearform] Sei $F: V \times V \rightarrow \mathbb{K}$ eine Sesquilinearform auf einem endlichdimensionalen $\mathbb{K}$ -Vektorraum V und $B = (v_1, \dots, v_n)$ eine Basis von V . Für

$$a_{ij} := F(v_j, v_i) \quad \text{für } i, j = 1, \dots, n$$

heißt $[F]_B := A := (a_{ij}) \in M_n(\mathbb{K})$ die *Matrix von F bzgl. der Basis B* .

Die Zahl $D_B(F) := \det([F]_B) \in \mathbb{K}$ heißt *Diskriminante* von F .

Satz 9.1.5. Sei $B = (v_1, \dots, v_n)$ eine Basis von V . Dann gilt:

(i) Jede Sesquilinearform $F: V \times V \rightarrow \mathbb{K}$ ist eindeutig durch ihre Matrix $A = [F]_B$ bestimmt:

$$F\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) = \sum_{i,j=1}^n a_{ji} x_i \overline{y_j}. \quad (9.1)$$

(ii) Für jede Matrix $A \in M_n(\mathbb{K})$ existiert genau eine Sesquilinearform $F: V \times V \rightarrow \mathbb{K}$ mit $[F]_B = A$:

$$F\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) := \sum_{i,j=1}^n a_{ji} x_i \overline{y_j}. \quad (9.2)$$

Beweis. (i) folgt sofort aus

$$F\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\right) = \sum_{i=1}^n x_i F\left(v_i, \sum_{j=1}^n y_j v_j\right) = \sum_{j=1}^n y_j \overline{F(v_i, v_j)} = \sum_{j=1}^n y_j \overline{a_{ji}} = \sum_{j=1}^n x_j \overline{y_j} a_{ji}.$$

(ii) Hierzu hat man nur zu verifizieren, dass durch (9.2) eine Sesquilinearform definiert wird (vgl. Beispiel 9.1.3(c)). $\square$

Definition 9.1.6. Wie im Komplexen definieren wir für Matrizen $A = (a_{ij}) \in M_n(\mathbb{K})$:

$$\overline{A} := (\overline{a_{ij}}) \quad \text{und} \quad A^* := \overline{A}^\top = (\overline{a_{ji}}).$$

Satz 9.1.7. [Transformationsformel] Sei F eine Sesquilinearform auf V sowie $B = (v_1, \dots, v_n)$ und $B' = (v'_1, \dots, v'_n)$ Basen von V . Sei $A = (F(v_j, v_i))$ die Matrix von F bzgl. B und $A' = (F(v'_j, v'_i))$ die Matrix von F bzgl. B' sowie $S = (s_{ij})$ die Transformationsmatrix, deren j . Spalte der Koordinatenvektor $[v'_j]_B$ ist, bzw. $v'_j = \sum_{i=1}^n s_{ij} v_i$. Dann gilt

$$A' = S^* A S.$$

Beweis. Das folgt sofort aus

$$a'_{ij} = F(v'_j, v'_i) = F\left(\sum_{k=1}^n s_{kj} v_k, \sum_{\ell=1}^n s_{\ell i} v_\ell\right) = \sum_{k=1}^n \sum_{\ell=1}^n s_{kj} \overline{s_{\ell i}} \underbrace{F(v_k, v_\ell)}_{a_{\ell k}} = \sum_{k, \ell=1}^n \overline{s_{\ell i}} a_{\ell k} s_{kj}. \quad \square$$

Bemerkung 9.1.8. Für die Diskriminante ergibt sich mit der Transformationsformel insbesondere die Relation

$$D_{B'}(F) = \det(A') = \det(S^* A S) = \det(S^*) \det(S) \det(A) = \overline{\det(S)} \det(S) D_B(F).$$

Sie ist also i.a. nicht unabhängig von der Wahl der Basis.

Bemerkung 9.1.9. Man beachte, dass die Transformationsformel für die Matrix einer Bilinearform verschieden ist von der Transformationsformel für die Matrix einer linearen Abbildung. Nur wenn $S^{-1} = S^*$ gilt, also wenn S unitär ist, sind die beiden Transformationsformeln gleich. Für $\alpha = \text{id}$ bedeutet dies, dass S orthogonal ist. Das ist insbesondere der Fall für Koordinatentransformationen zwischen Orthonormalbasen B und B' in $\mathbb{K}^n$ bzgl. dem Standardskalarprodukt. Wir werden später sehen, wie sich dieser Sachverhalt bei der Hauptachsentransformation anwenden lässt.

Beispiel 9.1.10. Sei $V = \mathbb{R}^2$, $B = (e_1, e_2)$ die Standardbasis und

$$F(x, y) = 2x_1 y_2 + x_2 y_1 \quad \text{für} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2.$$

Dann ist

$$F(x, y) = x^\top A^\top y \quad \text{mit} \quad A = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}.$$

Weiter sei B' die Basis $v'_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v'_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Dann ist die zugehörige Transformationsmatrix $S = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Sind $\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}$, $\begin{pmatrix} y'_1 \\ y'_2 \end{pmatrix}$ die Koordinaten von x und y bzgl. der neuen Basis, so ist F gegeben durch $F(x, y) = (x'_1 \ x'_2) S^\top A^\top S \begin{pmatrix} y'_1 \\ y'_2 \end{pmatrix}$ und

$$(A')^\top = S^\top A^\top S = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ -1 & -3 \end{pmatrix}.$$

Damit ist $F(x, y) = 3x'_1 y'_1 + x'_1 y'_2 - x'_2 y'_1 - 3x'_2 y'_2$ bzgl. der neuen Basis.

Definition 9.1.11. Sei $F: V \times V \rightarrow \mathbb{K}$ eine Sesquilinearform. Für $M \subseteq V$ definieren wir den F -Orthogonalraum

$$M^{\perp_F} = \{v \in V : (\forall m \in M) F(v, m) = 0\}.$$

Die Form F heißt *nicht entartet*, wenn $V^{\perp_F} = \{0\}$ ist, d.h.,

$$(\forall w \in V) F(v, w) = 0 \quad \Rightarrow \quad v = 0.$$

Lemma 9.1.12. Sei $F: V \times V \rightarrow \mathbb{K}$ eine Sesquilinearform und $B = (b_1, \dots, b_n)$ eine Basis von V . Dann gilt:

- (i) F ist genau dann hermitesch (schiefhermitesch), wenn die Matrix $A = [F]_B$ hermitesch (schiefhermitesch) ist.
- (ii) Für $\alpha = \text{id}_{\mathbb{K}}$ ist F genau dann symmetrisch (schiefsymmetrisch), wenn die Matrix $A = [F]_B$ symmetrisch (schiefsymmetrisch) ist.
- (iii) F ist genau dann nicht entartet, wenn die Matrix $A = [F]_B$ invertierbar ist.

Beweis. (i) Wir diskutieren den hermiteschen Fall. Der andere Fall lässt sich analog beweisen. Ist F hermitesch, so folgt sofort

$$a_{ij} = F(v_j, v_i) = \overline{F(v_i, v_j)} = \overline{a_{ji}}.$$

Ist umgekehrt A hermitesch und $B = (b_1, \dots, b_n)$, so erhalten wir für beliebige Elemente $x = \sum_{j=1}^n x_j b_j, y = \sum_{j=1}^n y_j b_j \in V$:

$$\begin{aligned} F(x, y) &= F\left(\sum_{i=1}^n x_i b_i, \sum_{j=1}^n y_j b_j\right) = \sum_{i,j=1}^n x_i \overline{y_j} F(b_i, b_j) = \sum_{i,j=1}^n x_i \overline{y_j} a_{ji} = \sum_{i,j=1}^n x_i \overline{y_j} \overline{a_{ij}} \\ &= \overline{\sum_{i,j=1}^n y_j \overline{x_i} F(b_j, b_i)} = \overline{F\left(\sum_j y_j b_j, \sum_i x_i b_i\right)} = \overline{F(y, x)}. \end{aligned} \quad \square$$

(ii) folgt aus (i) für $\alpha = \text{id}_{\mathbb{K}}$.

(iii) Wir erinnern uns an die Rekonstruktionsformel (9.1)

$$F(v, w) = (A[v]_B)^\perp [w]_B \quad \text{für } v, w \in V.$$

Also ist $v \in V^\perp$ äquivalent zu $(A[v]_B)^\perp y = 0$ für alle $y \in \mathbb{K}^n$. Für $y = e_1, \dots, e_n$ sieht man, dass dies äquivalent ist zu $A[v]_B = 0$. Daher ist $\ker(A) = \{0\}$ (was zur Invertierbarkeit von A äquivalent ist) gleichwertig dazu, dass F nicht entartet ist.

Beispiel 9.1.13. (a) Ist $n = p + q$, so ist die Form

$$F_{p,q}(x, y) := x_1 \overline{y_1} + \dots + x_p \overline{y_p} - x_{p+1} \overline{y_{p+1}} - \dots - x_n \overline{y_n}$$

auf $\mathbb{K}^n$ hermitesch und nicht entartet, denn bzgl. der Standardbasis $B = (e_1, \dots, e_n)$ ist ihre Matrix gegeben durch die invertierbare hermitesche Matrix

$$A = \begin{pmatrix} \mathbf{1}_p & 0 \\ 0 & -\mathbf{1}_q \end{pmatrix} \in M_n(\mathbb{K}).$$

(b) Die schiefsymmetrische Form

$$F(x, y) = \sum_{j=1}^n x_j y_{n+j} - x_{n+j} y_j$$

auf $\mathbb{K}^{2n}$ nicht entartet, denn bzgl. der Standardbasis $B = (e_1, \dots, e_{2n})$ ist ihre Matrix gegeben durch die invertierbare schiefsymmetrische Matrix

$$A = \begin{pmatrix} 0 & \mathbf{1}_n \\ -\mathbf{1}_n & 0 \end{pmatrix} \in M_{2n}(\mathbb{K}).$$

(c) Auf dem reellen Vektorraum $V = C^1([0, 1], \mathbb{R})$ der stetig differenzierbaren Funktionen betrachten wir die reell bilineare Form

$$F(f, g) = \int_0^1 f(x) g'(x) dx.$$

Dann erhalten wir mittels partieller Integration

$$F(g, f) = \int_0^1 g(x)f'(x) dx = - \int_0^1 g'(x)f(x) dx + [fg]_0^1 = -F(f, g) + f(1)g(1) - f(0)g(0).$$

Für $f(x) = x$ und $g(x) = f'(x) = 1$ ergibt sich daher $F(g, f) \neq -F(f, g)$, so dass F nicht schiefssymmetrisch ist.

Schränken wir uns auf den Unterraum

$$V_1 := \{f \in V : f(0) = f(1)\}$$

ein, so ist die Einschränkung $F|_{V_1}$ schiefssymmetrisch, denn für $f, g \in V_1$ ist $f(1)g(1) - f(0)g(0) = 0$.

9.1.2 Automorphismen

Definition 9.1.14. Sei F eine Sesquilinearform auf V . Eine invertierbare lineare Abbildung $\varphi: V \rightarrow V$ heißt *Automorphismus von (V, F)* oder *F -unitär*, wenn

$$F(\varphi(v), \varphi(w)) = F(v, w) \quad \text{für alle } v, w \in V$$

gilt. Wir schreiben $\text{Aut}(V, F) = \text{U}(V, F)$ für die Menge der F -unitären Abbildungen. Man verifiziert leicht, dass dies eine Untergruppe von $\text{GL}(V)$ ist; die *unitäre Gruppe von (V, F)* .

Ist $\alpha = \text{id}$ und F symmetrisch, so schreiben wir auch

$$\text{O}(V, F) := \text{U}(V, F)$$

und sprechen von der *orthogonalen Gruppe* von (V, F) .

Ist $\alpha = \text{id}$ und F schiefssymmetrisch, so schreiben wir auch

$$\text{Sp}(V, F) := \text{U}(V, F)$$

und sprechen von der *symplektischen Gruppe* von (V, F) .

Lemma 9.1.15. Sei $F: V \times V \rightarrow \mathbb{K}$ eine Sesquilinearform und $B = (b_1, \dots, b_n)$ eine Basis von V sowie $A = [F]_B$ die Matrix von F bzgl. B . Dann ist $\varphi \in \text{GL}(V)$ genau dann unitär, wenn die Matrix $S := [\varphi]_B$ der Relation

$$S^*AS = A$$

genügt.

Beweis. Für die Sesquilinearform $G(v, w) := F(\varphi(v), \varphi(w))$ ist φ genau dann unitär, wenn $F = G$ gilt. Wir berechnen dazu die Matrix von G bzgl. B :

$$\begin{aligned} g_{ij} &= G(b_j, b_i) = F(\varphi(b_j), \varphi(b_i)) = F\left(\sum_{k=1}^n s_{kj}b_k, \sum_{\ell=1}^n s_{\ell i}b_\ell\right) = \sum_{k, \ell=1}^n s_{kj}\overline{s_{\ell i}}F(b_k, b_\ell) \\ &= \sum_{k, \ell=1}^n s_{kj}\overline{s_{\ell i}}a_{\ell k} = \sum_{k, \ell=1}^n \overline{s_{\ell i}}a_{\ell k}s_{kj}, \end{aligned}$$

also erhalten wir $[G]_B = S^*AS$. Daher ist $G = F$ äquivalent zu $S^*AS = A$. $\square$

Definition 9.1.16. Ist $A \in M_n(\mathbb{K})$, so rechnet man leicht nach, dass

$$\text{U}_n(\mathbb{K}, A) := \{S \in \text{GL}_n(\mathbb{K}) : S^*AS = A\}$$

eine Untergruppe von $\text{GL}_n(\mathbb{K})$ ist; die *unitäre Gruppe zur Matrix A* . Für $A = \mathbf{1}_n$ erhalten wir die unitäre Gruppe $\text{U}_n(\mathbb{K}, \mathbf{1}_n) = \text{U}_n(\mathbb{K})$.

Lemma 9.1.15 besagt, dass für $A = [F]_B$ und $\varphi \in \text{GL}(V)$ die Bedingung $\varphi \in \text{U}(V, F)$ äquivalent ist zu $[\varphi]_B \in \text{U}_n(\mathbb{K}, A)$ auf der Ebene der zugehörigen Matrizen.

9.1.3 Orthogonalräume

Wir haben inzwischen mehrere Variationen des Konzepts des Orthogonalraum gesehen. Für eine Teilmenge $U \subseteq V$ und eine Sesquilinearform F auf V ist $U^{\perp_F}$ eine Teilmenge von V . Andererseits haben wir in Abschnitt 8.2 für eine Teilmenge A des Dualraums V^* (also eine Menge von linearen Funktionalen auf V) den Orthogonalraum $A^\perp = \bigcap_{\alpha \in A} \ker \alpha$ in V definiert. Im folgenden Beweis werden beide Konzepte zusammenkommen.

Bevor wir einen Blick auf Unterräume $U \subseteq V$ werfen, machen wir uns klar, dass die Einschränkung einer nicht entarteten Form auf einen Unterraum auch entartet sein kann.

Beispiel 9.1.17. (a) Am eklatantesten ist dieses Phänomen, wenn F alternierend ist, denn dann ist F auf jedem eindimensionalen Unterraum von V entartet. In der Tat ist $F(v, v) = 0$ für alle $v \in V$.

(b) Das Phänomen der Entartung auf Unterräumen tritt aber auch für symmetrische Formen auf. Für $\mathbb{K} = \mathbb{R}^2$ und die symmetrische Bilinearform

$$F(x, y) = x_1 y_1 - x_2 y_2$$

ist

$$F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = 0,$$

so dass F auf $U := \mathbb{R} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ entartet ist. Die Matrix von F bzgl. der Standardbasis $B = (e_1, e_2)$ ist $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, also invertierbar und F daher nicht entartet auf $\mathbb{R}^2$. Es gilt sogar $U^{\perp_F} = U$.

Bemerkung 9.1.18. Ist $U \subseteq V$ ein linearer Unterraum und $F: V \times V \rightarrow \mathbb{K}$ eine Sesquilinearform, so ist die Einschränkung $F|_U: U \times U \rightarrow \mathbb{K}$ ebenfalls eine Sesquilinearform. Ein Element $u_0 \in U$ erfüllt genau dann die Bedingung

$$(\forall u \in U) \quad F(u, u_0) = 0,$$

wenn $u_0 \in U \cap U^{\perp_F}$ ist. Daher ist $F|_U$ genau dann nicht entartet, wenn

$$U \cap U^{\perp_F} = \{0\}$$

ist.

Satz 9.1.19. Sei $F: V \times V \rightarrow \mathbb{K}$ orthosymmetrisch und $U \subseteq V$ ein endlichdimensionaler linearer Unterraum.

(i) Ist F nicht entartet, so ist

$$\text{codim}(U^{\perp_F}) = \dim U.$$

(ii) Ist $F|_U$ nicht entartet, so ist

$$V = U \oplus U^{\perp_F}.$$

Beweis. (i) Sei $u_1, \dots, u_k$ eine Basis von U . Wir betrachten die linearen Funktionalen

$$\beta_j(v) := F(v, u_j), \quad j = 1, \dots, k.$$

Dann ist

$$U^{\perp_F} = \bigcap_{j=1}^k \ker \beta_j = A^\perp \quad \text{für} \quad A := \text{span}\{\beta_1, \dots, \beta_k\} \subseteq V^*$$

(mit $\perp$ im Sinne von Definition 8.2.14), so dass aus Lemma 8.2.15(vi) die Relation

$$\text{codim}(U^{\perp_F}) = \text{codim}(A^\perp) = \dim A \leq k = \dim U$$

folgt.

Wir zeigen noch, dass $\beta_1, \dots, \beta_k$ lineare unabhängig sind, also $\dim A = k$ gilt. Hieraus folgt dann (i). Ist $\sum_{j=1}^k \lambda_j \beta_j = 0$, so erhalten wir

$$0 = \sum_{j=1}^k \lambda_j \beta_j(v) = \sum_{j=1}^k \lambda_j F(v, u_j) = F\left(v, \sum_{j=1}^k \overline{\lambda_j} u_j\right)$$

für alle $v \in V$. Da F orthosymmetrisch und nicht entartet ist, folgt hieraus $\sum_{j=1}^k \overline{\lambda_j} u_j = 0$, also $\overline{\lambda_j} = 0$ für alle j und damit auch $\lambda_1 = \dots = \lambda_k = 0$.

(ii) Ist $F|_U$ nicht entartet, so ist $U \cap U^{\perp_F} = \{0\}$ (Bemerkung 9.1.18), d.h. die Einschränkung der Quotientenabbildung

$$q: U \rightarrow V/U^{\perp_F}, \quad v \mapsto v + U^{\perp_F}$$

is injektiv. Also ist $\text{codim}(U^{\perp_F}) = \dim(V/U^{\perp_F}) \geq \dim U$. Im ersten Teil des Beweises von (i) haben wir andererseits eingesehen, dass $\text{codim}(U^{\perp_F}) \leq \dim U$ ist. Daher ist

$$\text{codim}(U^{\perp_F}) = \dim(U).$$

Folglich ist die injektive lineare Abbildung q zwischen Räumen gleicher Dimension ein Isomorphismus. Aus ihrer Surjektivität erhalten wir unmittelbar $V = U \oplus U^{\perp_F}$. $\square$

9.1.4 Normalformen

Definition 9.1.20. Die Basis B von V heißt *Orthonormalbasis* bzgl. F , wenn die Matrix $[F]_B$ die Einheitsmatrix ist. Sie heißt *Orthogonalbasis*, wenn $[F]_B$ eine Diagonalmatrix ist.

Wir werden jetzt sehen, dass in vielen Fällen eine Orthogonalbasis für F existiert.

Theorem 9.1.21. (Orthogonalbasen für hermitesche Formen) *In folgenden Fällen existiert für eine hermitesche Form F auf dem n -dimensional $\mathbb{K}$ -Vektorraum V eine Orthogonalbasis $B = (v_1, \dots, v_n)$:*

(i) $\alpha = \text{id}$, d.h., F ist symmetrisch und $2 \neq 0$ in $\mathbb{K}$:

(a) *Ist $\mathbb{K}$ algebraisch abgeschlossen, so lässt sich sogar $F(v_j, v_j) \in \{0, 1\}$ erreichen. Wir erhalten also eine Matrix der Blockgestalt*

$$[F]_B = \begin{pmatrix} \mathbf{1}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

F ist genau dann nicht entartet, wenn $r = n$. In diesem Fall existiert eine Orthonormalbasis.

(b) *Ist $\mathbb{K} = \mathbb{R}$, so lässt sich $F(v_j, v_j) \in \{-1, 0, 1\}$ erreichen. Wir erhalten also eine Matrix der Blockgestalt*

$$[F]_B = \begin{pmatrix} \mathbf{1}_p & 0 & 0 \\ 0 & -\mathbf{1}_q & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(ii) *Ist $\mathbb{K} = \mathbb{C}$ und $\alpha(z) = \bar{z}$, so lässt sich $F(v_j, v_j) \in \{-1, 0, 1\}$ erreichen. Wir erhalten also eine Matrix der Blockgestalt*

$$[F]_B = \begin{pmatrix} \mathbf{1}_p & 0 & 0 \\ 0 & -\mathbf{1}_q & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Wenn F nicht entartet ist, ist $F(v_j, v_j) \in \{-1, 1\}$.

Beweis. (i) Wir zeigen die Behauptung durch Induktion nach der Dimension n von V . Für $n = 0$ ist nichts zu zeigen. Wir nehmen nun an, dass die Behauptung für Räume der Dimension $n - 1$ richtig ist.

Ist $F = 0$, so erfüllt jede Basis unsere Bedingungen für $r = 0$ und $d = \dim V$. Wir dürfen daher $F \neq 0$ annehmen. Da $2 \neq 0$ ist, haben wir die Polarisierungsformel

$$F(v, w) = \frac{1}{2} (F(v + w, v + w) - F(v, v) - F(w, w)).$$

Wegen $F \neq 0$ existiert daher ein $v'_1 \in V$ mit $F(v'_1, v'_1) \neq 0$.

Ist $\mathbb{K}$ algebraisch abgeschlossen, so existiert ein $\lambda \in \mathbb{K}$ mit $\lambda^2 = F(v'_1, v'_1)$. Dann setzen wir $v_1 := \lambda^{-1}v'_1$ und erhalten $F(v_1, v_1) = 1$.

Ist $\mathbb{K} = \mathbb{R}$, so existiert ein $\lambda \in \mathbb{K}$ mit $\lambda^2 = |F(v'_1, v'_1)|$. Dann setzen wir $v_1 := \lambda^{-1}v'_1$ und erhalten $F(v_1, v_1) \in \{\pm 1\}$.

Da F auf dem eindimensionalen Unterraum $U := \mathbb{K}v_1$ nicht entartet ist, haben wir $V = U \oplus U^\perp$ mit $\dim U^\perp = n - 1$ (Satz 9.1.19(ii)), so dass wir die Induktionsannahme auf $U^\perp$ anwenden können. Hieraus folgt die Behauptung.

(ii) In diesem Fall argumentieren wir analog, indem wir die Polarisierungsformel

$$F(v, w) = \frac{1}{4} \sum_{k=0}^3 i^k F(v + i^k w, v + i^k w)$$

für $F \neq 0$ verwenden, um ein Element v'_1 mit $F(v'_1, v'_1) \neq 0$ zu finden. Da F hermitesch ist, ist $F(v'_1, v'_1)$ reell und es existiert ein $\lambda \in \mathbb{R}$ mit $\lambda^2 = |F(v'_1, v'_1)|$. Für $v_1 := \lambda^{-1}v'_1$ ist dann $F(v_1, v_1) \in \{\pm 1\}$. $\square$

Satz 9.1.22. (Sylvesterscher Trägheitssatz) Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Ist F eine hermitesche Form auf dem $\mathbb{K}$ -Vektorraum V und $v_1, \dots, v_n$ eine Orthogonalbasis mit

$$F(v_j, v_j) = 1, \quad 1 \leq j \leq p; \quad F(v_j, v_j) = -1, \quad p + 1 \leq j \leq p + q; \quad F(v_j, v_j) = 0, \quad j > p + q,$$

so sind die Zahlen p und q unabhängig von der Wahl der Basis.

Die Zahl (p, q) heißt *Signatur der Form F* und die Differenz $p - q$ heißt *Trägheitsindex*.

Beweis. Ist $B = (v_1, \dots, v_n)$, so ist

$$A := [F]_B = \begin{pmatrix} \mathbf{1}_p & 0 & 0 \\ 0 & -\mathbf{1}_q & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

also $p + q = \text{rank}([F]_B)$, und diese Zahl ist unabhängig von der Wahl der Basis, denn für $S \in \text{GL}_n(\mathbb{K})$ hat die Matrix S^*AS den gleichen Rang (vgl. Satz 9.1.7).

Auf dem Unterraum

$$U := \text{span}\{v_1, \dots, v_p, v_{p+q+1}, \dots, v_n\}$$

ist F positiv semidefinit und auf $W := \text{span}\{v_{p+1}, \dots, v_{p+q}\}$ ist F negativ definit, d.h., $-F|_W$ ist positiv definit.

Ist W' ein beliebiger Unterraum von V auf dem F negativ definit ist, so ist $F(w, w) < 0$ für $0 \neq w \in W'$. Insbesondere ist also $W' \cap U = \{0\}$, also

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

und somit $\dim W' \leq \dim W$. Daher ist q die maximale Dimension eines Unterraums von V auf dem F negativ definit ist. Diese Zahl ist unmittelbar durch F definiert, hängt also nicht von der Wahl einer Basis ab. $\square$

Theorem 9.1.23. (Kanonische Basen für alternierende Formen) *Ist F alternierend und V endlichdimensional, so existiert eine Basis*

$$p_1, \dots, p_r, \quad q_1, \dots, q_r, \quad v_1, \dots, v_d$$

von V , so dass

$$(\forall v \in V) F(v, v_j) = 0, \quad F(p_j, p_j) = F(q_j, q_j) = 0, \quad F(p_j, q_i) = \delta_{ij} \quad \text{für } i, j = 1, \dots, r.$$

Beweis. Auch hier zeigen wir die Behauptung durch Induktion nach der Dimension n von V . Für $n = 0$ ist nichts zu zeigen. Wir nehmen nun an, dass die Behauptung für Räume der Dimension $\leq n - 1$ richtig ist.

Ist $F = 0$, so erfüllt jede Basis unsere Bedingungen. Wir dürfen daher $F \neq 0$ annehmen. Dann existieren $p_1 \in V$ und $q'_1 \in V$ mit $\lambda := F(p_1, q'_1) \neq 0$. Für $q_1 := \lambda^{-1} q'_1$ ist dann $F(p_1, q_1) = 1$. Da F alternierend ist, ist $q_1 \notin \mathbb{K}p_1$. Also ist $U := \mathbb{K}p_1 + \mathbb{K}q_1$ ein zweidimensionaler Unterraum. Die Einschränkung auf U hat bzgl. der Basis (p_1, q_1) die Matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

ist also nicht entartet.

Da F auf U nicht entartet ist, haben wir $V = U \oplus U^\perp$ mit $\dim U^\perp = n - 2$ (Satz 9.1.19), so dass wir die Induktionsannahme auf $U^\perp$ anwenden können. Hieraus folgt die Behauptung. $\square$

Aufgaben zu Abschnitt 9.1

Aufgabe 9.1.1. Ist V ein $\mathbb{K}$ -Vektorraum und $\alpha \in \text{Aut}(\mathbb{K})$ ein involutiver Automorphismus, so ist die Menge $\text{Sesq}_\alpha(V, \mathbb{K})$ der α -Sesquilinearformen ein $\mathbb{K}$ -Vektorraum bzgl. der Operationen

$$(F_1 + F_2)(v, w) := F_1(v, w) + F_2(v, w) \quad \text{und} \quad (\lambda F)(v, w) := \lambda F(v, w).$$

Aufgabe 9.1.2. Sei V ein $\mathbb{K}$ -Vektorraum und V^* sein Dualraum. Zeigen Sie:

(a) Ist $F: V \times V \rightarrow \mathbb{K}$ bilinear, so definiert

$$F^\sharp: V \rightarrow V^*, \quad F^\sharp(v)(w) := F(v, w)$$

eine lineare Abbildung.

(b) Ist $G: V \rightarrow V^*$ linear, so definiert

$$G^\flat: V \times V \rightarrow \mathbb{K}, \quad G^\flat(v, w) := G(v)(w)$$

eine Bilinearform.

(c) Die Abbildung $\Gamma: \text{Hom}(V, V^*) \rightarrow \text{Bil}(V, \mathbb{K}), G \mapsto G^\flat$ ist ein linearer Isomorphismus auf den Vektorraum $\text{Bil}(V, \mathbb{K})$ der Bilinearformen auf V (vgl. Aufgabe 9.1.1).

Aufgabe 9.1.3. Zeigen Sie: Für jede invertierbare symmetrische komplexe Matrix $A \in M_n(\mathbb{C})$ existiert ein $S \in \text{GL}_n(\mathbb{C})$ mit $S^\top A S = \mathbf{1}_n$. Ist A nicht invertierbar, so existiert ein $S \in \text{GL}_n(\mathbb{C})$ mit

$$S^\top A S = \begin{pmatrix} \mathbf{1}_r & 0 \\ 0 & 0 \end{pmatrix} \quad \text{für } r = \text{rank}(A).$$

Aufgabe 9.1.4. Zeigen Sie: Für jede hermitesche komplexe Matrix $A \in M_n(\mathbb{C})$ existiert ein $S \in \text{GL}_n(\mathbb{C})$ und $p, q \in \mathbb{N}_0$ mit

$$S^* A S = \begin{pmatrix} \mathbf{1}_p & 0 & 0 \\ 0 & -\mathbf{1}_q & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{für } p + q = \text{rank}(A).$$

Aufgabe 9.1.5. Zeigen Sie: Für jede reelle symmetrische Matrix $A \in M_n(\mathbb{R})$ existiert ein $S \in \text{GL}_n(\mathbb{R})$ und $p, q \in \mathbb{N}_0$ mit

$$S^\top AS = \begin{pmatrix} \mathbf{1}_p & 0 & 0 \\ 0 & -\mathbf{1}_q & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{für } p + q = \text{rank}(A).$$

Aufgabe 9.1.6. Sei $\mathbb{K} = \mathcal{F}_2$ der zweielementige Körper und $A \in M_n(\mathbb{K})$. Für welche Matrizen A ist die Bilinearform $F_A(x) := x^\top Ax$ symmetrisch/schiefsymmetrisch/alternierend? Geben Sie jeweils Beispiele für $n = 2$ an.

9.2 Quadratische Formen

Sei V ein Vektorraum über einem Körper $\mathbb{K}$. Wir nehmen an, dass $2 := 1 + 1 \neq 0$ in dem Körper $\mathbb{K}$ gilt (wir lassen also z.B. den zweielementigen Körper $\mathbb{Z}/2\mathbb{Z}$ nicht zu).

In diesem Abschnitt werden wir quadratische Formen als Mittel zum Studium symmetrischer Bilinearformen kennenlernen. Auch wenn die Definition einer quadratischen Form zunächst etwas abstrakt anmutet, werden wir doch gleich sehen, was sich konkret dahinter verbirgt.

Definition 9.2.1. [Quadratische Formen] Sei V ein $\mathbb{K}$ -Vektorraum. Eine Abbildung

$$Q: V \rightarrow \mathbb{K}$$

heißt *quadratische Form*, wenn

$$Q(\lambda v) = \lambda^2 Q(v)$$

für alle $\lambda \in \mathbb{K}$ und $v \in V$ gilt sowie die Abbildung

$$F: V \times V \rightarrow \mathbb{K}, \quad (x, y) \mapsto Q(x + y) - Q(x) - Q(y)$$

bilinear ist. Man beachte, dass F definitionsgemäß symmetrisch ist, und dass sich Q durch

$$F(x, x) = Q(2x) - 2Q(x) = 4Q(x) - 2Q(x) = 2Q(x)$$

aus F zurückgewinnen lässt.

Frage: Wo haben wir die Annahme $1 + 1 \neq 0$ benutzt?

Bemerkung 9.2.2. Sei F eine symmetrische Bilinearform auf V . Die Abbildung

$$Q: V \rightarrow \mathbb{K}, \quad Q(x) = F(x, x), \quad x \in V$$

erfüllt trivialerweise

$$Q(\lambda x) = F(\lambda x, \lambda x) = \lambda F(x, \lambda x) = \lambda^2 F(x, x) = \lambda^2 Q(x).$$

Weiter gilt

$$\begin{aligned} Q(x + y) - Q(x) - Q(y) &= F(x + y, x + y) - F(x, x) - F(y, y) \\ &= F(x, x) + F(x, y) + F(y, x) + F(y, y) - F(x, x) - F(y, y) \\ &= F(x, y) + F(y, x) = 2F(x, y). \end{aligned}$$

Also ist Q eine quadratische Form und F die zugehörige Bilinearform. Wir nennen Q die zu F *assoziierte quadratische Form*. Wir erhalten so eine bijektive Korrespondenz zwischen quadratischen Formen und symmetrischen Bilinearformen. Beides sind verschiedene Sichtweisen auf die gleiche mathematische Struktur.

Beispiel 9.2.3. Die zu dem Standardskalarprodukt assoziierte Form auf $\mathbb{R}^n$ ist

$$Q(x) = \langle x, x \rangle = \|x\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Sei A die Matrix einer symmetrischen Bilinearform auf $\mathbb{K}^n$ (bzgl. der Standardbasis). Man erinnere sich, dass A eine symmetrische Matrix ist. Wegen

$$F(x, y) = x^\top A y = \sum_{i,j=1}^n a_{ij} x_i y_j$$

ist die assoziierte quadratische Form Q gegeben durch

$$Q(x) = F(x, x) = x^\top A x = \sum_{i,j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i<j} 2a_{ij} x_i x_j.$$

Wir haben hier die Ausdrücke $a_{ij} x_i x_j + a_{ji} x_j x_i$ zu $2a_{ij} x_i x_j$ für alle $i < j$ zusammengefasst. Das können wir tun, da $a_{ij} = a_{ji}$ aufgrund der Symmetrie von A gilt.

Umgekehrt sei $Q : \mathbb{K}^n \rightarrow \mathbb{K}$ eine Funktion folgender Art

$$Q(x) = \sum_{i=1}^n a_i x_i^2 + \sum_{i<j} b_{ij} x_i x_j,$$

Dann ist Q eine quadratische Form. Die assoziierte symmetrische Bilinearform hat die Matrix

$$A = \begin{pmatrix} a_1 & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1n} \\ \frac{1}{2}b_{12} & a_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{2}b_{1n} & \cdots & \cdots & a_n \end{pmatrix}$$

oder in anderen Worten $Q(x) = x^\top A x$.

Beispiel 9.2.4. Für $Q(x) = 2x_1^2 + 4x_1x_2 - x_2^2$ ist die Matrix $A = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$.

9.3 Hauptachsentransformation

In diesem Abschnitt betrachten wir nur den euklidischen Vektorraum $V = \mathbb{R}^n$ mit dem Standardskalarprodukt.

Satz 9.3.1. Sei F eine symmetrische Bilinearform auf $\mathbb{R}^n$ und Q die assoziierte quadratische Form $Q(x) = F(x, x)$. Dann gibt es eine Orthonormalbasis

$$B' = (v'_1, \dots, v'_n)$$

bzgl. dem Standardskalarprodukt von $\mathbb{R}^n$, so dass die Matrix $A' = [F]_{B'}$ von F bzgl. der Basis B' eine Diagonalmatrix ist:

$$A' = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix},$$

d.h. für $x = x'_1 v'_1 + \cdots + x'_n v'_n$, $y = y'_1 v'_1 + \cdots + y'_n v'_n$ gilt

$$\begin{aligned} F(x, y) &= \lambda_1 x'_1 y'_1 + \cdots + \lambda_n x'_n y'_n, \\ Q(x) &= \lambda_1 x_1'^2 + \cdots + \lambda_n x_n'^2. \end{aligned}$$

Die eindimensionalen linearen Teilräume, die von den neuen Basisvektoren $v'_1, \dots, v'_n$ aufgespannt werden, heißen Hauptachsen von F bzw. Q . Der Übergang von der kanonischen Basis B zur Basis B' heißt Hauptachsentransformation.

Beweis. Sei $A = [F]_B$ die Matrix von F bzgl. der kanonischen Basis $B = (e_1, \dots, e_n)$. Da F symmetrisch ist, ist die Matrix A eine reelle symmetrische Matrix (vgl. Lemma 9.1.12). Aufgrund des Hauptsatzes für reelle symmetrische Matrizen 7.4.12 gibt es eine Orthonormalbasis $B' = (v'_1, \dots, v'_n)$ aus Eigenvektoren von A . Die Transformationsmatrix Q ist orthogonal. Weiterhin ist $A' := Q^{-1}AQ = Q^\top AQ$ eine Diagonalmatrix (die Diagonaleinträge sind die Eigenwerte λ_i von A). Nach Satz 9.1.7 ist $A' := Q^\top AQ$ die Matrix von F bzgl. der Basis B' . $\square$

Der Satz über die Hauptachsentransformation kann auch so verstanden werden, dass durch die Basis B' gleichzeitig das Standardskalarprodukt und die symmetrische Bilinearform F diagonalisiert werden.

Eine abstrakte Variante des Satz aus der das klarer hervorgeht ist die folgende. Da jeder endlichdimensionale euklidische Raum isometrisch isomorph zu $\mathbb{R}^n$ ist, folgt dieser Satz sofort aus Satz 9.3.1.

Satz 9.3.2. *Ist V ein endlichdimensionaler euklidischer Vektorraum und $F: V \times V \rightarrow \mathbb{R}$ eine symmetrische Bilinearform auf V , so existiert eine ONB von V , die auch bzgl. F eine Orthogonalbasis ist.*

Definition 9.3.3. [Quadratische Hyperflächen] Für eine quadratische Form Q auf $\mathbb{R}^n$ heißt

$$\{x \in \mathbb{R}^n : Q(x) = 1\}$$

quadratische Hyperfläche (auch einfach *Quadrik*). Ist $A = (a_{ij})$ die Matrix von Q bzgl. der Basis B , so kann die Gleichung (1) in der Form

$$\sum_{i=1}^n a_{ii}x_i^2 + 2 \sum_{i < j} a_{ij}x_i x_j = 1 \quad (2)$$

geschrieben werden, wobei die x_i die Koordinaten von $x \in \mathbb{R}^n$ bzgl. B sind. Ersetzt man die 1 auf der rechten Seite durch irgendein $a > 0$, so ändert sich nichts Wesentliches.

Nach Satz 9.3.1 gibt es eine Orthonormalbasis $B' = (v'_1, \dots, v'_n)$, so dass die Gleichung (2) auf die Form

$$\lambda_1 x_1'^2 + \dots + \lambda_n x_n'^2 = 1 \quad (3)$$

gebracht werden kann, wobei die x'_i die Koordinaten von $x \in \mathbb{R}^n$ bzgl. der neuen Basis B' sind. Die 1-dimensionalen linearen Teilräume, die von den Vektoren $v'_1, \dots, v'_n$ aufgespannt werden, heißen *Hauptachsen* der Quadrik. Die Transformation der Gleichung von der Form (2) in die Form (3) heißt *Hauptachsentransformation*.

9.3.4. [Hauptachsentransformation] Wir nehmen an, eine Quadrik im $\mathbb{R}^n$ sei durch eine Gleichung der Form

$$\sum_{i=1}^n a_i x_i^2 + \sum_{1 \leq i < j \leq n} b_{ij} x_i x_j = 1 \quad (Q)$$

gegeben, wobei die x_i die Koordinaten von $x \in \mathbb{R}^n$ bzgl. der Standardbasis sind. Wir wollen dieselbe Quadrik bzgl. einer neuen Basis beschreiben, die aus den Hauptachsen der Quadrik besteht.

- (1) Bilde die Matrix

$$A = \begin{pmatrix} a_1 & \frac{1}{2}b_{12} & \dots & \frac{1}{2}b_{1n} \\ \frac{1}{2}b_{12} & a_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}b_{1n} & \dots & \dots & a_n \end{pmatrix}.$$

Diese Matrix ist symmetrisch und reell.

- (2) Gleichung
- (Q)
- kann nun in folgender Form geschrieben werden:

$$x^\top Ax = 1. \quad (Q_1)$$

- (3) Berechne die Eigenwerte $\lambda_1, \dots, \lambda_n$ von A , jeden mit seiner algebraischen Vielfachheit. (Aus Satz 7.4.11 wissen wir, dass alle Eigenwerte von A reell sind.)
- (4) Bestimme eine Orthonormalbasis $B' = (v'_1, \dots, v'_n)$ von Eigenvektoren von A in $\mathbb{R}^n$. Die Existenz einer solchen Basis folgt aus dem Hauptsatz über symmetrische Matrizen 7.4.12.
- (5) Sei $S = (v'_1, \dots, v'_n)$ die Matrix, bei der die Spalten die Vektoren v'_j der Orthonormalbasis B' sind. Dann ist S eine Orthogonalmatrix. Mit $S^{-1} = S^\top$ wird die Matrix A auf die Matrix

$$A' := S^{-1}AS = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

transformiert und zwar bzgl. der Basis B' .

Wir behaupten, dass diese neue Matrix die Quadrik bzgl. der neuen Koordinaten beschreibt: Seien $x'_1, \dots, x'_n$ die Koordinaten von $x \in \mathbb{R}^n$ bzgl. der neuen Basis B' . Dann gilt

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = S \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}.$$

Einsetzen in Gleichung (Q_1) ergibt

$$x^\top Ax = (x'_1 \dots x'_n) S^\top AS \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = (x'_1 \dots x'_n) A' \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = (x')^\top A' x'.$$

Bzgl. der neuen Basis B' wird die Quadrik durch die folgende Gleichung beschrieben:

$$\lambda_1 x'^2_1 + \lambda_2 x'^2_2 + \dots + \lambda_n x'^2_n = 1.$$

9.4 Typen reeller Quadriken

Wir geben nun eine Klassifikation der quadratischen Formen in den Dimensionen 2 und 3. Dabei hängt die Gestalt einer Quadrik nur von den Vorzeichen der Eigenwerte λ_i ab. Insbesondere ist der Typ einer Quadrik durch die Eigenwerte der Matrix A vollständig bestimmt.

Wir haben gesehen, dass bzgl. einer geeignet gewählten Orthonormalbasis jede quadratische Form im $\mathbb{R}^n$ wie folgt geschrieben werden kann:

$$Q(x) = \lambda_1 x^2_1 + \dots + \lambda_n x^2_n.$$

Für eine Klassifikation genügt es, quadratische Formen dieses Typs zu betrachten. Ist einer der Koeffizienten $\lambda_i = 0$, dann heißt die quadratische Form *entartet*. Ist $\lambda_i \neq 0$ für alle i , dann heißt sie *nicht entartet*. Zur Klassifikation betrachten wir die Quadriken mit der Gleichung

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 = 1.$$

9.4.1 Quadratische Kurven

In diesem Abschnitt ist $n = 2$, die entsprechenden Quadriken sind also quadratische Kurven.

Nach Hauptachsentransformation erhält die quadratische Gleichung die Form:

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1. \quad (Q)$$

Wir betrachten zunächst die nicht entarteten Fälle $\lambda_1, \lambda_2, \lambda_3 \neq 0$ und setzen

$$a := \frac{1}{\sqrt{|\lambda_1|}} \quad \text{und} \quad b := \frac{1}{\sqrt{|\lambda_2|}}.$$

(a) $\lambda_1 > 0$ und $\lambda_2 > 0$: Dann wird Gleichung (Q) zu

$$\boxed{\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1.}$$

Dies ist die Gleichung einer *Ellipse* mit den *Halbachsen* a, b .

(b) $\lambda_1 > 0, \lambda_2 < 0$:

Gleichung (Q) wird zu

$$\boxed{\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1.}$$

Dies ist die Gleichung einer *Hyperbel* mit den *Asymptoten* $x_2 = \pm \frac{b}{a} x_1$.

Die Asymptoten werden auf folgende Art erhalten: Die Gleichung der Hyperbel kann in der Form

$$\frac{x_1^2}{a^2} = \frac{x_2^2}{b^2} + 1.$$

geschrieben werden. Sind x_1 und x_2 groß (im Betrag), kann der Wert 1 vernachlässigt werden, denn wir erhalten zum Beispiel für den positiven Zweig die Funktionsgleichung $x_1 = \frac{a}{b} \sqrt{x_2^2 + b^2}$ und somit

$$\lim_{x_2 \rightarrow \infty} \frac{x_1}{x_2} = \lim_{x_2 \rightarrow \infty} \frac{a}{b} \sqrt{1 + \frac{b^2}{x_2^2}} = \frac{a}{b}$$

sowie

$$\lim_{x_2 \rightarrow \infty} x_1 - \frac{a}{b} x_2 = \lim_{x_2 \rightarrow \infty} \frac{a}{b} \left(\sqrt{x_2^2 + b^2} - x_2 \right) = \lim_{x_2 \rightarrow \infty} \frac{a}{b} \frac{b^2}{\sqrt{x_2^2 + b^2} + x_2} = 0.$$

Die Gerade mit der Gleichung $x_1 = \frac{a}{b} x_2$ ist also eine Asymptote der quadratischen Kurve. Analog erhält man für den negativen Zweig die Gerade $x_1 = -\frac{a}{b} x_2$ als Asymptote. Das asymptotische Geradenpaar der Hyperbel ist also durch die quadratische Gleichung

$$\frac{x_1^2}{a^2} = \frac{x_2^2}{b^2}$$

gegeben.

(c) $\lambda_1 < 0, \lambda_2 < 0$: Die Gleichung (Q) hat keine reelle Lösung.

(d) $\lambda_1 > 0, \lambda_2 = 0$:

Hier vereinfacht sich (Q) zu $\lambda_1 x_1^2 = 1$, also gilt

$$x_1 = \pm \frac{1}{\sqrt{\lambda_1}} = \pm a$$

In diesem degenerierten Fall besteht die quadratische Kurve aus zwei parallelen Geraden (in der Zeichnung punktiert gezeichnet).

Beispiel 9.4.1. Wir betrachten zunächst die quadratische Gleichung:

$$x_1^2 + 6x_1x_2 + x_2^2 = 1.$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$.

Charakteristische Gleichung: $\det(A - \lambda \mathbf{1}) = \begin{vmatrix} 1 - \lambda & 3 \\ 3 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - 9 = 0$.

Eigenwerte: $\lambda_1 = 4, \lambda_2 = -2$.

Eigenvektoren (normiert):

$$v'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Transformationsmatrix: $S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \cos \frac{\pi}{4} & -\sin \frac{\pi}{4} \\ \sin \frac{\pi}{4} & \cos \frac{\pi}{4} \end{pmatrix}$ (dies ist eine Drehung um $\frac{\pi}{4}$).

Für die Koordinaten bzgl. der neuen Basis v'_1, v'_2 ergibt sich die Gleichung der quadratischen Kurve zu:

$$4x_1'^2 - 2x_2'^2 = 1$$

oder gleichwertig dazu $\frac{x_1'^2}{(\frac{1}{2})^2} - \frac{x_2'^2}{(\frac{1}{\sqrt{2}})^2} = 1$. Das bedeutet, es liegt eine Hyperbel mit den Asymptoten $x'_2 = \pm \sqrt{2}x'_1$ vor.

Die Koordinatentransformation zwischen den Koordinatensystemen wird beschrieben durch

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}, \quad \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = S^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Das bedeutet

$$\begin{aligned} x_1 &= \frac{1}{\sqrt{2}}(x'_1 - x'_2), & x'_1 &= \frac{1}{\sqrt{2}}(x_1 + x_2), \\ x_2 &= \frac{1}{\sqrt{2}}(x'_1 + x'_2), & x'_2 &= \frac{1}{\sqrt{2}}(x_1 - x_2). \end{aligned}$$

Beispiel 9.4.2. Wir betrachten nun die quadratische Gleichung

$$7x_1^2 + 6\sqrt{3}x_1x_2 + 13x_2^2 = 16.$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} 7 & 3\sqrt{3} \\ 3\sqrt{3} & 13 \end{pmatrix}$

Charakteristisches Polynom:

$$\det(A - X\mathbf{1}) = \begin{vmatrix} 7 - X & 3\sqrt{3} \\ 3\sqrt{3} & 13 - X \end{vmatrix} = (7 - X)(13 - X) - 27 = X^2 - 20X + 64.$$

Eigenwerte: $\lambda_{1/2} = 10 \pm \sqrt{100 - 64} = 10 \pm 6$, also $\lambda_1 = 4$ und $\lambda_2 = 16$.

Eigenvektoren (normiert) zu diesen Eigenwerten:

$$v'_1 = \frac{1}{2} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}, \quad v'_2 = \frac{1}{2} \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}.$$

Transformationsmatrix: $S = \frac{1}{2} \begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} = \begin{pmatrix} \cos \frac{\pi}{6} & -\sin \frac{\pi}{6} \\ \sin \frac{\pi}{6} & \cos \frac{\pi}{6} \end{pmatrix}$ (dies ist eine Drehung um $\frac{\pi}{6}$).

Bzgl. der neuen Koordinaten hat die quadratische Kurve die Gleichung $4x_1'^2 + 16x_2'^2 = 16$ oder

$$\boxed{\frac{x_1'^2}{4} + x_2'^2 = 1.}$$

Dies ist eine Ellipse, die um $\frac{\pi}{6}$ um die Halbachse $a = 2$, $b = 1$ gedreht wurde.

Wir schreiben wieder die Hauptachsentransformation explizit aus:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix}, \quad \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = S^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

oder

$$\begin{aligned} x_1 &= \frac{1}{2}(\sqrt{3} x'_1 - x'_2), & x'_1 &= \frac{1}{2}(\sqrt{3} x_1 + x_2), \\ x_2 &= \frac{1}{2}(x'_1 + \sqrt{3} x'_2), & x'_2 &= \frac{1}{2}(-x_1 + \sqrt{3} x_2). \end{aligned}$$

9.4.2 Quadratische Flächen

In diesem Abschnitt ist $n = 3$, so dass die Quadriken quadratische Flächen im $\mathbb{R}^3$ sind. Bzgl. einer Hauptachsenbasis sind sie Lösungsmenge einer Gleichung der Form

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 = 1.$$

Wir diskutieren hier zunächst die nichtentarteten Fälle $\lambda_1, \lambda_2, \lambda_3 \neq 0$ und setzen

$$a := \frac{1}{\sqrt{|\lambda_1|}}, \quad b := \frac{1}{\sqrt{|\lambda_2|}} \quad \text{und} \quad c := \frac{1}{\sqrt{|\lambda_3|}}.$$

(a) $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$:

Dann wird Gleichung (Q) zu

$$\boxed{\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1}$$

Dies ist die Gleichung eines *Ellipsoids* mit den Halbachsen a, b, c .

(b) $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0$:

Dann wird (Q) zu

$$\boxed{\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 1.}$$

Dies ist die Gleichung eines *einschaligen Hyperboloids*.

Analog behandelt man die anderen Fälle, bei denen zwei der $\lambda_1, \lambda_2, \lambda_3$ positiv und der dritte Wert negativ ist.

Unter den Fällen, bei denen zwei der $\lambda_1, \lambda_2, \lambda_3$ negativ und der dritte Wert nicht negativ ist, wählen wir den folgenden Fall aus:

- (c) $\lambda_1 < 0, \lambda_2 < 0, \lambda_3 > 0$:
 (Q) wird dann zu

$$-\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$$

Dies ist ein *zweischaliges Hyperboloid*.

- (d) $\lambda_1 < 0, \lambda_2 < 0, \lambda_3 < 0$: Die Gleichung (Q) hat keine reelle Lösung.

Bemerkung 9.4.3. Die beiden Hyperboloide in (b) und (c) haben einen gemeinsamen *asymptotischen Kegel*: Für große x_1, x_2, x_3 kann man die rechte Seite 1 vernachlässigen und in beiden Fällen erhält man asymptotisch die Gleichung

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = \frac{x_3^2}{c^2}.$$

Dies ist die Gleichung einer Kegelfläche.

Bemerkung 9.4.4. Ist $\lambda_3 = 0$, so erhalten wir die Quadrik

$$F := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \lambda_1 x_1^2 + \lambda_2 x_2^2 = 1\} = B \times \mathbb{R}, \quad B := \{(x_1, x_2) \in \mathbb{R}^2 : \lambda_1 x_1^2 + \lambda_2 x_2^2 = 1\}.$$

Die Fläche F ist dann ein *Zylinder mit der Basis B* , wobei B eine quadratische Kurve in $\mathbb{R}^2$ ist.

Ist B eine Ellipse/Hyperbel, so nennt man F einen *elliptischen/hyperbolischen Zylinder*.

Bemerkung 9.4.5. Im Fall $\lambda_1 = \lambda_2 > 0$ ergeben sich sogenannte *Rotationsflächen*, die durch Drehung einer Kurve in der x_1 - x_3 -Ebene um die x_3 -Achse entstehen. Die Kurve, die gedreht wird, ist in Abhängigkeit von den Vorzeichen von λ_3 eine Ellipse ($\lambda_3 > 0$), eine Hyperbel ($\lambda_3 < 0$) oder ein Geradenpaar ($\lambda_3 = 0$). Die Schnittmengen der Fläche mit horizontalen Ebenen (parallel zur x_1 - x_2 -Ebene) sind Kreise.

Beispiel 9.4.6. Wir betrachten die quadratische Gleichung

$$x_1 x_2 + x_1 x_3 - x_2 x_3 = 1. \quad (Q)$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{pmatrix}$

Charakteristisches Polynom:

$$\begin{aligned} \det(A - X\mathbf{1}) &= \begin{vmatrix} -X & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -X & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -X \end{vmatrix} = -X^3 - \frac{1}{8} - \frac{1}{8} + \frac{X}{4} + \frac{X}{4} + \frac{X}{4} \\ &= -X^3 + \frac{3}{4}X - \frac{1}{4} = (X+1)\left(-X^2 + X - \frac{1}{4}\right) = -(X+1)\left(X - \frac{1}{2}\right)^2. \end{aligned}$$

(dabei haben wir die Wurzel -1 geraten, um $X+1$ ausklammern zu können).

Eigenwerte: $\lambda_1 = -1, \lambda_2 = \lambda_3 = \frac{1}{2}$

Eigenvektoren zu den Eigenwerten:

- $\lambda_1 = -1$: Gauß-Jordan-Elimination ergibt

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{4} & -\frac{3}{4} \\ 0 & -\frac{3}{4} & \frac{3}{4} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{4} & -\frac{3}{4} \\ 0 & 0 & 0 \end{pmatrix}$$

Also ist $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ ein Eigenvektor, und durch normieren erhalten wir $v'_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

- $\lambda_2 = \lambda_3 = \frac{1}{2}$: Wir verwenden wieder Gauß-Jordan-Elimination

$$\begin{array}{ccc|ccc} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \rightsquigarrow & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & & 0 & 0 & 0 \end{array}$$

Damit können wir

$$b_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \text{ und } b_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

als Fundamentalsystem von Lösungen wählen. Wir wenden Gram-Schmidt an, um ein Orthonormalsystem von Lösungen zu erhalten:

$$v'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, v'_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

Zur Berechnung von v'_3 verwenden wir

$$b_3 - \langle b_3, v'_2 \rangle v'_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}.$$

Damit ergibt sich die neue Orthogonalbasis zu v'_1, v'_2, v'_3 . Wir erhalten die Transformationsmatrix

$$Q = \begin{pmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{pmatrix}.$$

Die Gleichung unserer quadratischen Fläche bzgl. der neuen Basis v'_1, v'_2, v'_3 ist

$$\boxed{-x_1'^2 + \frac{1}{2}x_2'^2 + \frac{1}{2}x_3'^2 = 1}.$$

Da die beiden Eigenwerte λ_2 und λ_3 gleich sind, ist die Hyperfläche, die durch Q in den x' -Koordinaten beschrieben wird, invariant unter Rotation um die x'_1 -Achse. Ihr Schnitt mit der (x'_1, x'_3) -Ebene ist die Hyperbel mit der Gleichung $-x_1'^2 + \frac{1}{2}x_3'^2 = 1$. Durch Drehung um die x'_1 -Achse erhalten wir also ein einschaliges Hyperboloid, denn jeder der beiden Zweige der Hyperbel geht durch Rotation um den Winkel π um die x'_1 -Achse (Spiegelung an der x'_1 -Achse in der (x'_1, x'_3) -Ebene) in den jeweils anderen Zweig über.

Bemerkung 9.4.7. Wir brauchen nur die Eigenwerte $\lambda_1, \lambda_2, \lambda_3$ von A , um die obige Gleichung zu erhalten, die uns etwas über die Form der quadratischen Fläche sagt. Die Eigenvektoren v'_1, v'_2, v'_3 werden nur gebraucht, wenn wir wissen wollen, wie die Fläche im $\mathbb{R}^3$ plaziert ist.

Beispiel 9.4.8. Wir betrachten nun die Gleichung

$$\boxed{-7x_1^2 - x_2^2 - x_3^2 + 8x_1x_2 + 8x_1x_3 + 16x_2x_3 = 9.} \quad (Q)$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} -7 & 4 & 4 \\ 4 & -1 & 8 \\ 4 & 8 & -1 \end{pmatrix}$

Charakteristisches Polynom:

$$\begin{aligned}\det(A - X\mathbf{1}) &= \begin{vmatrix} -7-X & 4 & 4 \\ 4 & -1-X & 8 \\ 4 & 8 & -1-X \end{vmatrix} \\ &= (-7-X)(-1-X)(-1-X) + 128 + 128 + 16(1+X) + 16(1+X) + 64(7+X) \\ &= -(7+X)(1+X)^2 + 96X + 736 = -X^3 - 9X^2 + 81X + 729 \\ &= -(X-9)(X+9)^2\end{aligned}$$

Eigenwerte: $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = -9$.

Nach Koordinatentransformation lautet die Gleichung $9x_1'^2 - 9x_2'^2 - 9x_3'^2 = 9$ oder

$$x_1'^2 - x_2'^2 - x_3'^2 = 1.$$

Dies ist ein zweischaliges Hyperboloid: Es entsteht durch Rotation der Hyperbel mit der Gleichung $x_1'^2 - x_3'^2 = 1$ in der (x_1', x_3') -Ebene um die x_1' -Achse. Da beide Zweige der Hyperbel durch Spiegelung an der x_1' -Achse in sich übergehen, erhalten wir ein zweischaliges Hyperboloid. Die beiden Schalen entstehen durch Rotation der beiden Hyperbeläste um die x_1' -Achse.

Wollen wir das neue Koordinatensystem und die Transformationsmatrix berechnen, müssen wir eine Orthonormalbasis von Eigenvektoren von A bestimmen. (Übung.)

Beispiel 9.4.9. Wir betrachten abschließend

$$7x_1^2 + 6x_2^2 + 5x_3^2 - 4x_1x_2 - 4x_2x_3 = 3. \quad (Q)$$

Zugehörige Matrix der Bilinearform: $A = \begin{pmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{pmatrix}$

Charakteristisches Polynom:

$$\begin{aligned}\det(A - X\mathbf{1}) &= \begin{vmatrix} 7-X & -2 & 0 \\ -2 & 6-X & -2 \\ 0 & -2 & 5-X \end{vmatrix} = (7-X)(6-X)(5-X) - 4(7-X) - 4(5-X) \\ &= (7-X)(6-X)(5-X) - 8(6-X) = (6-X)((7-X)(5-X) - 8) \\ &= (6-X)(X^2 - 12X + 27) = (6-X)(X-3)(X-9).\end{aligned}$$

Eigenwerte: $\lambda_1 = 3$, $\lambda_2 = 6$, $\lambda_3 = 9$.

Die Gleichung nach Hauptachsentransformation ist $3x_1'^2 + 6x_2'^2 + 9x_3'^2 = 3$ oder

$$x_1'^2 + 2x_2'^2 + 3x_3'^2 = 1.$$

Dies ist die Gleichung eines Ellipsoids mit den Hauptachsen

$$a = 1, \quad b = \frac{1}{\sqrt{2}}, \quad c = \frac{1}{\sqrt{3}}.$$

Die Hauptachsen erhält man durch die Berechnung (normierter) Eigenvektoren zu den Eigenwerten λ_1 , λ_2 , λ_3 :

$$v_1' = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad v_2' = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \quad v_3' = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}.$$

Die Transformationsmatrix ist $Q = \frac{1}{3} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix}.$

9.4.3 Quadratische Gleichungen mit linearen Termen

Wir betrachten nun allgemeiner eine Gleichung der Form

$$\sum_{i=1}^n a_{ii}x_i^2 + \sum_{1 \leq i < j \leq n} a_{ij}x_i x_j + \sum_{i=1}^n b_i x_i + c = 0, \quad (9.3)$$

die sich in Matrixschreibweise liest als

$$x^\top A x + b^\top x = c$$

mit einer symmetrischen Matrix A und einem Vektor b .

Beispiel 9.4.10. Für den Fall $n = 1$ reduziert sich diese Gleichung auf

$$ax^2 + bx + c = 0, \quad a, b, c \in \mathbb{R}.$$

Dies ist nur dann eine quadratische Gleichung, wenn $a \neq 0$ ist. Das nehmen wir an. Dann erhalten wir durch quadratische Ergänzung

$$ax^2 + bx = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a}.$$

Für $x' := x + \frac{b}{2a}$ ergibt sich also die rein quadratische Gleichung

$$(x')^2 + c - \frac{b^2}{4a} = 0.$$

Beispiel 9.4.11. Das Verfahren aus dem eindimensionalen Fall lässt sich wie folgt verallgemeinern. Wir betrachten zuerst eine Koordinatentransformation durch Translation: $x' := x + v$, $v \in \mathbb{R}^n$. Damit ergibt sich für den quadratischen und den linearen Term

$$\begin{aligned} x^\top A x + b^\top x &= (x' - v)^\top A (x' - v) + b^\top (x' - v) = (x')^\top A x' - 2(Av)^\top x' + b^\top x' + v^\top A v - b^\top v \\ &= (x')^\top A x' + (b - 2Av)^\top x' + v^\top A v - b^\top v. \end{aligned}$$

Nach Verschieben ändert sich der lineare Anteil also derart, dass wir b durch $b' := b - 2Av$ ersetzen müssen.

Ist die Matrix A invertierbar und $v := \frac{1}{2}A^{-1}b$, so erhalten wir eine Gleichung ohne linearen Term. Die Fläche ist also eine der oben betrachteten Flächen, allerdings um $-v$ verschoben. Das gleiche gilt, falls A nicht invertierbar ist, aber trotzdem $b \in \text{im}(A)$ gilt.

Im allgemeinen zerlegen wir b gemäß der orthogonalen Zerlegung $\mathbb{R}^n = \text{im}(A) \oplus \text{im}(A)^\perp$ in $b = 2Av + b'$ und dürfen nach der Transformation $x' := x + v$ annehmen, dass $b' \perp \text{im}(A)$ gilt.

Jetzt betrachten wir den allgemeinen Fall. Nach dem Satz über die Hauptachsentransformation finden wir eine ONB $(v_1, \dots, v_n)$ von $\mathbb{R}^n$, so dass (9.3) in diesbezüglichen Koordinaten die Gestalt

$$\sum_{i=1}^r a_i x_i^2 + \sum_{i=1}^n b_i x_i + c = 0, \quad (9.4)$$

annimmt, wobei $r = \text{rank}(A)$ der Rang von A ist und $a_1, \dots, a_r$ alle von 0 verschieden sind. Wie wir in Beispiel 9.4.10 gesehen haben, dürfen wir nach einer Verschiebung der Koordinaten annehmen, dass $b' \perp \text{im}(A)$ gilt, also $b_1 = \dots = b_r = 0$. Wir erhalten damit die einfachere Gleichung des Typs

$$\sum_{i=1}^r a_i x_i^2 + \sum_{i=r+1}^n b_i x_i + c = 0, \quad (9.5)$$

Ist $b = 0$, so ist dies schon eine rein quadratische Gleichung. Ist dies nicht der Fall, so ändern wir die ONB $(v_1, \dots, v_n)$ derart ab, dass $v'_{r+1} := \frac{1}{\|b\|} b$ wird. Da alle Vektoren $v_{r+1}, \dots, v_n$ im Kern

von A liegen, ändert sich dadurch nicht der quadratische Anteil der Gleichung, aber der lineare Anteil vereinfacht sich zu

$$\sum_{i=1}^r a_i x_i^2 + b x_{r+1} + c = 0, \quad b \in \mathbb{R}^\times. \quad (9.6)$$

Wir haben damit den folgenden Satz bewiesen.

Satz 9.4.12. *Durch eine affine isometrische Koordinatentransformation $x' = Qx + v$, $Q \in O_n(\mathbb{R})$ lässt sich die Gleichung $x^\top A x + b^\top x + c = 0$ zu einer Gleichung des folgenden Typs transformieren:*

- (a) $\sum_{i=1}^r a_i x_i^2 + c = 0$ mit $a_1, \dots, a_r \in \mathbb{R}^\times$ (falls $b \in \text{im}(A)$),
- (b) $\sum_{i=1}^r a_i x_i^2 + b x_{r+1} + c = 0$, $a_1, \dots, a_r, b \in \mathbb{R}^\times$ (falls $b \notin \text{im}(A)$).

Lassen wir auch Reskalierungen der Koordinatenachsen zu, so erhält man eine noch einfachere Form:

Satz 9.4.13. *Durch eine affine Koordinatentransformation $x' = Qx + v$, $Q \in GL_n(\mathbb{R})$ lässt sich die Gleichung $x^\top A x + b^\top x + c = 0$ zu einer Gleichung des folgenden Typs transformieren:*

- (a) $\sum_{i=1}^r \varepsilon_i x_i^2 + c = 0$ mit $\varepsilon_i \in \{\pm 1\}$ (falls $b \in \text{im}(A)$),
- (b) $\sum_{i=1}^r \varepsilon_i x_i^2 + \varepsilon_{r+1} x_{r+1} + c = 0$, $\varepsilon_j \in \{\pm 1\}$ (falls $b \notin \text{im}(A)$).

9.5 Definitheit von Formen und von Matrizen

Sei V ein reeller Vektorraum, $F: V \times V \rightarrow \mathbb{R}$ eine symmetrische Bilinearform auf V und $Q: V \rightarrow \mathbb{R}$ die zugehörige quadratische Form, also $Q(v) = F(v, v)$. Insbesondere ist dann $Q(0) = F(0, 0) = 0$.

Definition 9.5.1. Die quadratische Form Q und entsprechend die Bilinearform F heißt

- *positiv definit*, wenn $Q(x) > 0$ für alle $x \neq 0$;
- *negativ definit*, wenn $Q(x) < 0$ für alle $x \neq 0$;
- *positiv semidefinit*, wenn $Q(x) \geq 0$ für alle x ;
- *negativ semidefinit*, wenn $Q(x) \leq 0$ für alle x ;
- *indefinit*, wenn Q keine der obigen Eigenschaften hat, d.h. wenn Q positive und negative Werte hat.

Nun sei A eine symmetrische $n \times n$ -Matrix mit reellen Einträgen. Diese definiert eine symmetrische Bilinearform auf $\mathbb{R}^n$ durch

$$F(x, y) := x^\top A y \quad \text{für alle } x, y \in \mathbb{R}^n$$

und die zugehörige quadratische Form:

$$Q(x) := x^\top A x \quad \text{für alle } x \in \mathbb{R}^n.$$

Wir nennen eine Matrix A *positiv [negativ] (semi-)definit* oder *indefinit*, wenn die zugehörige quadratische Form diese Eigenschaften hat.

Bemerkung 9.5.2. In Kapitel 7 haben wir einen euklidischen Vektorraum definiert als einen reellen Vektorraum V mit einer bilinearen, symmetrischen, positiv definiten Form. Ist also eine symmetrische Matrix A positiv definit, so definiert

$$F(x, y) = x^\top A y = \langle x, A y \rangle = \langle A x, y \rangle \quad \text{für } x, y \in \mathbb{R}^n,$$

ein Skalarprodukt auf $\mathbb{R}^n$, das vom Standardskalarprodukt verschieden ist, falls $A \neq \mathbf{1}_n$ ist. Umgekehrt ist jedes Skalarprodukt auf $\mathbb{R}^n$ auf diese Weise durch eine positiv definite symmetrische Matrix definiert.

Aufgabe 9.5.3. Sei B eine $n \times n$ -Matrix mit reellen Einträgen. Zeige, dass die Matrix $A := B^\top B$ symmetrisch und positiv semidefinit ist. Ist B invertierbar, so ist A sogar positiv definit.

Beispiele 9.5.4. Sei $V = \mathbb{R}^2$ und $Q(x) = \lambda_1 x_1^2 + \lambda_2 x_2^2$.

- (a) $\lambda_1 > 0, \lambda_2 > 0$: Q ist *positiv definit*.

Die durch die Gleichung

$$x_3 = Q(x)$$

gegebene Fläche nennt man ein *elliptisches Paraboloid*, denn die Schnitte mit den Ebenen, die die x_3 -Achse enthalten, sind jeweils Parabeln und die Schnitte mit den Ebenen parallel zur x_3 -Achse sind Ellipsen.

- (b) $\lambda_1 < 0, \lambda_2 < 0$: Q ist *negativ definit*.

Die durch die Gleichung

$$x_3 = Q(x)$$

gegebene Fläche ist wieder ein elliptisches Paraboloid (nach unten geöffnet).

- (c) $\lambda_1 > 0, \lambda_2 < 0$: Q ist *indefinit*.

In der Tat ist $Q \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \lambda_1 > 0$, $Q \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda_2 < 0$. Die durch die Gleichung

$$x_3 = Q(x)$$

gegebene Fläche ist ein *hyperbolisches Paraboloid*, denn die Schnitte mit den Ebenen, die die x_3 -Achse enthalten, sind Parabeln, wobei wir uns den Fall einer Geraden (für welche Ebene tritt er auf?) als entartete Parabel denken. Dagegen sind die Schnitte mit den Ebenen parallel zur x_3 -Ebene Hyperbeln.

- (d) $\lambda_1 > 0, \lambda_2 = 0$: Q ist *positiv semidefinit*.

Die durch die Gleichung

$$x_3 = Q(x) = \lambda_1 x_1^2$$

gegebene Fläche ist ein *parabolischer Zylinder*; man erhält ihn durch Ziehen der Parabel $x_3 = \lambda_1 x_1^2$ in der (x_1, x_3) -Ebene entlang der x_2 -Achse. In diesem Fall erhalten wir wieder Parabeln als Schnitte mit den Ebenen, die die x_3 -Achse enthalten, und für die Ebenen parallel zur x_3 -Ebene ergeben sich parallele Geradenpaare, die zu einer Geraden entarten können, oder auch zur leeren Menge.

- (e) $\lambda_1 < 0, \lambda_2 = 0$: Q ist *negativ semidefinit*.

Der parabolische Zylinder ist nun nach unten offen.

9.5.5. [Quadratische Formen auf $\mathbb{R}^2$] Wir betrachten auf $\mathbb{R}^2$ eine beliebige quadratische Form

$$Q(x) = ax_1^2 + 2bx_1x_2 + cx_2^2 \quad (Q)$$

(bzgl. der Standardbasis). Die zugehörige Matrix ist $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Mittels der Hauptachsentransformation finden wir eine Orthonormalbasis v'_1, v'_2 , so dass (Q) bzgl. der neuen Basis die Gestalt

$$Q(x') = \lambda_1 x_1'^2 + \lambda_2 x_2'^2$$

erhält, wobei λ_1 und λ_2 die Eigenwerte der Matrix A sind.

In 9.5.4 haben wir die quadratischen Formen dieses Typs klassifiziert. Man erkennt, dass die Definitheitseigenschaften von Q und A nur vom Vorzeichen der Eigenwerte von A abhängen:

Q ist positiv definit	$\iff$	die Eigenwerte λ_1, λ_2 von A sind beide > 0 ;
Q ist negativ definit	$\iff$	die Eigenwerte λ_1, λ_2 von A sind beide < 0 ;
Q ist indefinit	$\iff$	einer der Eigenwerte ist > 0 , der andere < 0 .
Q ist definit	$\iff$	$\lambda_1 \cdot \lambda_2 > 0$
und Q ist indefinit	$\iff$	$\lambda_1 \cdot \lambda_2 < 0$
Q ist positiv semidefinit	$\iff$	die Eigenwerte von A sind beide ≥ 0 ;
Q ist negativ semidefinit	$\iff$	die Eigenwerte von A sind beide ≤ 0 .

Da wir A mit Hilfe einer orthogonalen Übergangsmatrix S diagonalisieren können, wissen wir, dass

$$S^T A S = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

gilt und damit $\det A = \det(S^T A S) = \lambda_1 \cdot \lambda_2$ ist.

Wir folgern:

- Ist $\det A < 0$, so gilt $\lambda_1 \cdot \lambda_2 < 0$, d.h. Q ist indefinit.
- Ist $\det A > 0$, so gilt $\lambda_1 \cdot \lambda_2 > 0$, d.h. Q ist definit.

Um zu erkennen, ob positive oder negative Definitheit vorliegt, betrachten wir den Spezialfall $x_1 = 1, x_2 = 0$. Es folgt

$$Q(x) = ax_1^2 + 2bx_1x_2 + cx_2^2 = a$$

und damit $Q(x) \geq 0$ für alle x genau dann, wenn $a \geq 0$ ist; ebenso gilt $Q(x) \leq 0$ für alle x genau dann, wenn $a \leq 0$ ist.

Also ist Q im ersten Fall positiv definit und negativ definit im zweiten Fall. Beachte, dass $a = 0$ in diesem Fall nicht auftreten kann, da $a = 0$ schon $\det A = -b^2 \leq 0$ impliziert.

Insgesamt haben wir bewiesen:

Satz 9.5.6. Die quadratische Form $Q(x) = ax_1^2 + 2bx_1x_2 + cx_2^2$ ist

- *positiv definit* $\iff \det A = ac - b^2 > 0$ und $a > 0$;
- *negativ definit* $\iff \det A = ac - b^2 > 0$ und $a < 0$;
- *indefinit* $\iff \det A = ac - b^2 < 0$.

Im Folgenden verallgemeinern wir diese Betrachtungen von $n = 2$ auf beliebiges n ; die Bedeutung des Kriteriums besteht darin, dass wir damit die Definitheitseigenschaften einer Matrix bestimmen können, ohne die Eigenwerte berechnen zu müssen:

Hurwitzkriterium

Satz 9.5.7. [Kriterium für positive Definitheit] Sei $A = (a_{ij})$ eine symmetrische reelle $n \times n$ -Matrix. Dann sind die folgenden Eigenschaften äquivalent

- A ist positiv definit;
- $\lambda_i > 0$ für jeden Eigenwert λ_i von A ;
- $\det(A_r) > 0$ für $r = 1, 2, \dots, n$, wobei die Matrizen

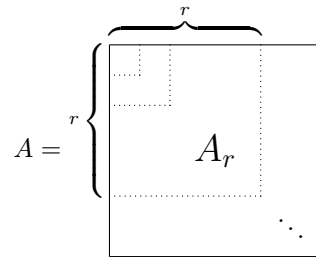
$$A_r := \begin{pmatrix} a_{11} & \dots & a_{1r} \\ \vdots & & \vdots \\ a_{r1} & \dots & a_{rr} \end{pmatrix}$$

die Hauptminoren von A sind.

Analog sind die folgenden Eigenschaften äquivalent:

- (a') A ist negativ definit;
- (b') $\lambda_i < 0$ für alle Eigenwerte λ_i von A ;
- (c') $(-1)^r \det(A_r) > 0$ für $r = 1, 2, \dots, n$.

Wir können die Hauptminoren schematisch wie folgt darstellen:



Beweis. Die Äquivalenz von (a'), (b') und (c') folgt aus der Äquivalenz von (a), (b) und (c) durch Ersetzung von A mit $-A$, denn A ist genau dann negativ definit, wenn $-A$ positiv definit ist.

Wir zeigen nun, dass (a) $\iff$ (b), (a) $\implies$ (c) und (c) $\implies$ (b) gilt. Dies impliziert alle obigen Äquivalenzen.

- (a) $\iff$ (b) Die Matrix A ist symmetrisch und damit diagonalisierbar. Mittels der Hauptachsentransformation können wir eine Orthonormalbasis von $\mathbb{R}^n$ finden, so dass bzgl. der Koordinaten x'_i in dieser neuen Basis gilt:

$$Q(x) := x^\top A x = \lambda_1 x_1'^2 + \dots + \lambda_n x_n'^2.$$

Dabei sind die $\lambda_1, \lambda_2, \dots, \lambda_n$ die Eigenwerte von A . In dieser Form erkennt man direkt, dass $Q(x)$ genau dann positiv definit ist, wenn alle λ_i positiv sind.

- (a) $\implies$ (c) Ist A symmetrisch und positiv definit, so sind auch alle Matrizen A_r symmetrisch und positiv definit. Die Werte von $y^\top A_r^\top y$, $y \in \mathbb{R}^r$, erhält man durch Auswertung von $Q(x)$ auf Vektoren $x \in \mathbb{R}^n$, deren letzte $n - r$ Komponenten verschwinden.

Damit ist jede Matrix A_r orthogonal diagonalisierbar zu

$$A'_r = Q^\top A_r Q = Q^{-1} A_r Q = \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_r \end{pmatrix},$$

wobei die Werte auf der Diagonalen die Eigenwerte von μ_r sind. Mittels obiger Begründung ergibt sich, dass alle μ_i positiv sind, da A_r positiv definit ist.

Die Determinante ist invariant unter Basiswechsel:

$$\begin{aligned} \det A_r &= \det(Q A'_r Q^{-1}) = \det Q \cdot \det A'_r \cdot \det Q^{-1} = \det A'_r \\ &= \mu_1 \cdot \mu_2 \cdot \dots \cdot \mu_r > 0. \end{aligned}$$

Damit gilt $\det A_r > 0$ für alle Hauptminoren.

- (c) $\implies$ (b) Wir beweisen mittels Induktion nach n , dass eine $n \times n$ -Matrix A mit der Eigenschaft, dass die Determinanten ihrer Hauptminoren echt positiv sind, positiv definit ist. Für $n = 1$ ist die Matrix A eine einfache Zahl, die mit ihrer Determinante übereinstimmt. Wir nehmen $n > 0$ an und dass die Implikation für Matrizen der Größe $r < n$ gilt. Die Hauptminoren der Matrix A_{n-1} sind unter den Hauptminoren von A , so dass nach der Induktionsannahme

A_{n-1} positiv definit ist. Das heißt aber, dass die zugehörige symmetrische Bilinearform F_A , eingeschränkt auf den Unterraum

$$U = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \mid x_n = 0 \right\},$$

positiv definit ist.

Wir haben zu zeigen, alle Eigenwerte $\lambda_1, \dots, \lambda_n$ von A positiv sind. Wegen $0 < \det(A) = \lambda_1 \cdots \lambda_n$ sind sie alle $\lambda_i \neq 0$. Hierzu nehmen wir an, dass ein Eigenwert $\lambda_i < 0$ ist. Dann existiert wegen $\det A > 0$ ein zweiter Eigenwert $\lambda_j < 0$. Also ist die Summe W der Eigenräume zu den negativen Eigenwerten mindestens zweidimensional.

Eingeschränkt auf W ist die Form F_A negativ definit. Dies folgt aus der Implikation b) $\Rightarrow$ a), da die Einschränkung von A auf W nur nichtpositive Eigenwerte besitzt. Ist $x \in U \cap W$ und $x \neq 0$, so folgt $x^\top A x > 0$ und $x^\top A x < 0$, was nicht möglich ist, also ist $U \cap W = \{0\}$, d.h. die Summe von U und W ist direkt. Aus $\text{codim}(U) = 1$ folgt damit insbesondere $\dim W \leq 1$, im Widerspruch zu $\dim W \geq 2$. $\square$

Beispiel 9.5.8. (für $n = 4$) Betrachte

$$Q(x) = 3x_1^2 + 2x_2^2 + 2x_3^2 + 6x_4^2 + 4x_1x_2 - 2x_1x_3 - 2x_3x_4.$$

Die zugehörige Matrix ist

$$A = \begin{pmatrix} 3 & 2 & -1 & 0 \\ 2 & 2 & 0 & 0 \\ -1 & 0 & 2 & -1 \\ 0 & 0 & -1 & 6 \end{pmatrix}$$

und die Hauptminoren sind

$$\det A_1 = 3, \quad \det A_2 = \begin{vmatrix} 3 & 2 \\ 2 & 2 \end{vmatrix} = 2, \quad \det A_3 = \begin{vmatrix} 3 & 2 & -1 \\ 2 & 2 & 0 \\ -1 & 0 & 2 \end{vmatrix} = 2, \quad \det A_4 = \det A = 10.$$

Also ist $\det A_r > 0$ für $r = 1, 2, 3, 4$. Nach obigem Kriterium ist $Q(x)$ positiv definit.

Satz 9.5.9. [Die Choleskyzerlegung] Für jede positiv definite, reellwertige, symmetrische Matrix A gibt es eine obere reelle Dreiecksmatrix R mit

$$A = R^\top R$$

und positiven Diagonaleinträgen $r_{ii} > 0$.

Beweis. Da A reell und symmetrisch ist, gibt es eine orthogonale Matrix S mit

$$S^\top A S = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

wobei $\lambda_1, \dots, \lambda_n$ die Eigenwerte von A sind.

Da A positiv definit ist, gilt $\lambda_i > 0$ für alle i nach 9.5.7. Wir können die Matrix

$$D = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix} \quad \text{bilden mit} \quad D \cdot D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Also gilt $S^\top AS = D \cdot D$ und damit $A = S \cdot D \cdot D \cdot S^\top$. Die Matrix $B := D \cdot S^\top$ ist invertierbar, d.h. die Spalten sind linear unabhängig. Ihre Q - R -Zerlegung ist $B = Q \cdot R$, wobei Q eine orthogonale Matrix ist und R obere Dreiecksgestalt hat, wobei die Diagonaleinträge positiv sind. Weiter gilt $B^\top = (DS^\top)^\top = S^\top D^\top = SD$. Also folgt

$$A = (SD)(DS^\top) = B^\top B = (QR)^\top QR = R^\top \underbrace{Q^\top Q}_E R = R^\top R,$$

was die gewünschte Zerlegung von A liefert. $\square$

Beispiel 9.5.10. Für praktische Zwecke berechnet man die Choleskyzerlegung nicht wie im Beweis. Wir zeigen das Vorgehen an einem Beispiel:

$$A = \begin{pmatrix} 4 & -2 & 4 & 2 \\ -2 & 2 & -3 & -3 \\ 4 & -3 & 6 & 2 \\ 2 & -3 & 2 & 13 \end{pmatrix}$$

Zeige, dass A in der Tat positiv definit ist (gehe wie in 9.5.8 vor). Nach Satz 9.5.9 gibt es eine obere Dreiecksmatrix R mit

$$A = R^\top \cdot R = \begin{pmatrix} r_{11} & 0 & 0 & 0 \\ r_{12} & r_{22} & 0 & 0 \\ r_{13} & r_{23} & r_{33} & 0 \\ r_{14} & r_{24} & r_{34} & r_{44} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ 0 & r_{22} & r_{23} & r_{24} \\ 0 & 0 & r_{33} & r_{34} \\ 0 & 0 & 0 & r_{44} \end{pmatrix}$$

Wir berechnen nun dieses Produkt Zeile für Zeile und bestimmen schrittweise r_{ij} unter Benutzung der bereits berechneten Werte:

- Erste Zeile: $r_{11}r_{11} = 4$, $r_{11}r_{12} = -2$, $r_{11}r_{13} = 4$, $r_{11}r_{14} = 2$.
Also gilt $r_{11} = \sqrt{4} = 2$, $r_{12} = -2/2 = -1$, $r_{13} = 4/2 = 2$, $r_{14} = 2/2 = 1$.
- Zweite Zeile: $r_{12}^2 + r_{22}^2 = 2$, $r_{12}r_{13} + r_{22}r_{23} = -3$, $r_{12}r_{14} + r_{22}r_{24} = -3$.
Also gilt $r_{22} = \sqrt{2 - (-1)^2} = 1$, $r_{23} = (-3 - (-1) \cdot 2)/1 = -1$ und $r_{24} = (-3 - (-1) \cdot 1)/1 = -2$.
- Dritte Zeile: $r_{13}^2 + r_{23}^2 + r_{33}^2 = 6$, $r_{13}r_{14} + r_{23}r_{24} + r_{33}r_{34} = 2$.
Also gilt $r_{33} = \sqrt{6 - 2^2 - (-1)^2} = 1$, $r_{34} = (2 - 2 \cdot 1 - (-1) \cdot (-2))/1 = -2$.
- Vierte Zeile: $r_{14}^2 + r_{24}^2 + r_{34}^2 + r_{44}^2 = 13$.
Es folgt $r_{44} = \sqrt{13 - 1^2 - (-2)^2 - (-2)^2} = 2$.

Wir erhalten $R = \begin{pmatrix} 2 & -1 & 2 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 2 \end{pmatrix}$. (Rechnen Sie $R^\top R = A$ nach.)

9.5.11. [Anwendung auf lineare Gleichungssysteme] Wir betrachten ein System linearer Gleichungen

$$Ax = b \tag{L}$$

Unter der Annahme, A sei *positiv definit* und *symmetrisch*, können wir eine obere Dreiecksmatrix R finden, so dass $A = R^\top R$ ist. Anstatt (L) direkt zu lösen, können wir die beiden Systeme

$$R^\top y = b \tag{1}$$

$$Rx = y \tag{2}$$

nacheinander lösen. Es ist leicht, solche Systeme zu lösen, da beide Matrizen R und $R^\top$ Dreiecksmatrizen sind.

Beispiel 9.5.12. Sei A wie in 9.5.10 und $b = \begin{pmatrix} -2 \\ 4 \\ 2 \\ -5 \end{pmatrix}$.

Um (L) $Ax = b$ zu lösen, lösen wir nacheinander

$$R^\top y = b, \text{ d.h.} \quad \begin{array}{ll} 2y_1 = -2 & y_1 = -1 \end{array} \quad (1)$$

$$-y_1 + y_2 = 4 \quad y_2 = 3$$

$$2y_1 - y_2 + y_3 = 2 \quad y_3 = 7$$

$$y_1 - 2y_2 - 2y_3 + 2y_4 = -5 \quad y_4 = 8$$

$$Rx = y, \text{ d.h.} \quad \begin{array}{ll} 2y_4 = 8 & y_4 = 4 \end{array} \quad (2)$$

$$y_3 - 2y_4 = 7 \quad y_3 = 15$$

$$y_2 - y_3 - 2y_4 = 3 \quad y_2 = 26$$

$$2y_1 - y_2 + 2y_3 + y_4 = -1 \quad y_1 = -9/2$$

Dieses Verfahren funktioniert nur, wenn die Koeffizientenmatrix A symmetrisch und positiv definit ist. Zur allgemeineren Verwendbarkeit benutzen wir folgenden Trick:

Angenommen, A sei invertierbar. Aus Aufgabe 9.5.13 wissen wir, dass $B := A^\top A$ positiv definit und symmetrisch ist. Um

$$Ax = b \quad (L)$$

zu lösen, sei $B = A^\top A$ und $c = A^\top b$ und man betrachte das äquivalente System

$$A^\top Ax = A^\top b \quad \text{oder} \quad Bx = c, \quad (L')$$

welches dieselbe Lösung hat. Auf dieses System wenden wir obige Prozedur an.

Natürlich kann (L) mit dem Gauß-Jordan-Algorithmus gelöst werden. Allerdings gibt es Matrizen, bei denen der Gauß-Jordan-Algorithmus aufgrund großer Rundungsfehler nicht praktikabel ist. In solchen Fällen ist die Choleskyzerlegung oft wesentlich besser.

Aufgaben zu Abschnitt 9.5

Beispiel 9.5.13. Zeigen Sie: Ist $B \in M_n(\mathbb{R})$, so ist die Matrix $A := B^\top B$ positiv semidefinit. Ist B sogar invertierbar, so ist A positiv definit.

Kapitel 10

Tensorprodukte

In diesem Kapitel bezeichnet $\mathbb{K}$ einen beliebigen Körper.

10.1 Tensorprodukte von Vektorräumen

10.1.1 Tensorprodukte von zwei Vektorräumen

Definition 10.1.1. Seien V und W zwei $\mathbb{K}$ -Vektorräume. Ein paar (η, U) aus einem Vektorraum U und einer bilinearen Abbildung $\eta: V \times W \rightarrow U$ heißt *Tensorprodukt von V und W* , wenn es folgende universelle Eigenschaft besitzt. Zu jeder bilinearen Abbildung

$$f: V \times W \rightarrow X$$

in einen Vektorraum X existiert genau eine lineare Abbildung $\tilde{f}: U \rightarrow X$ mit $\tilde{f} \circ \eta = f$.

Lemma 10.1.2. (Eindeutigkeit von Tensorprodukten) *Sind (η, U) und (η', U') zwei Tensorprodukte von V und W , so existiert ein linearer Isomorphismus $\Gamma: U \rightarrow U'$ mit $\Gamma \circ \eta = \eta'$.*

Beweis. Da die Abbildung $\eta': V \times W \rightarrow U'$ bilinear ist, existiert gemäß der universellen Eigenschaft von (η, U) genau eine lineare Abbildung $\alpha: U \rightarrow U'$ mit $\alpha \circ \eta = \eta'$. Ebenso finden wir mit der universellen Eigenschaft von (η', U') eine lineare Abbildung $\beta: U' \rightarrow U$ mit $\beta \circ \eta' = \eta$. Dann ist $\beta \circ \alpha: U \rightarrow U$ eine lineare Abbildung mit

$$\beta \circ \alpha \circ \eta = \beta \circ \eta' = \eta = \text{id}_U \circ \eta.$$

Aus der Eindeutigkeit in der universellen Eigenschaft von (η, U) folgt daher $\beta \circ \alpha = \text{id}_U$. Analog erhalten wir $\alpha \circ \beta = \text{id}_{U'}$. Also ist $\alpha: U \rightarrow U'$ ein linearer Isomorphismus mit $\alpha^{-1} = \beta$. $\square$

Das obige Lemma besagt, dass Tensorprodukte bis auf Isomorphie eindeutig bestimmt sind. Nach wir die Eindeutigkeit des Tensorprodukts eingesehen haben, beweisen wir nun seine Existenz.

Definition 10.1.3. Sei M eine Menge. Wir schreiben $F(M) := \mathbb{K}^{(M)}$ für den *freien Vektorraum über M* . Das ist der Unterraum des kartesischen Produkts $\mathbb{K}^M$, der aus den Elementen $(z_j)_{j \in M}$ besteht, für die nur endlich viele Komponenten z_j von Null verschieden sind.

Lemma 10.1.4. *Die Abbildung $\eta_M: M \rightarrow F(M)$ mit $\eta_M(j) = \delta_j := (\delta_{ij})_{i \in M}$ hat die folgende universelle Eigenschaft. Zu jeder Abbildung $f: M \rightarrow V$ in einen Vektorraum V existiert genau eine lineare Abbildung $F(f): F(M) \rightarrow V$ mit $F(f) \circ \eta_M = f$.*

Beweis. Die Eindeutigkeit folgt aus der Tatsache, dass $F(M)$ von $\eta_M(M)$ als Vektorraum erzeugt wird. Die Existenz folgt aus der Beobachtung, dass $\eta_M(M)$ eine Basis von $F(M)$ ist. $\square$

Im folgenden werden wir die Abbildung η_M unterdrücken und die Elemente $m \in M$ mit den entsprechenden Elementen $\eta_M(m) = \delta_m$ von $F(M)$ identifizieren. Andernfalls würde die Notation schnell sehr schwerfällig werden.

Definition 10.1.5. Seien V und W Vektorräume. In dem freien Vektorraum $F(V \times W)$ betrachten wir den Unterraum U , der von allen Elementen der Gestalt

$$(v_1 + v_2, w) - (v_1, w) - (v_2, w), \quad (v, w_1 + w_2) - (v, w_1) - (v, w_2),$$

und

$$(\lambda v, w) - (v, \lambda w), \quad \lambda(v, w) - (\lambda v, w)$$

erzeugt wird. Wir definieren

$$V \otimes W := F(V \times W)/U \quad \text{und} \quad v \otimes w := (v, w) + U.$$

Lemma 10.1.6. Für die Abbildung $b: V \times W \rightarrow V \otimes W, (v, w) \mapsto v \otimes w$ ist $(b, V \otimes W)$ ein Tensorprodukt von V und W .

Beweis. Die Bilinearität von b folgt aus der Definition von $V \otimes W$. Z.B. ist

$$b(v_1 + v_2, w) = (v_1 + v_2, w) + U = (v_1, w) + (v_2, w) + U = b(v_1, w) + b(v_2, w)$$

und

$$b(\lambda v, w) = (\lambda v, w) + U = \lambda(v, w) + U.$$

Ebenso folgt die Linearität im zweiten Argument.

Sei nun $f: V \times W \rightarrow X$ eine bilineare Abbildung. Da $F(V \times W)$ von der Teilmenge $V \times W$ erzeugt wird (wir erinnern uns daran, dass wir $V \times W$ als eine Teilmenge von $F(V \times W)$ auffassen), wird der Quotientenraum $V \otimes W$ von $b(V \times W) = (V \times W) + U$ erzeugt. Hieraus folgt sofort die Eindeutigkeit von $\tilde{f}$.

Um die Existenz von $\tilde{f}$ einzusehen, verwenden wir zuerst die universelle Eigenschaft von $F(V \times W)$ (Lemma 10.1.4), um eine lineare Abbildung $F(f): F(V \times W) \rightarrow X$ mit $F(f)(v, w) = f(v, w)$ für alle $(v, w) \in V \times W$ zu erhalten. Aus der Bilinearität von f folgt nun $U \subseteq \ker F(f)$. Also faktorisiert $F(f)$ zu einer linearen Abbildung $\tilde{f}: V \otimes W \rightarrow X$ mit $\tilde{f}(v \otimes w) = f(v, w)$ für alle $(v, w) \in V \times W$. $\square$

Bemerkung 10.1.7. In Aufgabe 10.1.14 wird gezeigt, wie man eine Basis des Tensorprodukts $V \otimes W$ erhält. Dazu sei $B_V = \{e_i: i \in I\}$ bzw. $B_W = \{f_j: j \in J\}$ eine Basis von V bzw. W . Dann ist

$$B_V \otimes B_W = \{e_i \otimes f_j: i \in I, j \in J\}$$

eine Basis des Vektorraums $V \otimes W$. Sind V und W endlichdimensional, so erhalten wir insbesondere die Formel

$$\dim(V \otimes W) = \dim V \cdot \dim W.$$

Bemerkung 10.1.8. (a) Seien X, Y und Z Vektorräume. Dann erhalten wir eine Abbildung

$$\Gamma: \text{Hom}(X \otimes Y, Z) \rightarrow \text{Bil}(X \times Y, Z), \quad \Gamma(\varphi)(x, y) := \varphi(x \otimes y),$$

wobei $\text{Bil}(X \times Y, Z)$ den Raum der bilinearen Abbildungen von $X \times Y$ nach Z bezeichnet. Die universelle Eigenschaft des Tensorprodukts bedeutet, dass diese Abbildung Γ bijektiv ist. Sie übersetzt bilinear in lineare Abbildungen. Da Γ bzgl. der kanonischen Vektorraumstrukturen linear ist, ist Γ sogar ein linearer Isomorphismus.

Für den Spezialfall $Z = \mathbb{K}$ erhalten wir so einen Isomorphismus

$$\Gamma: (X \otimes Y)^* \rightarrow \text{Bil}(X \times Y, \mathbb{K}), \quad \Gamma(\alpha)(x, y) := \alpha(x \otimes y).$$

(b) Von besonderem Interesse ist auch die Beziehung zwischen dem Tensorprodukt $X^* \otimes Y$ und dem Raum $\text{Hom}(X, Y)$ der linearen Abbildungen von X nach Y . Wir beobachten zunächst, dass wir eine natürliche lineare Abbildung

$$\Gamma: X^* \otimes Y \rightarrow \text{Hom}(X, Y), \quad \Gamma(\alpha \otimes y)(x) := \alpha(x)y$$

haben. Um die Existenz dieser Abbildung einzusehen, verwenden wir die universelle Eigenschaft des Tensorprodukts. Dazu muss man zeigen, dass die Abbildung

$$\tilde{\Gamma}: X^* \times Y \rightarrow \text{Hom}(X, Y), \quad \tilde{\Gamma}(\alpha, y)(x) := \alpha(x)y$$

bilinear ist, was sich aus einer trivialen Rechnung ergibt.

Das Tensorprodukt $X^* \otimes Y$ besteht ja aus endlichen Summen der Gestalt $A := \sum_{j=1}^k \alpha_j \otimes y_j$. Für die zugeordnete lineare Abbildung $\Gamma(A)$ ist dann $\text{im}(\Gamma(A)) \subseteq \text{span}\{y_1, \dots, y_k\}$. Insbesondere ist das Bild von $\Gamma(A)$ immer endlichdimensional. Ist $X = Y$ unendlichdimensional, so liegt id_X nicht im Bild von Γ . Insbesondere ist Γ nicht surjektiv.

Wir nehmen nun an, dass X und Y endlichdimensional sind. Dann ist zunächst

$$\dim(X^* \otimes Y) = \dim X^* \cdot \dim Y = \dim X \dim Y = \dim \text{Hom}(X, Y).$$

Ist $b_1, \dots, b_n$ eine Basis von X , $b_1^*, \dots, b_n^*$ die duale Basis von X^* und $c_1, \dots, c_m$ eine Basis von Y , so erhalten wir

$$\Gamma(b_j^* \otimes c_i) \left(\sum_{k=1}^n x_k b_k \right) = b_j^* \left(\sum_{k=1}^n x_k b_k \right) c_i = x_j c_i.$$

Die Matrix der linearen Abbildung $\Gamma(b_j^* \otimes c_i)$ ist also die Basismatrix $[\Gamma(b_j^* \otimes c_i)]_B^C = E_{ij} = (\delta_{ii'} \delta_{jj'})$. Insbesondere sehen wir, dass die Basis $b_j^* \otimes c_i$ in eine Basis von $\text{Hom}(X, Y) \cong M_{m,n}(\mathbb{K}) \cong \mathbb{K}^{m \times n}$ abgebildet wird. Also liefert Γ in diesem Fall einen Isomorphismus

$$X^* \otimes Y \cong \text{Hom}(X, Y).$$

Für $X = Y$ und $b_i = c_i$ ergibt sich insbesondere

$$\Gamma\left(\sum_{j=1}^n b_j^* \otimes b_j\right) = \text{id}_X,$$

denn die zugehörige Matrix ist $\mathbf{1}_n$.

In diesem Fall erhalten wir also insbesondere einen Isomorphismus

$$\text{End}(X) \cong X^* \otimes X.$$

Beispiel 10.1.9. Die Abbildung

$$\Gamma: \mathbb{K}^m \times \mathbb{K}^n \rightarrow M_{m,n}(\mathbb{K}), \quad (x, y) \mapsto x \cdot y^\top$$

ist bilinear, induziert also eine lineare Abbildung

$$\tilde{\Gamma}: \mathbb{K}^m \otimes \mathbb{K}^n \rightarrow M_{m,n}(\mathbb{K}), \quad x \otimes y \mapsto x \cdot y^\top.$$

Für die Standardbasis erhalten wir $\Gamma(e_i \otimes e_j) = E_{ij}$, so dass $\tilde{\Gamma}$ insbesondere ein linearer Isomorphismus ist.

10.1.2 Tensorprodukte von endlich vielen Vektorräumen

Definition 10.1.10. Sind $V_1, \dots, V_n$ Vektorräume und $n \geq 3$, so definieren wir rekursiv

$$V_1 \otimes \dots \otimes V_n := (V_1 \otimes \dots \otimes V_{n-1}) \otimes V_n$$

und analog

$$v_1 \otimes \dots \otimes v_n := (v_1 \otimes \dots \otimes v_{n-1}) \otimes v_n.$$

Lemma 10.1.11. Sind $V_1, \dots, V_n$ Vektorräume, so ist

$$m: V_1 \times \dots \times V_n \rightarrow V_1 \otimes \dots \otimes V_n, \quad (v_1, \dots, v_n) \mapsto v_1 \otimes \dots \otimes v_n$$

eine n -lineare Abbildung, d.h. linear in jedem Argument. Zu jeder n -linearen Abbildung

$$f: V_1 \times \dots \times V_n \rightarrow X$$

in einen Vektorraum X existiert genau eine lineare Abbildung

$$\tilde{f}: V_1 \otimes \dots \otimes V_n \rightarrow X$$

mit $\tilde{f} \circ m = f$.

Beweis. Zuerst zeigen wir die n -Linearität von $m_n := m$. Wir setzen hierzu

$$m_{n-1}: V_1 \times \dots \times V_{n-1} \rightarrow V_1 \otimes \dots \otimes V_{n-1}, \quad (v_1, \dots, v_{n-1}) \mapsto v_1 \otimes \dots \otimes v_{n-1}.$$

Ist $n = 2$, so ist nichts mehr zu zeigen (Lemma 10.1.6). Wir dürfen daher induktiv annehmen, dass die Abbildung m_{n-1} in jedem Argument linear ist. Nach Lemma 10.1.6 ist die Abbildung

$$b: (V_1 \otimes \dots \otimes V_{n-1}) \times V_n \rightarrow (V_1 \otimes \dots \otimes V_{n-1}) \otimes V_n, \quad (x, w) \mapsto x \otimes w$$

bilinear. Daher ist

$$m_n(v_1, \dots, v_{n-1}, v_n) = b(m_{n-1}(v_1, \dots, v_{n-1}), v_n)$$

in den ersten $n-1$ Argumenten als Komposition von zwei linearen Abbildungen linear. Im letzten Argument ist sie nach Lemma 10.1.6 linear, d.h. m_n ist n -linear.

Die Existenz und Eindeutigkeit von $\tilde{f}$ zeigt man ebenfalls induktiv. Wir dürfen daher annehmen, dass zu jedem $w \in V_n$ genau eine lineare Abbildung $\tilde{f}_w: V_1 \otimes \dots \otimes V_{n-1} \rightarrow X$ mit

$$\tilde{f}_w(v_1 \otimes \dots \otimes v_{n-1}) = f(v_1, \dots, v_{n-1}, w)$$

existiert, denn wenn wir das letzte Argument $v_n = w$ fixieren, wird f zu einer $(n-1)$ -linearen Abbildung. Wegen der Eindeutigkeit von $\tilde{f}_w$ für jedes w ist die Zuordnung

$$V_n \rightarrow \text{Hom}(V_1 \otimes \dots \otimes V_{n-1}, X), \quad w \mapsto \tilde{f}_w$$

linear, d.h. die Abbildung

$$(V_1 \otimes \dots \otimes V_{n-1}) \times V_n \rightarrow X, \quad (x, w) \mapsto \tilde{f}_w(x)$$

ist bilinear. Mit Lemma 10.1.6 finden wir daher eine eindeutig bestimmte lineare Abbildung

$$\tilde{f}: (V_1 \otimes \dots \otimes V_{n-1}) \otimes V_n = V_1 \otimes \dots \otimes V_{n-1} \otimes V_n \rightarrow X$$

mit

$$\begin{aligned} \tilde{f}(v_1 \otimes \dots \otimes v_{n-1} \otimes v_n) &= \tilde{f}((v_1 \otimes \dots \otimes v_{n-1}) \otimes v_n) \\ &= \tilde{f}_{v_n}(v_1 \otimes \dots \otimes v_{n-1}) = f(v_1, \dots, v_{n-1}, v_n). \end{aligned}$$

□

Definition 10.1.12. Sei V ein $\mathbb{K}$ -Vektorraum, $V^{\otimes n}$ das n -fache Tensorprodukt von V mit sich selbst, $V^{\otimes 0} = \mathbb{K}$ und $V^{\otimes 1} = V$. Sind $n, k \in \mathbb{N}_0$, so existiert genau eine bilineare Abbildung $m_{n,k}: V^{\otimes n} \times V^{\otimes k} \rightarrow V^{\otimes(n+k)}$ mit

$$m_{n,k}((v_1 \otimes \dots \otimes v_n), (v_{n+1} \otimes \dots \otimes v_{n+k})) = v_1 \otimes \dots \otimes v_{n+k}$$

für $v_1, \dots, v_{n+k} \in V$. In der Tat ist für jedes $x \in V^{\otimes n}$ die Abbildung

$$m_x: V^k \rightarrow V^{\otimes(n+k)}, \quad (w_1, \dots, w_k) \mapsto x \otimes w_1 \otimes \dots \otimes w_k$$

nach Lemma 10.1.11 eine k -lineare Abbildung, und wir erhalten eine eindeutig bestimmte lineare Abbildung

$$\tilde{m}_x: V^{\otimes k} \rightarrow V^{\otimes(n+k)} \quad \text{mit} \quad \tilde{m}_x(w_1 \otimes \dots \otimes w_k) = m_x(w_1, \dots, w_k).$$

Aus Lemma 10.1.11, angewandt auf $V^{\otimes n} \otimes V \otimes \dots \otimes V$, folgt, dass diese Abbildung ebenfalls linear in x ist. Daher erhalten wir eine eindeutig bestimmte bilineare Abbildung $m_{n,k}: V^{\otimes n} \times V^{\otimes k} \rightarrow V^{\otimes(n+k)}$ mit

$$\begin{aligned} & m_{n,k}((v_1 \otimes \dots \otimes v_n), (v_{n+1} \otimes \dots \otimes v_{n+k})) \\ &= \tilde{m}_{(v_1 \otimes \dots \otimes v_n)}(v_{n+1} \otimes \dots \otimes v_{n+k}) = v_1 \otimes \dots \otimes v_n \otimes v_{n+1} \otimes \dots \otimes v_{n+k}. \end{aligned}$$

Weiter definieren wir Abbildungen

$$m_{0,n}: V^{\otimes 0} \times V^{\otimes n} = \mathbb{K} \times V^{\otimes n} \rightarrow V^{\otimes n}, \quad (\lambda, v) \mapsto \lambda v$$

und

$$m_{n,0}: V^{\otimes n} \otimes V^{\otimes 0} = V^{\otimes n} \times \mathbb{K} \rightarrow V^{\otimes n}, \quad (v, \lambda) \mapsto \lambda v.$$

Setzen wir schließlich alle Abbildungen $m_{n,k}$, $n, k \in \mathbb{N}_0$, zusammen, so erhalten wir eine Multiplikation auf dem Vektorraum $\mathcal{T}(V) = \bigoplus_{n=0}^{\infty} V^{\otimes n}$. Die so erhaltene Algebra ist assoziativ und besitzt ein Einselement $\mathbf{1} \in V^{\otimes 0}$ (Nachweis!). Sie heißt *Tensoralgebra* über V .

Lemma 10.1.13. (Universelle Eigenschaft der Tensoralgebra) *Ist A eine assoziative $\mathbb{K}$ -Algebra mit Einselement $\mathbf{1}_A$ und V ein Vektorraum, so existiert zu jeder linearen Abbildung $f: V \rightarrow A$ genau ein Algebra-Homomorphismus $\tilde{f}: \mathcal{T}(V) \rightarrow A$ mit $\tilde{f}(\mathbf{1}) = \mathbf{1}_A$ und $\tilde{f}(v) = f(v)$ für alle $v \in V$.*

Beweis. Für die Eindeutigkeit bemerken wir zunächst, dass aus der Forderung, dass $\tilde{f}$ ein Algebra-Homomorphismus ist, die Beziehung

$$\tilde{f}(v_1 \otimes \dots \otimes v_n) = \tilde{f}(v_1) \cdots \tilde{f}(v_n) = f(v_1) \cdots f(v_n)$$

folgt. Damit ist die Eindeutigkeit gezeigt.

Für die Existenz fixieren wir zunächst ein $n \in \mathbb{N}$. Dann ist die Abbildung

$$f_n: V^n \rightarrow A, \quad (v_1, \dots, v_n) \mapsto f(v_1) \cdots f(v_n)$$

n -linear, und wir erhalten eine lineare Abbildung

$$\tilde{f}_n: V^{\otimes n} \rightarrow A \quad \text{mit} \quad \tilde{f}_n(v_1 \otimes \dots \otimes v_n) = f(v_1) \cdots f(v_n).$$

Wir setzen nun die linearen Abbildungen $\tilde{f}_n$ zu einer linearen Abbildung

$$\tilde{f}: \mathcal{T}(V) \rightarrow A \quad \text{mit} \quad \tilde{f}(\mathbf{1}) = \mathbf{1}_A$$

zusammen. Aus der Konstruktion folgt sofort $\tilde{f}(v) = f(v)$ für alle $v \in V$.

Es bleibt einzusehen, dass $\tilde{f}$ ein Homomorphismus von Algebren ist. Da $\mathcal{T}(V)$ von den Elementen der Gestalt $v_1 \otimes \dots \otimes v_n$ aufgespannt wird, reicht es aus, zu zeigen, dass

$$\tilde{f}((v_1 \otimes \dots \otimes v_n) \cdot (w_1 \otimes \dots \otimes w_k)) = \tilde{f}(v_1 \otimes \dots \otimes v_n) \tilde{f}(w_1 \otimes \dots \otimes w_k)$$

für alle $v_1, \dots, v_n, w_1, \dots, w_k \in V$ gilt. Hierzu muss man sich nur an die Definition von $\tilde{f}$ erinnern, die zeigt, dass beide Seiten der gewünschten Gleichung mit

$$f(v_1) \cdots f(v_n) \cdot f(w_1) \cdots f(w_k)$$

übereinstimmen. □

Aufgaben zu Abschnitt 10.1

Aufgabe 10.1.14. Ziel dieser Aufgabe ist es, eine bessere Vorstellung des Tensorprodukts zweier Vektorräume V und W zu bekommen. Dazu sei $B_V = \{e_i : i \in I\}$ bzw. $B_W = \{f_j : j \in J\}$ eine Basis von V bzw. W . Zeigen Sie:

- (a) Jede Funktion $f: B_V \times B_W \rightarrow \mathbb{K}$ lässt sich eindeutig zu einer bilinearen Abbildung $\tilde{f}: V \times W \rightarrow \mathbb{K}$ fortsetzen.
- (b) Die Menge $B_V \otimes B_W = \{e_i \otimes f_j : i \in I, j \in J\}$ ist eine Basis des Vektorraums $V \otimes W$.
- (c) Jedes Element $x \in V \otimes W$ lässt sich in eindeutiger Weise darstellen als eine endliche Summe $x = \sum_{i \in I} e_i \otimes w_i$ mit $w_i \in W$ (Man fasse die Darstellung eines Elements, die man aus (b) erhält, geeignet zusammen). Man folgere hieraus

$$V \otimes W \cong \bigoplus_{i \in I} e_i \otimes W \cong W^{(I)}.$$

- (d) Sind V_1 und V_2 Vektorräume, so gilt $(V_1 \oplus V_2) \otimes W \cong (V_1 \otimes W) \oplus (V_2 \otimes W)$ (Verwende (c)).
- (e) $V \otimes W$ und $W \otimes V$ sind isomorphe Vektorräume.

10.2 Symmetrische und alternierende Produkte

Definition 10.2.1. Sei V ein $\mathbb{K}$ -Vektorraum und $n \geq 2$. Wir definieren

$$S^n(V) := V^{\otimes n} / U,$$

wobei U der von den Elementen der Gestalt

$$v_1 \otimes \dots \otimes v_n - v_{\sigma(1)} \otimes \dots \otimes v_{\sigma(n)}, \quad \sigma \in S_n,$$

aufgespannte Untervektorraum ist. Der Raum $S^n(V)$ heißt *n-te symmetrische Potenz von V*. Wir setzen

$$v_1 \vee \dots \vee v_n := v_1 \otimes \dots \otimes v_n + U$$

und beachten, dass dieses Produkt in dem Sinne symmetrisch ist, dass

$$v_1 \vee \dots \vee v_n = v_{\sigma(1)} \vee \dots \vee v_{\sigma(n)}$$

für alle $\sigma \in S_n$ und $(v_1, \dots, v_n) \in V^n$ gilt. Für $n = 2$ bedeutet dies

$$v_1 \vee v_2 = v_2 \vee v_1$$

für $v_1, v_2 \in V$.

Für $n \leq 1$ setzen wir $S^0(V) = \mathbb{K}$ und $S^1(V) = V$.

Lemma 10.2.2. (Universelle Eigenschaft von $S^n(V)$) Sind V und X Vektorräume und $f: V^n \rightarrow X$ eine symmetrische n -lineare Abbildung, d.h. für jede Permutation $\sigma \in S_n$ gilt

$$f(v_1, \dots, v_n) = f(v_{\sigma(1)}, \dots, v_{\sigma(n)})$$

für alle $v \in V^n$, so existiert genau eine lineare Abbildung $\tilde{f}: S^n(V) \rightarrow X$ mit

$$\tilde{f}(v_1 \vee \dots \vee v_n) = f(v_1, \dots, v_n).$$

Beweis. Zunächst erhalten wir mit der universellen Eigenschaft des Tensorprodukts eine eindeutig bestimmte lineare Abbildung $f_0: V^{\otimes n} \rightarrow X$ mit

$$f_0(v_1 \otimes \cdots \otimes v_n) = f(v_1, \dots, v_n) \quad \text{für alle } v_1, \dots, v_n \in V.$$

Wegen der Symmetrieeigenschaft von f verschwindet f_0 auf dem Unterraum U , faktorisiert also zu einer linearen Abbildung $\tilde{f}: S^n(V) \rightarrow X$ mit der gewünschten Eigenschaft. $\square$

Definition 10.2.3. Sei V ein $\mathbb{K}$ -Vektorraum, $\text{char } \mathbb{K} \neq 2$ und $n \geq 2$. Sei $\text{sgn}: S_n \rightarrow \{1, -1\}$ der Signaturhomomorphismus. Wir definieren

$$\Lambda^n(V) := V^{\otimes n} / U,$$

wobei U von den Elementen der Gestalt

$$v_1 \otimes \cdots \otimes v_n - \text{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}, \quad \sigma \in S_n$$

aufgespannt wird. Der Raum $\Lambda^n(V)$ heißt *n-te äußere Potenz von V* . Wir setzen

$$v_1 \wedge \cdots \wedge v_n := v_1 \otimes \cdots \otimes v_n + U$$

und beachten, dass dieses Produkt in dem Sinne alternierend ist, dass

$$v_1 \wedge \cdots \wedge v_n = \text{sgn}(\sigma) v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(n)}$$

für alle $\sigma \in S_n$ und $(v_1, \dots, v_n) \in V^n$ gilt. Für $n = 2$ bedeutet dies

$$v_1 \wedge v_2 = -v_2 \wedge v_1$$

für $v_1, v_2 \in V$. Weiter definiert man $\Lambda^0(V) := \mathbb{K}$ und $\Lambda^1(V) := V$.

Lemma 10.2.4. (Universelle Eigenschaft von $\Lambda^n(V)$) Sind V und X Vektorräume und $f: V^n \rightarrow X$ eine alternierende n -lineare Abbildung, d.h. für jede Permutation $\sigma \in S_n$ gilt

$$f(v_1, \dots, v_n) = \text{sgn}(\sigma) f(v_{\sigma(1)}, \dots, v_{\sigma(n)})$$

für alle $v \in V^n$, so existiert genau eine lineare Abbildung $\tilde{f}: \Lambda^n(V) \rightarrow X$ mit

$$\tilde{f}(v_1 \wedge \cdots \wedge v_n) = f(v_1, \dots, v_n).$$

Beweis. Der Beweis ist vollkommen analog zu dem Beweis von Lemma 10.2.2. $\square$

Kapitel 11

Affine und konvexe Geometrie

11.1 Affine Geometrie

Wir haben bereits affine Unterräume von Vektorräumen kennengelernt. Das sind Teilmenge A eines $\mathbb{K}$ -Vektorraums V , von der Gestalt $A = x + U$ wobei $x \in V$ und $U \subseteq V$ ein linearer Unterraum ist. Natürlich sind lineare Unterräume Beispiele affiner Unterräume, aber wenn A nicht die 0 enthält, d.h. $x \notin U$, dann ist A natürlich kein Untervektorraum.

In diesem Abschnitt werden wir das Konzept eines affinen Raums kennenlernen und sehen, dass affine Unterräume eines Vektorraums insbesondere affine Räume sind. Der Begriff des affinen Raums ist insbesondere dadurch motiviert, dass man für geometrische Überlegungen klar zwischen Punkten und Vektoren unterscheiden möchte. Im Rahmen der Theorie der Vektorräume tut man das nicht, da man in diesem Kontext Vektoren mit den Elementen (Punkten) eines Vektorraums identifiziert. In der Geometrie ist das aber kein natürlicher Standpunkt. Insbesondere sind Translationsabbildungen $\tau_v(x) := x + v$ in geometrischen Kontexten typischerweise als erlaubte Koordinatentransformationen zugelassen, aber diese Abbildungen sind keine Vektorraumisomorphismen.

Um die Begriffsbildungen zu motivieren, betrachten wir zunächst einen Vektorraum V . Gilt $b = a + x$ für $a, b, x \in V$, so denken wir uns x repräsentiert durch das Paar der Punkte (a, b) und schreiben daher auch

$$\mathbf{x} := \overrightarrow{ab} \quad \text{und} \quad b = a + \overrightarrow{ab}.$$

Aus den Rechenregeln für die Addition in Vektorräumen erhalten wir damit sofort:

- (1) $\overrightarrow{aa} = 0$ für alle $a \in V$.
- (2) $\overrightarrow{ab} + \overrightarrow{bc} = \overrightarrow{ac}$ für $a, b, c \in V$.
- (3) $\overrightarrow{ba} = -\overrightarrow{ab}$ für $a, b \in V$.

Definition 11.1.1. Eine *affiner Raum* (über dem Körper $\mathbb{K}$) ist ein Tripel $(A, V, +)$ bestehend aus einer Menge A , einem $\mathbb{K}$ -Vektorraum V und einer Abbildung

$$+: A \times V \rightarrow A, \quad (a, \mathbf{x}) \mapsto a + \mathbf{x},$$

so dass folgende Bedingungen erfüllt sind:

- (A1) $a + \mathbf{0} = a$ für alle $a \in A$.
- (A2) $a + (\mathbf{x} + \mathbf{y}) = (a + \mathbf{x}) + \mathbf{y}$ für $a \in A, \mathbf{x}, \mathbf{y} \in V$.
- (A3) Zu $a, b \in A$ existiert genau ein $\mathbf{x} \in V$ mit $b = a + \mathbf{x}$. Dieses Element $\overrightarrow{ab} := \mathbf{x}$ nennen wir dann *Verschiebungsvektor von a nach b* .

Wir nennen V den *Verschiebungsraum* zu A . Man beachte, dass A auch leer sein darf. Ist das nicht der Fall, so nennen wir

$$\dim A := \dim V$$

die *Dimension* von A .

Eine *affine Gerade* ist ein affiner Raum der Dimension 1. Eine *affine Ebene* ist ein affiner Raum der Dimension 2.

Beispiel 11.1.2. (a) Ist V ein $\mathbb{K}$ -Vektorraum, $U \subseteq V$ ein linearer Teilraum und $a \in V$, so ist $A := a + U$ bzgl. der Additionsabbildung $A \times U \rightarrow A, (a, u) \mapsto a + u$ ein affiner Raum mit dem Verschiebungsraum U .

(b) $A = \emptyset$ ist ein affiner Raum für jeden Vektorraum V bzgl. der leeren Abbildung $+: A \times V \rightarrow A$.

(c) Ist $(A, V, +)$ ein affiner Raum der Dimension 0, so ist $V = \{0\}$ und aus (A1/3) folgt sofort, dass $A = \{a\}$ nur aus einem Element besteht. Punkte sind also affine Räume der Dimension 0.

(d) Ist V ein $\mathbb{K}$ -Vektorraum, so ist $(V, V, +)$ ein affiner Raum mit Verschiebungsraum V .

Für $V = \mathbb{K}^n$ erhalten wir so den *n -dimensional affinen (Koordinaten-)Raum*

$$\mathbb{A}^n := (\mathbb{K}^n, \mathbb{K}^n, +).$$

Bemerkung 11.1.3. Ist $(A, V, +)$ eine affine Gerade, so ist $V = \mathbb{K}\mathbf{v}$ eindimensional. Fixieren wir einen Punkt $a_0 \in A$, so erhalten wir mit (A3) für jeden anderen Punkt $a \in A$ genau ein $\lambda \in \mathbb{K}$ mit $a = a_0 + \lambda\mathbf{v}$. Das ist eine Parametrisierung von A .

Wählen wir stattdessen einen anderen Punkt a'_0 und einen anderen Basisvektor $\mathbf{v}' = s\mathbf{v} \in V$, so existiert ein $\lambda_0 \in \mathbb{K}$ mit $a'_0 = a_0 + \lambda_0\mathbf{v}$. Für a erhalten wir damit

$$a = a_0 + \lambda\mathbf{v} = a'_0 + \lambda'\mathbf{v}' \quad \text{mit} \quad \lambda = \lambda_0 + \lambda's.$$

Durch Wechsel des Koordinatenursprungs a_0 zu a'_0 und des Basisvektors $\mathbf{v}$ zu $\mathbf{v}'$ ändern sich also die Koordinaten auf A gemäß:

$$\lambda = \lambda_0 + \lambda's \quad \text{und} \quad \lambda' = s^{-1}\lambda - s^{-1}\lambda_0.$$

Das sind jeweils affine Abbildungen von $\mathbb{K}$ auf sich.

Definition 11.1.4. (Affine Koordinaten) Sei $(A, V, +)$ ein affiner Raum der Dimension n und $a_0 \in A$ sowie $B = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ eine Basis von V . Dann existieren für jeden Punkt $a \in V$ eindeutig bestimmte Zahlen $\lambda_1, \dots, \lambda_n$ mit

$$a = a_0 + \lambda_1\mathbf{v}_1 + \dots + \lambda_n\mathbf{v}_n.$$

Die Zahlen $\lambda_1, \dots, \lambda_n$ heißen *affine Koordinaten* von a bzgl. (a, B) . Der Punkt a_0 heißt *Koordinatenursprung* und das $(n+1)$ -Tupel $(a_0, \mathbf{v}_1, \dots, \mathbf{v}_n)$ ein *Koordinatenrahmen*.

Ist

$$a'_0 = a_0 + \sum_{j=1}^n \lambda_j^0 \mathbf{v}_j$$

ein neuer Koordinatenursprung und

$$\mathbf{v}'_j := \sum_{i=1}^n s_{ij} \mathbf{v}_i$$

eine neue Basis in V , so erhalten wir

$$a_0 + \sum_{j=1}^n \lambda_j \mathbf{v}_j = a = a'_0 + \sum_{j=1}^n \lambda'_j \mathbf{v}'_j = a_0 + \sum_{j=1}^n \lambda_j^0 \mathbf{v}_j + \sum_{i,j=1}^n s_{ij} \lambda'_j \mathbf{v}_i = a_0 + \sum_{j=1}^n \lambda_j^0 \mathbf{v}_j + \sum_{i,j=1}^n s_{ji} \lambda'_i \mathbf{v}_j$$

und damit

$$\lambda = \lambda^0 + S\lambda' \quad \text{für} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathbb{K}^n \quad \text{und} \quad S = (s_{ij}) \in \text{GL}_n(\mathbb{K}).$$

Durch Wechsel des Koordinatenrahmens (a, B) zu (a', B') ändern sich also die Koordinaten gemäß

$$\lambda = \lambda_0 + S\lambda' \quad \text{und} \quad \lambda' = S^{-1}\lambda - S^{-1}\lambda_0.$$

Das sind jeweils affine Abbildungen von $\mathbb{K}^n$ auf sich.

Definition 11.1.5. Sei $(A, V, +)$ ein affiner Raum:

(a) Eine Teilmenge $B \subseteq A$ heißt *affiner Unterraum*, wenn ein $a \in A$ und ein linearer Unterraum $U \subseteq V$ so existiert, dass $B = a + U$ ist. Es ist klar, dass $(B, U, +)$ bzgl. der eingeschränkten Abbildung $+: B \times U \rightarrow B$ dann selbst wieder ein affiner Raum mit dem Verschiebungsraum U ist. Wir definieren die *Kodimension von B* durch

$$\text{codim}(B) = \text{codim}(U) := \dim(V/U).$$

Diese Definition ist konsistent mit unserer bisherigen Terminologie, wenn wir sie auf den affinen Raum $(V, V, +)$ anwenden.

(b) Zwei affine Unterräume $B_1, B_2 \subseteq A$ heißen *parallel*, wenn sie den gleichen Verschiebungsraum $U \subseteq V$ besitzen, also wenn $a_1, a_2 \in A$ so existieren, dass

$$B_1 = a_1 + U \quad \text{und} \quad B_2 = a_2 + U$$

gilt.

Beispiel 11.1.6. (a) Punkte $B = \{a\}$ sind affine Unterräume der Dimension 0.

(b) Sind $a \neq b \in A$, so heißt $a + \overrightarrow{\mathbb{K}ab}$ die (*affine*) *Gerade durch a und b* . Das ist ein affiner Unterraum der Dimension 1.

Für den Fall $(A, V, +) = (V, V, +)$ können wir die Gerade durch $a, b \in V$ auch symmetrischer schreiben als die Menge der Punkte

$$\{a + \lambda(b - a) : \lambda \in \mathbb{K}\} = \{(1 - \lambda)a + \lambda b : \lambda \in \mathbb{K}\} = \{\mu a + \lambda b : \lambda, \mu \in \mathbb{K}, \lambda + \mu = 1\}.$$

Diese Überlegung kann man auf den allgemeinen Fall übertragen. Dazu wählen wir einen Koordinatenursprung $a_0 \in a + \overrightarrow{\mathbb{K}ab}$. Ein allgemeiner Punkt $c \in a + \overrightarrow{\mathbb{K}ab}$ hat dann die Gestalt

$$c = a_0 + \overrightarrow{a_0c} = a_0 + \overrightarrow{a_0a} + \overrightarrow{ac} = a_0 + \overrightarrow{a_0a} + \lambda \overrightarrow{ab} = a_0 + \overrightarrow{a_0a} + \lambda(\overrightarrow{a_0b} - \overrightarrow{a_0a}) = a_0 + (1 - \lambda)\overrightarrow{a_0a} + \lambda \overrightarrow{a_0b}.$$

Für $\lambda = 0$ ergibt sich der Punkt a und für $\lambda = 1$ der Punkt b .

Definition 11.1.7. Sei V ein $\mathbb{K}$ -Vektorraum.

(a) Sind $y_1, \dots, y_n$ und $t_1, \dots, t_n \in \mathbb{K}$ mit $\sum_{j=1}^n t_j = 1$, so heißt

$$t_1 y_1 + \dots + t_n y_n$$

eine *Affinkombination der Vektoren $y_1, \dots, y_n$* .

(b) Ist $(A, V, +)$ ein affiner Raum und $a_1, \dots, a_n \in A$ sowie $t_1, \dots, t_n \in \mathbb{K}$ mit $\sum_{j=1}^n t_j = 1$, so definieren wir die zugehörige Affinkombination durch

$$t_1 a_1 + \dots + t_n a_n := a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j},$$

wobei $a_0 \in A$ beliebig ist. Damit diese Definition sinnvoll ist, müssen wir zeigen, dass die rechte Seite nicht von der Wahl des Punktes a_0 abhängt. Für einen anderen Punkt $a'_0 \in A$ erhalten wir in der Tat

$$\begin{aligned} a'_0 + \sum_{j=1}^n t_j \overrightarrow{a'_0 a_j} &= a_0 + \overrightarrow{a_0 a'_0} + \sum_{j=1}^n t_j \overrightarrow{a'_0 a_j} = a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a'_0} + \sum_{j=1}^n t_j \overrightarrow{a'_0 a_j} \\ &= a_0 + \sum_{j=1}^n t_j (\overrightarrow{a_0 a'_0} + \overrightarrow{a'_0 a_j}) = a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j}. \end{aligned}$$

Satz 11.1.8. Sei $(A, V, +)$ ein affiner Raum und $\emptyset \neq B \subseteq A$. Dann sind äquivalent:

- (i) B ist ein affiner Unterraum.
- (ii) Zu $b_1, \dots, b_n \in B$ gehört auch jede Affinkombination dieser Elemente zu B .
- (iii) (falls $2 \in \mathbb{K}^\times$) Zu $b_1 \neq b_2 \in B$ enthält B auch die Verbindungsgerade.

Beweis. (i) $\Rightarrow$ (ii): Ist $B = a + U$ für einen linearen Unterraum $U \subseteq V$ sowie $b_j = a + \mathbf{u}_j \in B$, $j = 1, \dots, n$ und $t_1, \dots, t_n \in \mathbb{K}$ mit $\sum_j t_j = 1$, so ist

$$\sum_j t_j b_j = a + \sum_j t_j \mathbf{u}_j \in a + U.$$

(ii) $\Rightarrow$ (i): Sei $b_0 \in B$ und $U := \{\mathbf{v} \in V : b_0 + \mathbf{v} \in B\}$. Dann ist $0 \in U$ und die Menge U enthält mit $u_1, \dots, u_n$ auch jede Affinkombination dieser Elemente. Sind nun $u_1, \dots, u_n \in U$ und $t_1, \dots, t_n \in \mathbb{K}$ beliebig, so ist

$$\sum_{j=1}^n t_j u_j = (1 - \sum_{j=1}^n t_j) \cdot 0 + \sum_{j=1}^n t_j u_j \in U,$$

da

$$b_0 + \sum_{j=1}^n t_j u_j = (1 - \sum_{j=1}^n t_j) \cdot b_0 + \sum_{j=1}^n t_j \overrightarrow{b_0 u_j} \in B.$$

Also ist $U \subseteq V$ ein linearer Unterraum und daher $B = b_0 + U$ ein affiner Unterraum von A .

(ii) $\Rightarrow$ (iii) ist trivial, die die Elemente der Verbindungsgerade von b_1 und b_2 genau die Affinkombinationen von b_1 und b_2 sind.

(iii) $\Rightarrow$ (i): Sei $2 \in \mathbb{K}^\times$, $b_0 \in B$ und $U := \{\mathbf{v} \in V : b_0 + \mathbf{v} \in B\}$. Dann ist $0 \in U$. Ist $\lambda \in \mathbb{K}$ und $\mathbf{v} \in U$, so ist

$$b_0 + \lambda \mathbf{v} = (1 - \lambda)b_0 + \lambda(b_0 + \mathbf{v}) \in B,$$

also $\lambda \mathbf{v} \in U$. Für $\mathbf{v}_1, \mathbf{v}_2 \in U$ ergibt sich daher

$$b_0 + \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2) = \frac{1}{2}(b_0 + \mathbf{v}_1) + \frac{1}{2}(b_0 + \mathbf{v}_2) \in B,$$

also $\frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2) \in U$ und damit auch $\mathbf{v}_1 + \mathbf{v}_2 \in U$. Wir haben damit gezeigt, dass U ein linearer Unterraum von V ist. Aus der Konstruktion folgt $B = b_0 + U$. $\square$

Bemerkung 11.1.9. Ist $\mathbb{K} = \{0, 1\}$ der zweielementige Körper und $(A, V, +)$ ein affiner Raum über $\mathbb{K}$, so besteht die Verbindungsgerade zweier Elemente $a_1, a_2 \in A$ lediglich aus der Menge $\{a_1, a_2\}$, so dass jede Teilmenge $B \subseteq A$ der Bedingung genügt, dass sie mit zwei Punkten b_1, b_2 auch deren Verbindungsgerade enthält. Andererseits ist für $A = \mathbb{K}^2$ nicht jede Teilmenge ein affiner Unterraum. Ein konkretes Beispiel ist die Teilmenge

$$B := \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{K}^2.$$

Definition 11.1.10. Man sieht leicht ein, dass für jede Familie $(B_j)_{j \in J}$ affiner Unterräume eines affinen Raums $(A, V, +)$ auch der Durchschnitt $B := \bigcap_{j \in J} B_j$ wieder ein affiner Unterraum ist. Ist der Durchschnitt nicht leer und $a \in B$ so ist $B_j = a + U_j$ mit einem linearen Unterraum U_j , also $B = a + \bigcap_{j \in J} U_j$.

Für eine Teilmenge $M \subseteq A$ ist daher

$$\text{span}_a(M) := \bigcap \{B : M \subseteq B, B \text{ affiner Unterraum}\}$$

der kleinste affine Unterraum von A , der M enthält. Man nennt ihn den *affinen Spann von M* oder *den von M erzeugten affinen Unterraum*.

Bemerkung 11.1.11. Ist $M \subseteq A$ nicht leer, so fixieren wir ein Element $m_0 \in M$ und schreiben

$$M = m_0 + \overrightarrow{m_0 M}, \quad \overrightarrow{m_0 M} := \{\overrightarrow{m_0 m} : m \in M\}$$

und, wie wir im Beweis der Implikation (ii) $\Rightarrow$ (i) von Satz 11.1.8 gesehen haben, gilt daher

$$\text{span}_a(M) = m_0 + \overrightarrow{\text{span } m_0 M}.$$

Damit lässt sich das Erzeugen affiner Räume auf das Erzeugen linearer Unterräume zurückführen. Insbesondere besteht für eine Teilmenge M des affinen Raums die Menge $\text{span}_a(M)$ aus allen Affinkombinationen der Elemente von M .

Definition 11.1.12. Analog zur Begriffsbildung bei Vektorräumen nennt man eine Teilmenge M eines affinen Raums A *affin unabhängig*, wenn für jedes $m \in M$ gilt

$$m \notin \text{span}_a(M \setminus \{m\}).$$

Lemma 11.1.13. Eine Teilmenge $B \subseteq V \setminus \{0\}$ ist genau dann linear unabhängig, wenn die Teilmenge $B_0 := B \cup \{0\} \subseteq V$ affin unabhängig ist.

Beweis. Es ist klar, dass für $b \in B$ die Bedingung $b \in \text{span}_a(B_0 \setminus \{b\})$ äquivalent ist zu $b \in \text{span}(B \setminus \{b\})$. Also folgt die lineare Unabhängigkeit von B aus der affinen Unabhängigkeit von B_0 . Ist andererseits B linear unabhängig, so ist auch $0 \notin \text{span}_a B$, denn für $b_1, \dots, b_n \in B$ folgt aus $\sum_{i=1}^n t_i = 1$ die Relation $\sum_i t_i b_i \neq 0$.

Also folgt die affine Unabhängigkeit von B_0 auch aus der linearen Unabhängigkeit von B . $\square$

Bemerkung 11.1.14. Sei $m_0 \in M \subseteq A$. Aus Lemma 11.1.13 folgt, dass die Teilmenge

$$\overrightarrow{m_0 M} \setminus \{0\}$$

genau dann linear unabhängig ist, wenn M affin unabhängig ist.

Insbesondere sind paarweise verschiedene Elemente $a_0, \dots, a_n \in A$ genau dann affin unabhängig, wenn die Vektoren

$$\overrightarrow{a_0 a_1}, \dots, \overrightarrow{a_0 a_n} \in V$$

linear unabhängig sind.

Definition 11.1.15. Ist A ein affiner Raum der Dimension n und sind $a_0, \dots, a_n \in A$ affin unabhängig, dann besitzt jedes Element eine eindeutige Darstellung

$$\begin{aligned} a &= a_0 + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j} = a_0 + (1 - \sum_{j=1}^n t_j) \overrightarrow{a_0 a_0} + \sum_{j=1}^n t_j \overrightarrow{a_0 a_j} \\ &= (1 - \sum_{j=1}^n t_j) a_0 + \sum_{j=1}^n t_j a_j = \sum_{j=0}^n t_j a_j, \end{aligned}$$

wobei $t_0 = 1 - \sum_{j=1}^n t_j$ ist, also $\sum_{j=0}^n t_j = 1$. Man nennt $(t_0, t_1, \dots, t_n)$ die *baryzentrischen Koordinaten von a bzgl. der Punkte $a_0, \dots, a_n$* .

Definition 11.1.16. (a) Seien $(A, V, +)$ und $(B, W, +)$ zwei affine Räume. Eine Abbildung $\varphi: A \rightarrow B$ heißt *affin*, wenn eine lineare Abbildung $\varphi_L: V \rightarrow W$ existiert, so dass

$$\varphi(a + v) = \varphi(a) + \varphi_L(v) \quad \text{für } a \in A, v \in V$$

gilt. Es ist klar, dass $\varphi_L: V \rightarrow W$ dann eindeutig durch φ bestimmt ist.

Wir schreiben $\text{Aff}(A, B)$ für die Menge der affinen Abbildungen $\varphi: A \rightarrow B$ und $\text{Aff}(A) := \text{Aff}(A, A)$.

(b) Eine affine Abbildung $\varphi: A \rightarrow B$ heißt *Affinität* oder *Isomorphismus affiner Räume*, wenn eine affine Abbildung $\psi: B \rightarrow A$ existiert, so dass

$$\varphi \circ \psi = \text{id}_B \quad \text{und} \quad \psi \circ \varphi = \text{id}_A$$

gilt. Es ist eine leichte Übung einzusehen, dass diese Bedingung gleichbedeutend damit ist, dass φ bijektiv, denn dann ist die Umkehrabbildung $\varphi^{-1}: B \rightarrow A$ automatisch affin (Übung!). Wir schreiben $\text{Iso}(A, B)$ für die Menge der Affinitäten $A \rightarrow B$ und $\text{Aut}(A) := \text{Iso}(A, A)$ für die Menge der affinen Automorphismen von A .

Beispiel 11.1.17. (a) Seien V und W Vektorräume und $(V, V, +)$ bzw. $(W, W, +)$ die zugehörigen affinen Räume.

Ist $\psi: V \rightarrow W$ eine lineare Abbildung und $w \in W$, so wird durch

$$\varphi(v) := \psi(v) + w$$

eine affine Abbildung mit $\varphi_L = \psi$ definiert, denn für jedes $a \in V$ und $v \in V$ ist dann

$$\varphi(a + v) = \psi(a + v) + w = \psi(a) + w + \psi(v) = \varphi(a) + \psi(v).$$

Ist umgekehrt $\varphi: V \rightarrow W$ affin, so ist

$$\varphi(v) = \varphi(0 + v) = \varphi_L(v) + \varphi(0).$$

(b) Für $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$ folgt aus (a), dass die affinen Abbildungen $\varphi: \mathbb{K}^n \rightarrow \mathbb{K}^m$ die Gestalt

$$\varphi(x) = Ax + b$$

für eine Matrix $A \in M_{m,n}(\mathbb{K})$ und $b \in \mathbb{K}^m$ haben. Für $n = m = 1$ ergibt sich speziell die Gestalt

$$\varphi(x) = ax + b.$$

Bemerkung 11.1.18. (a) Kompositionen affiner Abbildungen sind affin. Insbesondere ist die Menge $\text{Aff}(A)$ ein Monoid bzgl. Komposition und neutralem Element id_A .

(b) Die Zuordnung

$$\text{Aff}(A) \rightarrow \text{End}(V), \quad \varphi \mapsto \varphi_L$$

ist ein Morphismus von Monoiden, d.h.,

$$(\text{id}_A)_L = \text{id}_V \quad \text{und} \quad (\varphi \circ \psi)_L = \varphi_L \circ \psi_L.$$

(c) Sei $a_0 \in A$ fixiert. Zu jeder linearen Abbildung $\psi: V \rightarrow V$ können wir dann eine Abbildung

$$\varphi: A \rightarrow A, \quad a_0 + v \mapsto a_0 + \psi(v)$$

definieren. Wir zeigen, dass φ affin ist mit $\varphi_L = \psi$. In der Tat gilt für $a = a_0 + w$ und $v \in V$ die Relation

$$\varphi(a + v) = \varphi(a_0 + w + v) = a_0 + \psi(w + v) = a_0 + (\psi(w) + \psi(v)) = \varphi(a_0 + w) + \psi(v) = \varphi(a) + \psi(v).$$

Lemma 11.1.19. (a) Für jedes $v \in V$ ist die Translation

$$\tau_v: A \rightarrow A, \quad a \mapsto a + v$$

eine affine Abbildung mit $(\tau_v)_L = \text{id}_V$ und jede affine Abbildung $\varphi: A \rightarrow A$ mit $\varphi_L = \text{id}_V$ ist eine Translation.

(b) Die Translationen bilden eine Untergruppe von $\text{Aut}(A)$ und für $v, w \in V$ gilt

$$\tau_v \circ \tau_w = \tau_{v+w} \quad \text{und} \quad \tau_v^{-1} = \tau_{-v}.$$

(c) Eine Abbildung $\varphi: A \rightarrow A$ ist genau dann affin, wenn eine lineare Abbildung $\varphi_L: V \rightarrow V$ existiert, so dass

$$\varphi \circ \tau_v = \tau_{\varphi_L(v)} \circ \varphi$$

für alle $v \in V$ gilt.

Beweis. (a) Um einzusehen, dass die Translationen τ_v affin sind, beachten wir, dass für alle $w \in V$ gilt:

$$\tau_v(a + w) = (a + w) + v = (a + v) + w = \tau_v(a) + w.$$

Also ist τ_v affin mit linearem Anteil id_V .

Sei umgekehrt $\varphi: A \rightarrow A$ affin mit $\varphi_L = \text{id}_V$. Wir dürfen o.B.d.A. $A \neq \emptyset$ annehmen. Wir fixieren ein Element $a_0 \in A$ und $v \in V$ mit $\varphi(a_0) = a_0 + v$. Für $a \in A$ existiert ein $w \in V$ mit $a = a_0 + w$. Dann ist

$$\varphi(a) = \varphi(a_0 + w) = \varphi(a_0) + w = a_0 + v + w = (a_0 + w) + v = \tau_v(a).$$

Also ist $\varphi = \tau_v$.

(b) ist trivial.

(c) ist eine Umformulierung der Definition einer affinen Abbildung. □

Bemerkung 11.1.20. (Struktur der Gruppe $\text{Aut}(A)$) Die Zuordnung

$$\text{Aut}(A) \rightarrow \text{GL}(V), \quad \varphi \mapsto \varphi_L$$

ist ein Gruppenhomomorphismus. Lemma 11.1.19 besagt gerade, dass der Kern von diesem Homomorphismus aus der Gruppe $\tau(V) := \{\tau_v: v \in V\}$ der Translationen besteht.

Wir wollen nun die Struktur der Gruppe $\text{Aut}(A)$ der Affinitäten des affinen Raums A beschreiben. Dazu wählen wir einen Punkt $a_0 \in A$. Ist $\psi \in \text{GL}(V)$, so erhalten wir eine affine Abbildung

$$\psi^{a_0}: A \rightarrow A, \quad a_0 + v \mapsto a_0 + \psi(v),$$

die a_0 fest lässt und für die $(\psi^{a_0})_L = \psi$ gilt (Bemerkung 11.1.18(c)). Ist $\varphi \in \text{Aut}(A)$ mit $\varphi(a_0) = a_0$, so erhalten wir

$$\varphi(a_0 + v) = \varphi(a_0) + \varphi_L(v) = a_0 + \varphi_L(v),$$

also $\varphi = (\varphi_L)^{a_0}$. Wir erhalten also eine Bijektion

$$\sigma_{a_0}: \text{GL}(V) \rightarrow \text{Aut}(A)_{a_0} := \{\varphi \in \text{Aut}(A): \varphi(a_0) = a_0\}, \quad \psi \mapsto \psi^{a_0}.$$

Aus $(\psi \circ \psi')^{a_0} = \psi^{a_0} \circ (\psi')^{a_0}$ (Nachweis als Übung) folgt, dass σ_{a_0} ein Isomorphismus von Gruppen ist.

Ist nun $\varphi \in \text{Aut}(A)$ beliebig und $\varphi(a_0) = a_0 + v$, so ist $(\tau_{-v}\varphi)(a_0) = a_0$. Also ist $\tau_{-v} \circ \varphi = (\varphi_L)^{a_0}$ und somit

$$\varphi = \tau_v \circ (\varphi_L)^{a_0}.$$

Jede affine Abbildung ist also eine Komposition einer affinen Abbildung, die a_0 fixiert und einer Translation. Aus $v = a_0\varphi(a_0)$ folgt die Eindeutigkeit dieser Zerlegung. Für $v \in V$ und $\psi \in \text{GL}(V)$ ist nun

$$\psi^{a_0} \circ \tau_v = \tau_{\psi(v)} \circ \psi^{a_0}$$

(Lemma 11.1.19(c)). Das Komposition zweier affiner Abbildung lässt sich bzgl. dieser Zerlegung also schreiben als

$$(\tau_{v_1} \circ \psi_1^{a_0}) \circ (\tau_{v_2} \circ \psi_2^{a_0}) = \tau_{v_1 + \psi_1(v_2)} \circ (\psi_1 \psi_2)^{a_0}.$$

In der Sprache der Gruppentheorie nennt man $\text{Aut}(A)$ daher ein *semidirektes Produkt* der Untergruppen $\tau(V)$ und $\text{GL}(V)^{a_0}$.

Fixiert man statt $a_0 \in A$ einen andere Punkt a_1 als Koordinatenursprung und ist $b := \overrightarrow{a_0 a_1}$, so ist

$$\tau_b \circ \psi^{a_0} \circ \tau_{-b} = \psi^{a_1}$$

die eindeutig bestimmte affine Abbildung φ , die a_1 fixiert mit $\varphi_L = \psi$.

Lemma 11.1.21. *Affine Abbildungen erhalten Affinkombinationen: Ist $\varphi: A \rightarrow B$ affin, $a_1, \dots, a_n \in A$ und $t_1, \dots, t_n \in \mathbb{K}$ mit $\sum_j t_j = 1$, so gilt*

$$\varphi\left(\sum_{j=1}^n t_j a_j\right) = \sum_{j=1}^n t_j \varphi(a_j). \quad (11.1)$$

Gilt (11.1) für alle Affinkombinationen in A für eine Abbildung $\varphi: A \rightarrow B$, so ist φ affin. Ist $2 \in \mathbb{K}^\times$, so reicht es sogar aus, dass dies für alle Affinkombinationen von jeweils 2 Elementen gilt.

Beweis. Wir fixieren einen Punkt $a_0 \in A$. Ist φ affin, so gilt $\varphi(a_0 + v) = \varphi(a_0) + \varphi_L(v)$. Schreiben wir $a_j = a_0 + v_j$, so erhalten wir insbesondere

$$\begin{aligned} \varphi\left(\sum_{j=1}^n t_j a_j\right) &= \varphi\left(a_0 + \sum_{j=1}^n t_j v_j\right) = \varphi(a_0) + \varphi_L\left(\sum_{j=1}^n t_j v_j\right) = \varphi(a_0) + \sum_{j=1}^n t_j \varphi_L(v_j) \\ &= \sum_{j=1}^n t_j (\varphi(a_0) + \varphi_L(v_j)) = \sum_{j=1}^n t_j \varphi(a_j). \end{aligned}$$

Sei nun $\varphi: A \rightarrow B$ eine Abbildung, die Affinkombinationen zweier Elemente erhält. Wir definieren $\varphi_L: V \rightarrow W$ durch

$$\varphi(a_0 + v) := \varphi(a_0) + \varphi_L(v), \quad \text{d.h.} \quad \varphi_L(v) = \overrightarrow{\varphi(a_0) \varphi(a_0 + v)}.$$

Wir müssen zeigen, dass φ_L linear ist. Für $\lambda \in \mathbb{K}$ erhalten wir aus

$$a_0 + \lambda v = (1 - \lambda)a_0 + \lambda(a_0 + v)$$

die Relation

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

also $\varphi_L(\lambda v) = \lambda\varphi_L(v)$ für $\lambda \in \mathbb{K}$, $v \in V$.

Seien nun $v, w \in V$ und $2 \in \mathbb{K}^\times$. Dann ist

$$\begin{aligned} \varphi(a_0) + \varphi_L(v + w) &= \varphi(a_0 + v + w) = \varphi\left(\frac{1}{2}(a_0 + 2v) + \frac{1}{2}(a_0 + 2w)\right) = \frac{1}{2}\varphi(a_0 + 2v) + \frac{1}{2}\varphi(a_0 + 2w) \\ &= \varphi(a_0) + \frac{1}{2}\varphi_L(2v) + \frac{1}{2}\varphi_L(2w) = \varphi(a_0) + \varphi_L(v) + \varphi_L(w). \end{aligned}$$

Hieraus folgt die Linearität von φ_L .

Ist $2 = 0$ in $\mathbb{K}$, so müssen wir anders argumentieren. Hierzu benötigen wir Affinkombinationen dreier Elemente. Mit

$$a_0 + v + w = (-1) \cdot a_0 + 1 \cdot (a_0 + v) + 1 \cdot (a_0 + w)$$

erhalten wir

$$\begin{aligned} \varphi(a_0) + \varphi_L(v + w) &= \varphi(a_0 + v + w) = (-1) \cdot \varphi(a_0) + 1 \cdot \varphi(a_0 + v) + 1 \cdot \varphi(a_0 + w) \\ &= (-1) \cdot \varphi(a_0) + 1 \cdot (\varphi(a_0) + \varphi_L(v)) + 1 \cdot (\varphi(a_0) + \varphi_L(w)) = \varphi(a_0) + \varphi_L(v) + \varphi_L(w). \end{aligned}$$

Damit haben wir auch im allgemeinen Fall die Linearität von φ_L gezeigt. $\square$

Bemerkung 11.1.22. Wie wir gerade gesehen haben, sind affine Abbildungen mit Affinkombinationen verträglich. Wenden wir dies auf affine Abbildungen $A^n = \mathbb{K}^n \rightarrow A$ an, so erhalten wir für einen affinen Raum A :

$$\dim A := \sup\{n \in \mathbb{N}_0 : (\exists a_0, \dots, a_n \in A) a_0, \dots, a_n \text{ affin unabh.}\}.$$

11.2 Konvexe Mengen

In diesem Abschnitt sei $\mathbb{K} = \mathbb{R}$ der Körper der reellen Zahlen und alle Vektorräume sind reell.

Definition 11.2.1. Sei A ein affiner Raum. Zu zwei Punkten $a, b \in A$ heißt

$$\overline{ab} := \{ta + (1-t)b : 0 \leq t \leq 1\}$$

die *Verbindungsstrecke* von a und b .

Eine Teilmenge $K \subseteq A$ heißt *konvex*, wenn sie mit je zwei Elementen $a, b \in K$ auch deren Verbindungsstrecke enthält:

$$(\forall a, b \in K) \quad \overline{ab} \subseteq K.$$

Die *Dimension einer konvexen Menge* wird definiert durch die Dimension

$$\dim K := \dim(\text{span}_a(K))$$

des affinen Unterraums, den K erzeugt.

Beispiele 11.2.2. (a) Jeder affine Unterraum $B \subseteq A$ ist konvex.

(b) Jede Strecke $\overline{ab}$ ist konvex: Für $c = ta + (1-t)b$ und $d = sa + (1-s)b$ mit $0 \leq t, s \leq 1$ sowie $\lambda \in [0, 1]$ ist

$$\lambda c + (1-\lambda)d = (\lambda t + (1-\lambda)s)a + (\lambda(1-t) + (1-\lambda)(1-s))b \in \overline{ab},$$

da

$$(\lambda t + (1-\lambda)s) + (\lambda(1-t) + (1-\lambda)(1-s)) = \lambda + (1-\lambda) = 1$$

ist sowie beide Summanden nichtnegativ.

(c) Sei $(V, \|\cdot\|)$ ein normierter Raum und $B_r(0) := \{v \in V : \|v\| \leq r\}$ für ein $r \geq 0$ die abgeschlossene Kugel vom Radius r um 0. Dann ist $B_r(0)$ konvex, denn für $a, b \in B_r(0)$ und $0 \leq t \leq 1$ ist

$$\|ta + (1-t)b\| \leq \|ta\| + \|(1-t)b\| = t\|a\| + (1-t)\|b\| \leq tr + (1-t)r = r.$$

Definition 11.2.3. Sei $(A, V, +)$ ein reeller affiner Raum. Eine Affinkombination

$$t_1 y_1 + \dots + t_n y_n, \quad t_1 + \dots + t_n = 1$$

von Elementen $y_1, \dots, y_n \in A$ heißt *Konvexkombination*, wenn zusätzlich $t_j \geq 0$ für jedes j gilt.

Satz 11.2.4. Für eine Teilmenge K eines affinen Raums A sind äquivalent:

- (i) K ist konvex.
- (ii) Mit endlich vielen Punkten $y_1, \dots, y_n$ enthält K auch jede Konvexkombination dieser Punkte.

Beweis. (ii) $\Rightarrow$ (i) ist trivial, da für zwei Punkte $a, b \in K$ ihre Verbindungsstrecke $\overline{ab}$ aus den Konvexkombinationen von a und b besteht.

(i) $\Rightarrow$ (ii): Seien $y_1, \dots, y_n \in K$. Wir zeigen durch Induktion nach n , dass alle Konvexkombinationen dieser Elemente in K enthalten sind. Für $n = 1$ ist nichts zu tun. Wir dürfen daher

annehmen, dass alle Konvexkombinationen von $y_1, \dots, y_{n-1}$ in K enthalten sind. Sei nun $t_j \geq 0$ und $\sum_{j=1}^n t_j = 1$. Ist $t_n = 1$, so ist

$$t_1 y_1 + \dots + t_n y_n = y_n \in K.$$

Sei nun $0 < t_n < 1$. Dann schreiben wir

$$t_1 y_1 + \dots + t_n y_n = (1 - t_n) \underbrace{\sum_{j=1}^{n-1} \frac{t_j}{1 - t_n} y_j}_{\in K} + t_n y_n \in K,$$

da $\sum_{j=1}^{n-1} t_j = 1 - t_n$ ist. □

Definition 11.2.5. Ist $(K_j)_{j \in J}$ eine Familie konvexer Teilmenge des reellen affinen Raums A , so ist auch ihr Durchschnitt

$$K := \bigcap_{j \in J} K_j$$

konvex. In der Tat, folgt aus $a, b \in K$, dass $a, b \in K_j$ für jedes $j \in J$ gilt. Dann ist aber auch $\overline{ab} \subseteq K_j$ für jedes j , folglich $\overline{ab} \subseteq K$.

Ist nun $M \subseteq A$ eine Teilmenge, so schreiben wir

$$\text{conv}(M) := \bigcap \{B \subseteq A : M \subseteq B, B \text{ konvex}\}$$

für die kleinste konvexe Teilmenge von A , die M enthält. Man nennt $\text{conv}(M)$ die *konvexe Hülle* von M .

Lemma 11.2.6. Für eine Teilmenge $M \subseteq A$ besteht $\text{conv}(M)$ aus allen Konvexkombinationen von Elementen von M .

Beweis. Sei $\widetilde{M}$ die Menge der Konvexkombinationen von Elementen von M . Aus $M \subseteq \text{conv}(M)$ und der Konvexität von $\text{conv}(M)$ folgt sofort $\widetilde{M} \subseteq \text{conv}(M)$ (Satz 11.2.4).

Um die umgekehrte Inklusion zu zeigen, müssen wir einsehen, dass $\widetilde{M}$ konvex ist, denn dann folgt aus der Definition von $\text{conv}(M)$ unmittelbar $\text{conv}(M) \subseteq \widetilde{M}$. Seien dazu $a := \sum_{i=1}^m t_i m_i$ und $b := \sum_{j=1}^n s_j m'_j$ zwei Elemente von $\widetilde{M}$. Für $0 < \lambda < 1$ ist dann

$$c := \lambda a + (1 - \lambda)b = \sum_{i=1}^m \lambda t_i m_i + \sum_{j=1}^n (1 - \lambda) s_j m'_j.$$

Wegen

$$\sum_{i=1}^m \lambda t_i + \sum_{j=1}^n (1 - \lambda) s_j = \lambda + (1 - \lambda) = 1$$

ist $c \in \widetilde{M}$. Also ist $\overline{ab} \subseteq \widetilde{M}$. Die Menge $\widetilde{M}$ ist also konvex. □

Definition 11.2.7. Ist $K \subseteq A$ eine konvexe Menge und $\emptyset \neq F \subseteq K$, so nennen wir F eine *Seite* von K , wenn F konvex ist und die folgende Eigenschaft besitzt:

$$(\forall a, b \in K)(\forall 0 < \lambda < 1) \quad \lambda a + (1 - \lambda)b \in F \quad \Rightarrow \quad a, b \in F.$$

D.h., wenn F das Innere einer Strecke in K schneidet, dann auch die beiden Endpunkte der Strecke.

Ein Punkt $c \in K$, für den $\{c\}$ eine Seite ist, heißt *Extremalpunkt* von K .

Wir schreiben $\text{Ext}(K)$ für die Menge der Extremalpunkte von K und $\mathcal{F}(K)$ für die Menge der Seiten von K .

Beispiele 11.2.8. (a) Ist K ein affiner Unterraum der Dimension $d > 0$, so ist $\text{Ext}(K) = \emptyset$. Um das einzusehen, seien $x \neq y \in K$. Zwei verschiedene Punkte existieren wegen $\dim K > 0$. Dann ist

$$x = \frac{1}{2}(x+y) + \frac{1}{2}(x-y)$$

der Mittelpunkt der Strecke $\overline{(x+y)(x-y)}$, also nicht extremal.

(b) Für ein abgeschlossenes beschränktes Intervall $[a, b] \subseteq \mathbb{R}$ ist $\text{Ext}([a, b]) = \{a, b\}$, wie man sich leicht klar macht. Für ein halbbeschränktes Intervall haben wir zum Beispiel $\text{Ext}([a, \infty]) = \{a\}$.

Lemma 11.2.9. Ist $F \subseteq K$ eine Seite der konvexen Menge K und $E \subseteq F$ eine Seite von F , so ist auch E eine Seite von K .

Beweis. Seien $a, b \in K$ und $0 < \lambda < 1$ mit $\lambda a + (1-\lambda)b \in E$. Aus $E \subseteq F$ und der Seiteneigenschaft von F folgt zunächst $a, b \in F$. Da E eine Seite von F ist, erhalten wir weiter $a, b \in E$. Also ist E eine Seite von K . $\square$

Das folgende Lemma zeigt, wie Seiten sehr natürlich im Kontext von Extremalproblemen auftreten. Konvexe Mengen können eine sehr komplizierte geometrische Struktur besitzen. Im nächsten Abschnitt werden wir die sogenannten Polyeder kennenlernen. Dies ist eine Klasse konvexer Mengen, die sich bereits durch kombinatorische Daten beschreiben lässt.

Lemma 11.2.10. Ist $K \subseteq A$ konvex und $f: A \rightarrow \mathbb{R}$ affin mit $c := \inf f(K)$, so ist

$$F := \{x \in K: f(x) = c\}$$

eine Seite von K .

Beweis. Seien $a, b \in K$ und $0 < \lambda < 1$ mit $\lambda a + (1-\lambda)b \in F$. Dann ist

$$c = f(\lambda a + (1-\lambda)b) = \lambda f(a) + (1-\lambda)f(b).$$

Ist $f(a) > c$ oder $f(b) > c$, so folgt aus $f(a), f(b) \geq c$, dass auch $\lambda f(a) + (1-\lambda)f(b) > c$ ist. Also ist $f(a) = f(b) = c$ und somit $a, b \in F$. $\square$

Beispiel 11.2.11. Für die abgeschlossene Einheitskugel $B := B_1(0)$ in einem normierten Raum $(V, \|\cdot\|)$ ist

$$\text{Ext}(B_1(0)) \subseteq S := \{v \in V: \|v\| = 1\},$$

aber im allgemeinen haben wir keine Gleichheit, wie man für $V = \mathbb{R}^2$ und $\|x\| := \max\{|x_1|, |x_2|\}$ leicht einsieht. In diesem Fall ist B ein Quadrat mit den Ecken $\{\pm e_1 \pm e_2\}$, so dass nur 4 Extremalpunkte vorliegen.

Ist V ein euklidischer Raum und $\|v\| = \sqrt{\langle v, v \rangle}$, so ist

$$\text{Ext}(B) = S,$$

denn für $x \in S$ erhalten wir aus der Diskussion der Gleichheit in der Cauchy-Schwarzschen Ungleichung (Bemerkung 7.1.10), dass

$$\{y \in B: \langle x, y \rangle = 1\} = \{x\}$$

gilt. Da diese Menge eine Seite ist (Lemma 11.2.10), ist x ein Extremalpunkt.

Bemerkung 11.2.12. (a) Ist $F \subseteq K$ eine Seite, so ist ihr Komplement $F^c := K \setminus F$ eine konvexe Menge: Für $a, b \in F^c$ und $0 < \lambda < 1$ kann nicht $\lambda a + (1-\lambda)b \in F$ gelten. Also ist $\overline{ab} \subseteq F^c$.

(b) Sind $a_0, \dots, a_n \in K$ und gilt

$$k := \sum_{j=0}^n \lambda_j a_j \in F$$

für eine Seite F und eine Konvexkombination der a_j , so folgt aus $\lambda_j > 0$ auch $a_j \in F$. Für den Nachweis nehmen wir $\lambda_0 > 0$ an. Ist $\lambda_0 = 1$, so ist $a_0 = \sum_{j=0}^n \lambda_j a_j \in F$. Ist $0 < \lambda_0 < 1$, so ist

$$k = \lambda_0 a_0 + (1 - \lambda_0)b \quad \text{für} \quad b := \sum_{j=1}^n \frac{\lambda_j}{1 - \lambda_0} a_j.$$

Da F eine Seite ist, folgt hieraus $a_0, b \in F$.

(c) Sind $(F_j)_{j \in J}$ Seiten von K , so ist auch der Durchschnitt $F := \bigcap_{j \in J} F_j$ eine Seite von K . In der Tat, erhalten wir aus $a, b \in K$, $0 < \lambda < 1$ und $\lambda a + (1 - \lambda)b \in F$ auch $\lambda a + (1 - \lambda)b \in F_j$ für jedes j und somit $a, b \in F_j$. Da $j \in J$ beliebig war, gilt $a, b \in F$.

Aufgaben zu Abschnitt 11.2

Aufgabe 11.2.1. Zeigen Sie: Sind A, B reelle affine Räume, $M \subseteq A$ eine Teilmenge sowie $\varphi: A \rightarrow B$ eine affine Abbildung, so gilt

$$\varphi(\text{conv}(M)) = \text{conv}(\varphi(M)).$$

11.3 Polyeder

Bisher haben wir in dieser Vorlesung immer Gleichungen $F(x) = y$ betrachtet, wobei $F: V \rightarrow W$ linear oder quadratisch war. Entsprechend hatten die Lösungsmengen dieser Gleichungen sehr verschiedenartige geometrische Strukturen. Ist F linear, so sind sie affine Unterräume, und ist F quadratisch und $W = \mathbb{K}$, so handelt es sich um Quadriken. In diesem Abschnitt wenden wir uns nun linearen Ungleichungen zu. Solche Ungleichungen treten in sehr natürlicher Weise als Nebenbedingungen bei Optimierungsproblemen auf.

In diesem Abschnitt sind alle Vektorräume reell, denn wir benötigen die Ordnung des Körpers $\mathbb{R}$ der reellen Zahlen, um Ungleichungen modellieren zu können.

Definition 11.3.1. Sei A ein reeller affiner Raum.

(a) Ist $\alpha: A \rightarrow \mathbb{R}$ eine nicht konstante affine Abbildung, so heißt

$$H := \{x \in A: \alpha(x) \geq 0\} = \alpha^{-1}([0, \infty])$$

ein (abgeschlossener) Halbraum in A . Die affine Hyperebene

$$\partial H := \{x \in V: \alpha(x) = 0\}$$

heißt Rand von H .

Für $A = (V, V, +)$ haben affine Abbildungen $\alpha: A \rightarrow \mathbb{R}$ die Gestalt $\alpha(x) = \alpha_L(x) - c$ für ein $c \in \mathbb{R}$, und α ist genau dann nicht konstant, wenn $\alpha_L \neq 0$ ist. Abgeschlossene Halbräume in V haben daher die Gestalt

$$H := \{x \in V: \alpha_L(x) \geq c\} = \alpha_L^{-1}([c, \infty]), \quad c \in \mathbb{R}, 0 \neq \alpha_L \in V^*.$$

(b) Ein Polyeder $P \subseteq A$ ist ein Durchschnitt von endlich vielen (abgeschlossenen) Halbräumen

$$P = H_1 \cap \dots \cap H_k.^1$$

(c) Ein Extrempunkt $e \in P$ eines Polyeders heißt Ecke. Eine eindimensionale Seite $K \subseteq P$ heißt Kante.

Bemerkung 11.3.2. Für $A = (V, V, +)$ werden Polyeder also beschrieben durch $\alpha_1, \dots, \alpha_k \in V^* \setminus \{0\}$ und $c_1, \dots, c_k \in \mathbb{R}$ mit

$$P = \{v \in V: (\forall j = 1, \dots, k) \alpha_j(v) \geq c_j\}.$$

D.h. ein Polyeder ist die Lösungsmenge endlich vieler linearer Ungleichungen in V .

¹Beachte: Polyeder sind sächlich: **das** Polyeder

In diesem Sinn verallgemeinern Polyeder die Lösungsmengen von Systemen linearer Gleichungen, von denen wir ja schon wissen, dass sie affine Räume sind. Unser Ziel in diesem Abschnitt ist es, die Struktur von Polyedern besser zu verstehen. Hierbei wird der Begriff der Seite von zentraler Bedeutung sein.

Beispiel 11.3.3. (a) Die Halbräume von $\mathbb{R}$ sind genau die abgeschlossenen halbbeschränkten Intervalle

$$[c, \infty[\quad \text{und} \quad] - \infty, c].$$

Der Punkt c ist jeweils eine Ecke dieses Polyeders und bildet die einzige echte Seite.

(b) In der Ebene $V = \mathbb{R}^2$ ist die rechte Halbebene

$$\{x \in \mathbb{R}^2 : x_1 \geq 0\}$$

ein Halbraum.

(c) In $V = \mathbb{R}^n$ ist der *positive Orthant*

$$P = (\mathbb{R}_+)^n := \{x \in \mathbb{R}^n : x_1 \geq 0, \dots, x_n \geq 0\}$$

ein Polyeder.

(d) In $V = \mathbb{R}^n$ ist der *Einheitswürfel*

$$W = \{x \in \mathbb{R}^n : (\forall j) 0 \leq x_j \leq 1\}$$

ein Polyeder.

(e) In $V = \mathbb{R}^n$ ist das *Kreuzpolytop*

$$K = \left\{x \in \mathbb{R}^n : \sum_{j=1}^n |x_j| \leq 1\right\} = \left\{x \in \mathbb{R}^n : (\forall \varepsilon \in \{\pm 1\}^n) \sum_{j=1}^n \varepsilon_j x_j \leq 1\right\}$$

ein Polyeder. Für $n = 3$ erhält man so ein Oktaeder.

(f) In $V = \mathbb{R}^n$ ist die Menge

$$S := \left\{x \in \mathbb{R}^n : (\forall j) x_j \geq 0, \sum_{j=1}^n x_j \leq 1\right\}$$

ein sogenanntes *Simplex*.

Eine symmetrischere Darstellung erhält man, wenn man in dem n -dimensionalen affinen Raum $\mathbb{A}^n = (\mathbb{R}^n, \mathbb{R}^n, 0)$ für $n+1$ affin unabhängige Punkte $a_0, \dots, a_n$ deren konvexe Hülle betrachtet

$$S(a_0, \dots, a_n) := \left\{\sum_{j=0}^n x_j a_j : (\forall j) x_j \geq 0, \sum_{j=0}^n x_j = 1\right\}.$$

Durch

$$\varphi: \mathbb{R}^n \rightarrow \mathbb{A}^n, \quad \varphi(0) = a_0, \quad \varphi(e_1) = a_1, \quad \dots, \quad \varphi(e_n) = a_n$$

erhalten wir dann eine Affinität, also einen Isomorphismus affiner Räume, der S auf $S(a_0, \dots, a_n)$ abbildet. Also ist auch $S(a_0, \dots, a_n)$ ein Polyeder.

Jeder Punkt $S(a) = \{a\}$ ist ein 0-dimensionales Simplex. Für $a_0 \neq a_1$ ist die Strecke $[a_0, a_1] = S(a_0, a_1)$ ein eindimensionales Simplex. Liegt der Punkt a_2 nicht auf der Geraden $\text{span}_a(a_0, a_1) = a_0 + \mathbb{R}\overrightarrow{a_0 a_1}$, so ist $S(a_0, a_1, a_2)$ ein 2-dimensionales Simplex (ein Dreieck).

Lemma 11.3.4. Ist $\varphi: A \rightarrow B$ eine affine Abbildung reeller affiner Räume und $K \subseteq B$ konvex, so ist $\varphi^{-1}(K)$ eine konvexe Teilmenge von A .

Beweis. Für $a, b \in \varphi^{-1}(K)$ ist $\varphi(\overline{ab}) = \overline{\varphi(a)\varphi(b)} \subseteq K$, also $\overline{ab} \subseteq \varphi^{-1}(K)$. □

Satz 11.3.5. (a) Jedes Polyeder P ist konvex.

(b) Der Durchschnitt eines Polyeders mit einer affinen Hyperebene ist ein Polyeder.

Beweis. (a) Da Durchschnitte konvexer Mengen konvex sind, reicht es aus zu zeigen, dass Halbräume konvex sind. Wegen Lemma 11.3.4 folgt das sofort aus der Konvexität der Halbstrahlen $[c, \infty[$ in $\mathbb{R}$.

(b) Sei $H \subseteq A$ eine affine Hyperebene und $P \subseteq A$ ein Polyeder. Dann ist H von der Gestalt $H = \{x \in A: \alpha(x) = 0\}$ für eine nicht konstante affine Abbildung $\alpha: A \rightarrow \mathbb{R}$. Damit ist

$$P \cap H = P \cap \{x \in A: \alpha(x) \geq 0\} \cap \{x \in A: -\alpha(x) \geq 0\}$$

ebenfalls Schnitt endlich vieler Halbräume und damit ein Polyeder. $\square$

Seitensatz für Polyeder

Satz 11.3.6. In $A = (V, V, +)$ sei für $\alpha_1, \dots, \alpha_k \in V^* \setminus \{0\}$ und $c_1, \dots, c_k \in \mathbb{R}$ das Polyeder

$$P = \{v \in V: (\forall j = 1, \dots, k) \alpha_j(x) \geq c_j\}$$

nicht leer. Für jede endliche Teilmenge $E \subseteq I := \{1, \dots, k\}$ ist dann

$$F_E := \{x \in P: (\forall j \in E) \alpha_j(x) = c_j\}$$

eine Seite von P und jede Seite von P sieht so aus. Insbesondere ist die Menge $\mathcal{F}(P)$ der Seiten von P endlich.

Beweis. Aus Lemma 11.2.10 folgt, dass F_E für jede einelementige Menge $E = \{j\}$ eine Seite von P ist. Da Durchschnitte von Seiten wieder Seiten sind (Bemerkung 11.2.12(c)), ist F_E für jede Teilmenge $E \subseteq I$ eine Seite.

Sei nun $F \subseteq P$ eine Seite und

$$E := \{j: \alpha_j(F) = \{c_j\}\}.$$

Dann gilt trivialerweise $F \subseteq F_E$ und F ist eine Seite des Polyeders F_E (Lemma 11.2.9). Für $j \in E^c = I \setminus E$ existiert ein $x_j \in F$ mit $\alpha_j(x_j) > c_j$. Für die Konvexkombination

$$x := \sum_{j \notin E} \frac{1}{|E^c|} x_j \in F$$

ist dann $\alpha_j(x) > c_j$ für alle $j \in E^c$, d.h., $\{j: \alpha_j(x) > c_j\} = E^c$. Sei $y \in F_E$ beliebig. Dann existiert ein $\lambda < 0$ mit

$$y' := \lambda y + (1 - \lambda)x \in P,$$

denn wir müssen $\lambda < 0$ nur so nahe bei 0 wählen, dass

$$\alpha_j(y') = \alpha_j(\lambda y + (1 - \lambda)x) = \lambda \alpha_j(y) + (1 - \lambda) \alpha_j(x) = \alpha_j(x) + \lambda(\alpha_j(y) - \alpha_j(x)) > c_j$$

für alle $j \in E^c$ gilt. Für $j \in E$ ist $\alpha_j(x) = \alpha_j(y) = c_j$ und daher ohnehin

$$\alpha_j(\lambda y + (1 - \lambda)x) = \lambda \alpha_j(y) + (1 - \lambda) \alpha_j(x) = c_j.$$

Aus

$$x = \frac{1}{1 - \lambda}(\lambda y + (1 - \lambda)x) + \frac{-\lambda}{1 - \lambda}y \in F \quad \text{und} \quad 0 < \frac{1}{1 - \lambda} < 1$$

folgt dann $y \in F$. Also ist $F = F_E$. $\square$

Der Seitensatz hat eine ganze Reihe interessanter Konsequenzen für die Struktur eines Polyeders.

Folgerung 11.3.7. Für ein Polyeder $P \subseteq V$ wie in Satz 11.3.6 gilt:

- (i) Jede Seite $F \subseteq P$ ist ein Polyeder.
(ii) Für $E_1 \subseteq E_2$ ist $F_{E_2} \subseteq F_{E_1}$. Insbesondere haben die maximalen echten Seiten von P die Gestalt

$$F_j := \{x \in P : \alpha_j(x) = c_j\}$$

für ein $j \in \{1, \dots, m\}$, für das α_j nicht konstant auf P ist, so dass $\dim F_j < \dim P$ gilt.

- (iii) P hat genau dann keine echten Seiten, wenn P ein affiner Unterraum ist.
(iv) Die minimalen Seiten von P haben die Gestalt F_E , wobei $E \subseteq I$ maximal ist mit der Eigenschaft, dass $F_E \neq \emptyset$ gilt. Jede minimale Seite ist ein affiner Unterraum.

Beweis. (i) Ist $F = F_E$ für $E \subseteq I$, so ist

$$F = P \cap \{x \in V : (\forall j \in E) -\alpha_j(x) \geq -c_j\}.$$

(ii) folgt unmittelbar aus Satz 11.3.6.

(iii) Hat P keine echten Seiten, so folgt aus (ii), dass alle α_j auf P konstant sind. Dann ist aber

$$P = \{x \in V : (\forall j \in I) \alpha_j(x) = c_j\}$$

als Lösungsmenge eines linearen Gleichungssystems ein affiner Unterraum.

Ist andererseits P ein affiner Unterraum, so ist jedes lineare Funktional α_j auf P nach unten beschränkt, also konstant (Nachweis!). Insbesondere folgt aus (ii), dass P keine echten Seiten besitzt.

(iv) Der erste Teil der Behauptung folgt unmittelbar aus Satz 11.3.6. Ist $F \subseteq P$ eine minimale Seite, so ist F ein Polyeder ohne echte Seite, also ein affiner Unterraum. $\square$

Beispiele 11.3.8. (a) Ist $H = \{x \in V : \alpha(x) \geq c\}$ ein Halbraum, so ist $\partial H = \{x \in V : \alpha(x) = c\}$ die einzige echte Seite von H . Einen Extrempunkt besitzt H also nur dann, wenn $\dim V = 1$ ist.

(b) Für den positive Orthanten

$$P = \{x \in \mathbb{R}^n : x_1 \geq 0, \dots, x_n \geq 0\}$$

erhalten wir für jede Teilmenge $E \subseteq I := \{1, \dots, n\}$ genau eine Seite

$$F_E = \{x \in P : j \in E \Rightarrow x_j = 0\} \cong \mathbb{R}_+^{E^c}.$$

Es gibt also 2^n Seiten und $F_I = \{0\}$ ist der einzige Extrempunkt.

(c) Den Einheitswürfel

$$W = \{x \in \mathbb{R}^n : (\forall j) 0 \leq x_j \leq 1\}$$

können wir mit $I := \{1, \dots, 2n\}$ und $\alpha_j(x) = x_j$ für $j \leq n$ und $\alpha_j(x) = 1 - x_{j-n}$ für $j > n$ beschreiben als

$$W = \{x \in \mathbb{R}^n : (\forall j \in I) \alpha_j(x) \geq 0\}.$$

Für $E \subseteq I$ ist die Seite $F_E \subseteq W$ genau dann nicht leer, wenn für kein $j \in \{1, \dots, n\}$ gilt, dass $\{j, j+n\} \subseteq E$ ist. Hieraus schließt man leicht, dass der Würfel W genau 3^n Seiten besitzt. Die maximalen Seiten sind Würfel der Dimension $n-1$ und davon gibt es $2n$. Die Extrempunkte entsprechen den Teilmengen $E \subseteq I$, die für jedes $j \in \{1, \dots, n\}$ entweder j oder $j+n$ enthalten. Es gibt 2^n solche Teilmengen, und die zugehörigen Extrempunkte sind die Ecken $(\varepsilon_1, \dots, \varepsilon_n) \in W$ mit $\varepsilon_j \in \{\pm 1\}$.

(d) Das Kreuzpolytop

$$K = \left\{x \in \mathbb{R}^n : \sum_{j=1}^n |x_j| \leq 1\right\} = \left\{x \in \mathbb{R}^n : (\forall \varepsilon \in \{\pm 1\}^n) \sum_{j=1}^n \varepsilon_j x_j \leq 1\right\}$$

wird durch 2^n Ungleichungen beschrieben: Für $\varepsilon \in I := \{\pm 1\}^n$ sei $\alpha_\varepsilon(x) := 1 - \sum_{j=1}^n \varepsilon_j x_j$. Dann ist

$$K = \{x \in \mathbb{R}^n : (\forall \varepsilon \in I) \alpha_\varepsilon(x) \geq 0\}.$$

Es gibt 2^n maximale Seiten der Dimension $n - 1$:

$$F_\varepsilon = \{x \in K : \alpha_\varepsilon(x) = 0\} = \text{conv}\{\varepsilon_1 e_1, \dots, \varepsilon_n e_n\},$$

die jeweils Simplexes der Dimension $n - 1$ sind.

Es gibt $2n$ Extrempunkte $\{\pm e_1, \dots, \pm e_n\}$.

(e) Sind $a_0, \dots, a_n \in \mathbb{R}^n$ affin unabhängig und

$$S := S(a_0, \dots, a_n) = \left\{ \sum_{j=0}^n x_j a_j : (\forall j) x_j \geq 0, \quad \sum_{j=0}^n x_j = 1 \right\}$$

das von ihnen erzeugte Simplex, so erhalten wir durch die baryzentrischen Koordinatenfunktionen $\alpha_0, \dots, \alpha_n$ bzgl. $(a_0, \dots, a_n)$, d.h.,

$$a = \sum_{j=0}^n \alpha_j(x) a_j,$$

$n + 1$ affine Funktionen mit

$$S = \{x \in \mathbb{R}^n : (\forall j \in I) \alpha_j(x) \geq 0\}, \quad I = \{0, \dots, n\}.$$

Für $E \subseteq I$ ist die zugehörige Seite

$$F_E = \{x \in S : (\forall j \in E) \alpha_j(x) = 0\} = \text{conv}\{a_j : j \in E^c = I \setminus E\}$$

ein Simplex der Dimension $n - |E| = |E^c| - 1$. Die Extrempunkte sind $\{a_0, \dots, a_n\}$.

Lemma 11.3.9. Für ein Polyeder $P \subseteq V$ wie in Satz 11.3.6 und $v \in V$ sind äquivalent:

(i) $x + \mathbb{R}_+ v \subseteq P$ für ein $x \in P$.

(ii) $\alpha_j(v) \geq 0$ für alle $j \in I$.

(iii) $P + \mathbb{R}_+ v \subseteq P$.

Beweis. (i) $\Rightarrow$ (ii): Sei $x \in P$ mit $x + \mathbb{R}_+ v \subseteq P$. If $\alpha_j(v) < 0$ für ein j , so existiert ein $t > 0$ mit $\alpha_j(x + tv) < c_j$, im Widerspruch zu $x + tv \in P$.

(ii) $\Rightarrow$ (iii): Gilt (ii), so erhalten wir für $x \in P$, $t \geq 0$ und jedes j :

$$\alpha_j(x + tv) = \alpha_j(x) + t\alpha_j(v) \geq \alpha_j(x) \geq c_j.$$

Also ist $x + tv \in P$.

(iii) $\Rightarrow$ (i) ist trivial. □

Definition 11.3.10. Für ein Polyeder P wie in Satz 11.3.6 definieren wir den *Verschiebungskegel*

$$C_P := \{v \in V : (\forall j \in I) \alpha_j(v) \geq 0\}$$

und den *Verschiebungsraum*

$$T_P := C_P \cap -C_P = \{v \in V : (\forall j \in I) \alpha_j(v) = 0\} = \{\alpha_1, \dots, \alpha_n\}^\perp.$$

Aus Lemma 11.3.9 folgt sofort $P + C_P \subseteq P$ und daher auch $P + T_P = P$.

Bemerkung 11.3.11. Aus der Relation

$$T_P = \{\alpha_1, \dots, \alpha_n\}^\perp$$

folgt sofort, dass T_P genau dann $\{0\}$ ist, wenn $\alpha_1, \dots, \alpha_n$ den Dualraum V^* aufspannen (vg. Lemma 8.2.15(vi)(a)).

Satz 11.3.12. *Jede minimale Seite $F \subseteq P$ ist ein affiner Unterraum mit dem Translationsraum T_P . Insbesondere haben alle minimale Seiten die gleiche Dimension und P besitzt genau dann eine Ecke (Extremalpunkt), wenn $T_P = \{0\}$ ist.*

Beweis. Aus Satz 11.3.6 folgt zunächst, dass jede Seite von P unter Translation mit T_P invariant ist. Ist F minimal und ein affiner Unterraum mit Translationsraum U , so ist $F + U \subseteq P$, also auch $U \subseteq T_P$ (Lemma 11.3.9). Hieraus folgt $U = T_P$ für jede minimale Seite. $\square$

Satz 11.3.13. *Sei $f: P \rightarrow \mathbb{R}$ eine affine Funktion, die auf P ein Minimum annimmt und $T_P = \{0\}$. Dann existiert ein Extremalpunkt $e \in P$ mit $f(e) = \min f(P)$.*

Beweis. Ist $c := \min f(P)$, so ist $F := \{x \in P: f(x) = c\}$ eine Seite von P . Da P nur endlich viele Seiten besitzt, existiert eine minimale Seite $F_0 \subseteq F$. Diese ist wegen $T_P = \{0\}$ ein Extremalpunkt. $\square$