Inhaltsverzeichnis

Vorwort

Dies ist ein partielles Skriptum zur Vorlesung “Lineare Algebra II”. Es beruht auf Vorlesungen zur Linearen Allgebra, 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 werden sich sehr eng an dieses Buch halten. Als Begleitlektüre zu den anderen ist dieses Skript gedacht.

Manche Teile dieses Skripts wurde schon im Rahmen der Linearen Algebra I behandelt. Ich habe sie trotzdem nicht herausgenommen, um leichter auf die verweisen zu können. Zum Beispiel beginnt der Stoff des zweiten Teils der Linearen Algebra etwa in Abschnitt 5.5.

Kapitel 5

Eigenvektoren und Eigenwerte

chap:5

In diesem Kapitel sei V ein n-dimensionaler Vektorraum über dem Körper $\mathbb{K}$. Wir betrachten eine lineare Abbildung $\varphi : V \to V$. 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 in diesem Semester 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 $t^2 + 1 = 0$ hat keine reelle Lösung, aber zwei komplexe Lösungen, $\pm i$. Dieses Verhalten steht in starkem Kontrast zu den Systemen linearer Gleichungen, deren Lösbarkeit —wie wir im 1. Semester gesehen haben— 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 Eigenvektoren und Eigenwerte

sec:5.1

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

5.1.1. [Motivation] Sei V ein n-dimensionaler Vektorraum über dem Körper $\mathbb{K}$ und $\varphi : V \to V$ 7.1.1
eine lineare Abbildung.

Um φ durch eine Matrix A darstellen zu können, wählen wir eine Basis

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

von V und berechnen die Bilder der Basisvektoren

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

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 $[\varphi]_B^C$ bzgl. zweier Basen B und C, die jeweils zu Bild und Urbildbereich von φ gehören.

Würden wir zwei Basen zulassen, so könnten wir mit Beispiel (c) in 5.5.4 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

$$A = [\varphi]_B^C = \left(\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & 1 & \ddots & \vdots & \vdots & & & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots & & & \vdots \\
0 & \cdots & 0 & 1 & \vdots & & & \vdots \\
\hline
0 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & 0 \\
\vdots & & & & \vdots & & & \vdots \\
0 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & 0
\end{array}\right)$$

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.

Wann ist eine Matrix "möglichst einfach"? Die einfachsten Matrizen sind Diagonalmatrizen

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

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

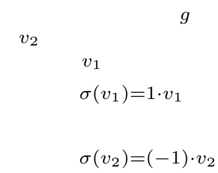
$$\varphi(v_j) = \lambda_j v_j \qquad \text{für } j = 1, \ldots, n.$$

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

bsp:7.1.2

Beispiele 5.1.3. 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



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

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 $\rho_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ 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:

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

Dabei ist $\operatorname{id} = \operatorname{id}_V$ 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.

Definition 5.1.4. [Eigenvektoren und Eigenwerte] 7.1.3

(a) Sei $\varphi \in \operatorname{End}(V)$ 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

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

heißt *Eigenraum zum Eigenwert* λ. Da der Kern einer linearen Abbildung immer ein linearer Teilraum ist (5.1.1), ist auch V_λ ein Untervektorraum von V. Die Dimension

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

[1] Darüber, 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.

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 $d_\lambda \geq 1$. 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

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

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

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

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 $\varphi : V \to V$ diagonalisierbar?

Bemerkung 5.1.5. Wir fassen noch einmal die wichtigsten Aspekte zusammen.

$$\begin{aligned} & \lambda \text{ ist ein Eigenwert von } \varphi \\ \Longleftrightarrow\ & \text{Die Gleichung } (\varphi - \lambda \operatorname{id})(v) = 0 \text{ hat eine Lösung } v \neq 0 \\ \Longleftrightarrow\ & \ker(\varphi - \lambda \operatorname{id}) \neq \{0\} \end{aligned}$$

Analog werden Eigenvektoren charakterisiert:

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

Bemerkung 5.1.6. Für $A \in M_n(\mathbb{K})$ ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von A, wenn das homogene System linearer Gleichungen

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

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:

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

wobei $E = E_n$ 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.

7.1.5

Satz 5.1.7. *Ist* $\varphi \in \operatorname{End}(V)$ *und* $\dim V < \infty$, *so ist* $\lambda \in \mathbb{K}$ *genau dann ein Eigenwert von* φ, *wenn* $\det(\varphi - \lambda \operatorname{id}) = 0$ *ist.*

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

$$\begin{aligned}\ker(\varphi - \lambda\,\mathrm{id}) = \{0\} &\iff \varphi - \lambda\,\mathrm{id} \text{ ist injektiv} && \text{(nach 5.1.3)}\\ &\iff \varphi - \lambda\,\mathrm{id} \text{ ist bijektiv} && \text{(nach 5.1.5)}\\ &\iff \det(\varphi - \lambda\,\mathrm{id}) \neq 0 && \text{(nach 6.4.5 (D9))}\end{aligned}$$

□

Folgerung 5.1.8. *Für $A \in M_n(\mathbb{K})$ gilt:* 7.2.3

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

Die letzte Formel heißt *charakteristische Gleichung* von A.

Beweis. Wir wissen aus 6.4.5 (D9), dass die Matrix $A - \lambda\mathbf{1}$ genau dann nicht invertierbar ist, wenn $\det(A - \lambda\mathbf{1}) = \mathbf{0}$ gilt. Wir können nun Satz 5.1.7 auf die lineare Abbildung $L_A \colon \mathbb{K}^n \to \mathbb{K}^n, x \mapsto Ax$ anwenden. Die Behauptung folgt dann aus

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

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

7.2.6

Beispiele 5.1.9. (a) Wie in Beispiel 5.1.3 (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 5.1.3(b) gilt:

Eigenwerte	$\lambda_1 = 1$	$\lambda_2 = 0$
Eigenvektoren, Eigenräume	wie oben	

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

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

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

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

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

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

führt auf die Eigenvektoren

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

7.2.5 **Beispiel 5.1.10.** (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

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

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

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

Eigenvektoren: Wir müssen beide Eigenwerte berücksichtigen.

- $\lambda_1 = 4$:
$$(A - \lambda_1\mathbf{1})\mathbf{v} = \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix}\mathbf{v} = \mathbf{0} \qquad \Longrightarrow \mathbf{v} \in \mathbb{R}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- $\lambda_2 = -2$:
$$(A - \lambda_2\mathbf{1})\mathbf{v} = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}\mathbf{v} = \mathbf{0} \qquad \Longrightarrow \mathbf{v} \in \mathbb{R}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

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 $\varphi_A : \mathbb{R}^2 \to \mathbb{R}^2$, 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).

Lineare Unabhängigkeit von Eigenvektoren

7.1.7

Satz 5.1.11. *Seien $\lambda_1, \lambda_2, \ldots, \lambda_r$ paarweise verschiedene Eigenwerte von $\varphi \in \mathrm{End}(V)$ und $v_1, \ldots, v_r$ zugehörige Eigenvektoren. Dann sind $v_1, \ldots, v_r$ linear unabhängig.*

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

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

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

Wenden wir φ auf diese Gleichung an, so erhalten wir

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

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

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

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

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

Wegen $\lambda_j \neq \lambda_k$ 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. □

7.1.7b

Folgerung 5.1.12. *Seien $\lambda_1, \lambda_2, \ldots, \lambda_r$ paarweise verschiedene Eigenwerte von φ. Dann ist die Summe der zugehörigen Eigenräume direkt:*

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

Beweis. Es seien

$$\begin{aligned} u &= u_1 + \ldots + u_{r-1} + u_r \\ u &= u_1' + \ldots + u_{r-1}' + u_r' \end{aligned} \tag{5.1}$$

zwei Darstellungen eines Vektors u durch Vektoren $u_i, u_i' \in V_{\lambda_i}$. Wir müssen zeigen, dass $u_i = u_i'$ für alle $i = 1, \ldots, r$ gilt. Aus (5.1) erhalten wir durch Subtraktion

$$0 = (u_1 - u_1') + \ldots + (u_{r-1} - u_{r-1}') + (u_r - u_r'). \tag{5.2}$$

Da die $u_j - u_j'$ entweder 0 oder Eigenvektoren zum Eigenwert λ_j sind, sind die von 0 verschiedenen Summanden nach Satz 5.1.11 linear unabhängig. Da ihre Summe 0 ergibt, müssen also alle Summanden verschwinden, und wir erhalten $u_j = u_j'$ für alle j. □

Bemerkung 5.1.13. Ein Endomorphismus φ eines n-dimensionalen Vektorraumes V kann höchstens n verschiedene Eigenwerte besitzen. Hat nämlich φ verschiedene Eigenwerte $\lambda_1, \ldots, \lambda_r$ mit zugehörigen Eigenvektoren $v_1, \ldots, v_r$, so sind diese nach 5.1.11 linear unabhängig. Also gilt $r \leq n$.

7.3.x

Satz 5.1.14. *Hat $A \in M_n(\mathbb{K})$ n verschiedene Eigenwerte, dann ist A diagonalisierbar.*

Beweis. Seien $\lambda_1, \ldots, \lambda_n$ verschiedene Eigenwerte von A und $v_1, \ldots, v_n$ zugehörige Eigenvektoren. Diese sind linear unabhängig 5.1.11 und bilden daher eine Basis von $\mathbb{K}^n$. Also ist A diagonalisierbar. □

7.1.9

Bemerkung 5.1.15. Seien $\lambda_1, \ldots, \lambda_r$ verschiedene Eigenwerte von φ und $d_1, \ldots, d_r$ deren geometrische Vielfachheiten. Wir finden dann jeweils eine d_i-elementige Basis von V_{λ_i}. Nehmen wir diese Basen zusammen, so erhalten wir eine Basis des Teilraums $V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_r}$. Gilt $d_1 + d_2 + \cdots + d_r = n$, 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:

$$\begin{array}{c} d_1\left\{\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right. \\ \\ \\ d_r\left\{\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right. \end{array} \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & & 0 \\ & \ddots & \\ 0 & & \boxed{\begin{matrix} \lambda_r & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{matrix}} \end{pmatrix}$$

7.2.7

Bemerkung 5.1.16. Die obigen Eigenschaften gelten entsprechend für Matrizen $A \in M_n(\mathbb{K})$, da man sie durch Anwendung auf die lineare Abbildung $L_A \colon \mathbb{K}^n \to \mathbb{K}^n, x \mapsto Ax$ übertragen kann: Seien $\lambda_1, \lambda_2, \ldots, \lambda_r$ paarweise verschiedene Eigenwerte einer $n \times n$-Matrix A und seien $d_1, d_2, \ldots, d_r$ ihre geometrischen Vielfachheiten.

a) Die Summe der Eigenräume von A ist direkt:

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

b) Sind $v_1, v_2, \ldots, v_r$ Eigenvektoren zu den paarweise verschiedenen Eigenwerten $\lambda_1, \lambda_2, \ldots, \lambda_r$, so sind diese Eigenvektoren linear unabhängig. Wir haben dies schon im Beweis von 7.1.14 verwendet. Hieraus folgern wir insbesondere $r \leq n$, d.h. eine $n \times n$-Matrix A hat höchstens n verschiedene Eigenwerte.

7.1.17

Aufgabe 5.1.17. 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 $\overline{v}$ ein Eigenvektor von A zum Eigenwert $\overline{\lambda}$. Hierbei steht $\overline{v} \in \mathbb{C}^n$ für den Vektor, der durch Konjugieren der Einträge von v entsteht.

5.2 Formale Polynome

sec:5.2

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

$$p(X) = a_0 + a_1 X + \ldots + a_n X^n,$$

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

Definition 5.2.1. Sei $\mathbb{K}$ ein Körper.

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

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

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

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

für $N \leq M$ als gleich, wenn

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

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

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

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*

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

(c) Ist $p(X) = a_0 + \ldots + a_N X^N \in \mathbb{K}[X]$ 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 Nullpolynome 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.

prop:deg

Satz 5.2.2. *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*

$$\deg(p(X)q(X)) = \deg(p(X)) + \deg(q(X)). \tag{5.3}$$

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

Remark 5.2.3. (a) Aus (5.3) 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 $p(X) = f(X)q(X) = \widetilde{f}(X)q(X)$ folgt $f(X) = \widetilde{f}(X)$ wenn $q(X) \neq 0$ ist, denn $(f(X) - \widetilde{f}(X))q(X) = 0$.

Bemerkung 5.2.4. Man kann jedem formalen Polynom

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

die Polynomfunktion

$$\widetilde{p}\colon \mathbb{K} \to \mathbb{K}, \quad x \mapsto a_0 + a_1 x + \ldots + a_n x^n$$

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

$$\Phi\colon \mathbb{K}[X] \to K[x], \quad p \mapsto \widetilde{p}$$

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 $\mathbb{K} \to \mathbb{K}$ 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 $q \colon \mathbb{K} \to \mathbb{K}$ verschwindet. Wie wir unten sehen werden, hat ein von Null verschiedenes Polynom vom Grad n maximal n verschiedene Nullstellen (Satz 5.2.7), 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 5.2.5. [Nullstellen von Polynomen] Ist $p(X) = a_0 + \ldots + a_N X^N \in \mathbb{K}[X]$, so heißt $\lambda \in \mathbb{K}$ *Nullstelle* oder *Wurzel* von $p(X)$, wenn

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

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

Lemma 5.2.6. *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

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

für gewisse $b_0, \ldots, b_d \in K$. 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 5.2.7. *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 $\mathrm{der}(q(X)) = n-1$. 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. □

Definition 5.2.8. 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 5.2.9. Es seien $\lambda_1, \ldots, \lambda_r$ die verschiedenen Nullstellen des Polynoms $p(X) \in \mathbb{K}[X]$ mit den Vielfachheiten $e_1, \ldots, e_r$.

Dann haben wir

$$p(X) = (X - \lambda_1)^{e_1} p_1(X)$$

und die Zahlen $\lambda_2, \ldots, \lambda_r$ sind keine Nullstellen von $(X - \lambda_1)^{e_1}$, 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 $r(X) = (X - \lambda_1)^{e_1} s(X)$ nicht in λ_i verschwindet. Also ist $e_i = f_i$. Induktiv erhalten wir so eine Faktorisierung

$$p(X) = (X - \lambda_1)^{e_1} \cdots (X - \lambda_r)^{e_r} q(X),$$

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

Theorem 5.2.10. (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:*

$$p(X) = c(X - \lambda_1)^{e_1} \cdots (X - \lambda_r)^{e_r}$$

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 5.2.11. 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.

Satz 5.2.12. [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

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

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

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

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

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

im Widerspruch zu Satz 5.2.2.

Existenz: Sei

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

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

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

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

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

Dann ist

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

und die Behauptung folgt mit

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

Aufgabe 5.2.13. Der obige Beweis liefert ein Verfahren zur Polynomdivision, bei dem der Grad von p bei jedem Schritt um 1 verkleinert wird, bis er kleiner als der Grad von q ist. Schreiben Sie dieses Verfahren als Algorithmus auf.

Beispiel 5.2.14. 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

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

5.3 Das charakteristische Polynom einer Matrix

sec:5.3

7.2.8 **Definition 5.3.1.** [Das charakteristische Polynom] Wir betrachten für eine $n{\times}n$-Matrix A die Determinante $\det(A - XE)$ als Element von $\mathbb{K}[X]$:

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

Das Polynom $\chi_A(X)$ heißt *charakteristisches Polynom* der Matrix A.

Beispiel 5.3.2. 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 5.3.3. *Das charakteristische Polynom $\chi_A(X)$ von A ist ein Polynom vom Grad n mit Koeffizienten aus $\mathbb{K}$. Weiter gilt*

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

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

Beweis. Berechnen wir $\det(A - XE)$ mittels der Formel von Leibniz nach 6.3.4, so erhalten wir

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

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

Bemerkung 5.3.4. Nach 5.1.8 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 (5.2.7), erhalten wir damit einen neuen Beweis für die Aussage in 5.1.16 (b), dass eine $n \times n$-Matrix höchstens n verschiedene Eigenwerte besitzt.

Definition 5.3.5. [Algebraische Vielfachheit] Wir nennen für einen Eigenwert λ von A die Vielfachheit e_λ der Nullstelle λ des charakteristischen Polynoms $\chi_A(X)$ die *algebraische Vielfachheit des Eigenwerts* λ. 7.2.9

Aufgabe 5.3.6. Für jeden Eigenwert λ einer Matrix A ist die geometrische Vielfachheit d_λ kleiner oder gleich der algebraischen Vielfachheit e_λ.

Beispiel 5.3.7. ($\mathbb{K} = \mathbb{R}$): Wir betrachten die Matrix $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Zur Berechnung der Eigenwerte betrachten wir das charakteristische Polynom 7.2.10

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

Also ist 1 der einzige Eigenwert, und er hat die algebraischen Vielfachheit $e_1 = 2$.

Eigenvektoren zum einzigen Eigenwert 1 erhalten wir als Lösungen von

$$(A - 1 \cdot E)v = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} v = 0 \qquad \Longrightarrow \qquad v \in \mathbb{R}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Also ist die geometrische Vielfachheit des Eigenwertes $\lambda_1 = 1$ gleich $d_1 = 1$. In diesem Beispiel ist $d_1 < e_1$; es gibt also keine Basis von Eigenvektoren.

Satz 5.3.8. (Bedingungen für Diagonalisierbarkeit von Matrizen) *Eine $n\times n$-Matrix A ist genau* beddiag
dann diagonalisierbar über $\mathbb{K}$, wenn ihr charakteristisches Polynom $\chi_A(X)$ in ein Produkt von Linearfaktoren

$$\chi_A(X) = (\lambda_1 - X)^{e_1} \dots (\lambda_r - X)^{e_r}$$

($\lambda_1, \dots, \lambda_r \in \mathbb{K}$) zerlegt werden kann und falls für jeden Eigenwert λ_j die geometrische Vielfachheit d_j gleich der algebraischen Vielfachheit e_j ist.

Beweis. Seien $\lambda_1, \dots, \lambda_r$ die verschiedenen Eigenwerte von A.

Ist A diagonalisierbar und sind $d_1, \dots, d_r$ die geometrischen Vielfachheiten von $\lambda_1, \dots, \lambda_r$, so existiert eine invertierbare Matrix S (deren Spalten die Eigenvektoren von A sind), so dass

$$A' := S^{-1}AS = \begin{matrix} d_1\left\{\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right. \\ \\ \\ d_r\left\{\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right. \end{matrix} \begin{pmatrix} \boxed{\begin{matrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_1\end{matrix}} & & 0 \\ & \ddots & \\ 0 & & \boxed{\begin{matrix}\lambda_r & & 0\\ & \ddots & \\ 0 & & \lambda_r\end{matrix}} \end{pmatrix}$$

gilt. Dann ist

$$\begin{aligned}\chi_A(X) &= \det(A - XE) = \det(SA'S^{-1} - XE) \\ &= \det(S(A' - tE)S^{-1}) = \det(A' - tE) = \chi_{A'}(X) = \prod_{j=1}^{r} (\lambda_j - X)^{d_j}.\end{aligned}$$

Insbesondere folgt $e_j = d_j$, die algebraischen und geometrischen Vielfachheiten stimmen also überein.

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

$$\chi_A(X) = (\lambda_1 - X)^{e_1} \cdots (\lambda_r - X)^{e_r}$$

und die geometrische Vielfachheit d_j des Eigenwerts λ_j mit der algebraischen Vielfachheit e_j übereinstimmt. Dann ist

$$n = \deg(\chi_A(X)) = e_1 + \dots + e_r = d_1 + \dots + d_r,$$

also folgt aus Bemerkung 5.1.15, dass A diagonalisierbar ist. □

7.3.9
7.3.5 **Folgerung 5.3.9.** [Spezialfall: $\mathbb{K} = \mathbb{C}$] *Eine komplexe Matrix $A \in M_n(\mathbb{C})$ ist genau dann diagonalisierbar, wenn für jeden Eigenwert $\lambda_j \in \mathbb{C}$ die algebraische und die geometrische Vielfachheit übereinstimmen.*

Beweis. Für eine komplexe Matrix A zerfällt das charakteristische Polynom $\chi_A(X) \in \mathbb{C}[t]$ nach dem Fundamentalsatz der Algebra in Linearfaktoren. Die Behauptung folgt also aus Satz 5.3.8. □

7.3.6 **Bemerkung 5.3.10.** [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 (5.3.9) 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 $A - \lambda E \in M_n(\mathbb{R})$ eine reelle Matrix. Sei $d_\lambda^{\mathbb{C}}$ die geometrische Vielfachheit von λ

als Eigenwert der komplexen Matrix $A \in M_n(\mathbb{C})$ und $d_\lambda^{\mathbb{R}}$ die geometrische Vielfachheit von λ als Eigenwert der reellen Matrix. Dann ist

$$d_\lambda^{\mathbb{R}} = n - \operatorname{rank}(A - \lambda E) = d_\lambda^{\mathbb{C}},$$

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

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

woraus mit Bemerkung 5.1.15 folgt, dass A reell diagonalisierbar ist.

5.4 Ähnlichkeit von Matrizen

sec:5.4

Definition 5.4.1. [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 $S \in \mathrm{GL}_n(\mathbb{K})$ gibt mit 7.3.1

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

Lemma 5.4.2. *Ähnlichkeit von Matrizen ist eine Äquivalenzrelation, d.h. für $A, A', A'' \in M_n(\mathbb{K})$ gilt:*

(i) $A \approx A$.

(ii) $A \approx A' \Rightarrow A' \approx A$.

(iii) $A \approx A', A' \approx A'' \Rightarrow A \approx A''$.

Beweis. (i) Für die Einheitsmatrix $S := E_n$ gilt $A = E_n^{-1}AE_n$, also $A \approx A$.

(ii) Gilt $A \approx A'$, so gibt es eine invertierbare Matrix S, so dass $A' = S^{-1}A'S$ 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 $A' \approx A$.

(iii) 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 $A \approx A''$. □

7.4.3a

Satz 5.4.3. [Bedeutung der Ähnlichkeit] *Zwei Matrizen A und A' in $M_n(\mathbb{K})$ sind genau dann ähnlich, wenn eine lineare Abbildung $\varphi : V \to V$ 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 $\varphi : V \to V$ bzgl. zweier Basen B und B' in V, so gibt es eine invertierbare Matrix $S = [\mathrm{id}]_{B'}^{B}$ (die Transformationsmatrix) mit $A' = S^{-1}AS$, d.h. $A \approx A'$.

Umgekehrt sei $A \approx A'$. Sei $V := \mathbb{K}^n$ und $\varphi\colon \mathbb{K}^n \to \mathbb{K}^n$ definiert durch $\varphi(v) := Av$. Dann gilt $[\varphi]_B = A$ bzgl. der Standardbasis B von $\mathbb{K}^n$. Nach Annahme gilt $A' = S^{-1}AS$ für eine invertierbare Matrix S. Die Spalten $v_1, \ldots, v_n$ von S bilden eine Basis B' von $\mathbb{K}^n$ mit $S = [\mathrm{id}]_B^{B'}$. 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. □

Aufgabe 5.4.4. Man beweise, dass ähnliche Matrizen dieselbe Determinante, dieselbe Spur und das gleiche charakteristische Polynom haben.

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.

Bemerkung 5.4.5. Eine Matrix $A \in M_n(\mathbb{K})$ ist genau dann diagonalisierbar, wenn es in $\mathbb{K}^n$ eine Basis $v_1, v_2, \ldots, v_n$ aus Eigenvektoren zu Eigenwerten $\lambda_2, \lambda_2, \ldots, \lambda_n$ von A gibt. Ist S die Transformationsmatrix mit den Eigenvektoren $v_1, \ldots, v_n$ als Spalten, so gilt

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

7.3.2 **Folgerung 5.4.6.** *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*

$$A' := S^{-1}AS \quad \textit{eine Diagonalmatrix ist.}$$

7.3.3 **Beispiele 5.4.7.** $[\mathbb{K} = \mathbb{R}, n = 2]$

(a) Die Matrix $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \in M_2(\mathbb{R})$ aus Beispiel 5.1.10 ist diagonalisierbar:

Eigenwerte: $\lambda_1 = 4, \lambda_2 = -2$.

Basis von Eigenvektoren: $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

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

(b) Die Matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in M_2(\mathbb{C})$ aus Beispiel 5.1.9 hat keine reellen Eigenwerte. Sie kann also über dem Körper der reellen Zahlen nicht diagonalisiert werden. A hat aber komplexe

$$\begin{array}{lll} \text{Eigenwerte} & \lambda_1 = 1 + i, & \lambda_2 = 1 - i \quad \text{und} \\ \text{Eigenvektoren} & v_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}, & v_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}. \end{array}$$

Also ist A über dem Körper der komplexen Zahlen diagonalisierbar:

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

Dieses Beispiel zeigt, dass Diagonalisierbarkeit vom betrachteten Körper $\mathbb{K}$ abhängen kann.

7.3.7 **5.4.8.** [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:

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

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

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

Das bedeutet, dass die Abbildung

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

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

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

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

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

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

$$A^n = SB^nS^{-1} \tag{v}$$

Beispiel 5.4.9. Wie in Beispiel 5.1.10 sei $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. Dann gilt $S^{-1}AS = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$ mit

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

Nun können wir Folgendes berechnen:

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

5.5 Trigonalisierung

sec:5.5

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

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

siehe Beispiele 5.3.7 und 5.4.7(c).

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 $e_1 = 2$ echt größer als seine geometrische Vielfachheit $d_1 = 1$ ist, denn $d_1 = 2 - \operatorname{rank}(A - E) = 1$.

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:

7.5.1a

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

Beweis. Nach dem Fundamentalsatz der Algebra hat das charakteristische Polynom $\chi_A(X) = \det(A - XE)$ eine Nullstelle, also A einen Eigenwert. □

Satz 5.5.2. [Trigonalisierung von 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 $S \in \mathrm{GL}_n(\mathbb{C})$, so dass* 7.4.2

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

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

$$T^{-1}AT = \left(\begin{array}{c|c} \lambda & * \cdots * \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right)$$

für eine Matrix $B \in M_{n-1}(\mathbb{C})$ hat. Anwendung der Induktionsannahme auf B ergibt eine Matrix V, so dass $V^{-1}BV$ obere Dreiecksgestalt hat. Nun sei

$$V' := \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & V \end{array}\right).$$

Damit ist $(V')^{-1}T^{-1}ATV'$ eine obere Dreiecksmatrix und $S := TV'$ ist die gesuchte Transformationsmatrix. □

7.4.1 **Folgerung 5.5.3.** [Trigonalisierung von Endomorphismen] *Sei $\varphi : V \to V$ ein Endomorphismus eines endlichdimensionalen Vektorraumes V über dem Körper $\mathbb{C}$ der* komplexen Zahlen*. Dann gibt es eine Basis $v_1, v_2, \ldots, v_n$ in $\mathbb{C}$, so dass die Matrix 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 5.4.3). □

5.6 Der Satz von Cayley–Hamilton

sec:5.6

Definition 5.6.1. [Anwendung eines Polynoms auf eine Matrix] Sei $A \in M_n(\mathbb{K})$ eine Matrix und

$$P = a_0 + a_1X + \ldots + a_nX^n \in \mathbb{K}[X]$$

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

$$P(A) := a_0E + a_1A + \ldots + a_nA^n.$$

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

Ist zum Beispiel $P = X^2 + 2X + 1$, so erhalten wir $P(A) = A^2 + 2A + E$. Ist $Q = X^2 + 4X + 4$, so gilt $Q(A) = A^2 + 4A + 4E$.

Man rechnet leicht nacht, dass die Abbildung

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

ein Homomorphismus von $\mathbb{K}$-Algebren ist, also eine lineare Abbildung, die zusätzlich mit der Multiplikation verträglich ist:

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

Definition 5.6.2. [Anwendung eines Polynoms auf eine lineare Abbildung] Genauso wie Matrizen, kann man auch lineare Abbildungen in Polynome einsetzen: Sei $\varphi : V \to V$ eine lineare Abbildung auf einem $\mathbb{K}$-Vektorraum und

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

ein Polynom, so definieren wir

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

Bemerkung 5.6.3. [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 $P(A) = (A - \mathbf{1})^{\mathbf{2}} = \mathbf{0}$. Andererseits ist $A - E \neq 0$. 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."

$$\boxed{\text{Matrizen haben Nullteiler; Polynome nicht!}}$$

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 λE, denn es gilt

$$P(\lambda E) = P(\lambda)E = 0.$$

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

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

Satz 5.6.4. [Cayley–Hamilton] *Sei $A \in M_n(\mathbb{K})$ und $\chi_A(X) = \det(A - XE)$ das charakteristisches Polynom. Die Einsetzung von A in das charakteristische Polynom $\chi_A(X)$ ergibt die Nullmatrix, d.h.* 7.5.1

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

Beweis. Sei $\operatorname{adj}(A - XE)$ die adjungierte Matrix von $A - XE$. Aus 6.3.6 folgt:

$$\operatorname{adj}(A - XE)^T \cdot (A - XE) = \det(A - tE) \cdot E = \chi_A(t) \cdot E \qquad (*)$$

im Sinne von Matrizen deren Einträge Polynome in t sind, also in $M_n(\mathbb{K}[X])$.

Die Einträge $b'_{ij} = (-1)^{i+j} \det B_{ij}$ in $\operatorname{adj}(A - XE)$ sind Polynome in t vom Grad $\leq n-1$, da B_{ij} die $(n-1) \times (n-1)$-Matrix ist, die man aus $(A - tE)$ durch Streichen der i-ten Zeile und der j-ten Spalte erhält. Wir erhalten also

$$\operatorname{adj}(A - XE)^\top = C_0 + C_1 t + \cdots + C_{n-1}t^{n-1}$$

mit $C_0, C_1, \ldots, C_{n-1} \in M_n(\mathbb{K})$, d.h. die C_j enthalten nicht die Variable t.

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

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

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

Aufgabe 5.6.5. Es ist instruktiv, sich zu überlegen, warum der folgende "einfache" Beweis des Satzes von Cayley-Hamilton nicht schlüssig ist:

$$p_A(A) = \det(A - A \cdot E) = \det(A - A) = 0.$$

Beispiele 5.6.6. (für $\mathbb{K} = \mathbb{R}$):

(a) Wie in 5.1.10 sei $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. Das charakteristische Polynom ist $\chi_A(t) = (1-t)^2 - 9$. Damit folgt aus dem Satz von Cayley–Hamilton $(E - A)^2 - 9E = 0$, d.h. es ist $A^2 - 2A - 8E = 0$ und damit $A^2 = 2A + 8E$. (Probe!)

Als Konsequenz kann man jede Potenz von A als Linearkombination der Matrizen A und E darstellen. Es gilt z.B.

$$A^3 = AA^2 = A(2A + 8E) = 2A^2 + 8A = 2(2A + 8E) + 8A = 12A + 16E.$$

(b) Wir betrachten wie in 5.1.9 die Matrix $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Das charakteristische Polynom ist $\chi_A(X) = (1-t)^2 + 1$. Nach dem Satz von Cayley–Hamilton folgt $\chi_A(A) = (E - A)^2 + E = 0$ und daraus $A^2 - 2A + 2E = 0$. Damit können wir $A^2 = 2(A - E)$ folgern. (Probe!)

Satz 5.6.7. *Für jede Matrix $A \in M_n(\mathbb{K})$ existiert genau ein Polynom M_A minimalen Grades mit Leitkoeffizient 1, für das*

$$M_A(A) = 0$$

gilt.

Das Polynom M_A heißt *Minimalpolynom der Matrix A*.

Beweis. Nach dem Satz von Cayley–Hamilton gilt $\chi_A(A) = 0$. Das Polynom $(-1)^n \chi_A$ hat den Leitkoeffizienten 1 und annulliert A. Also existiert ein Polynom M_A mit dem Leitkoeffizienten 1, das A annulliert, und unter allen Polynomen mit dieser Eigenschaft minimalen Grad d besitzt (Jede nichtleere Teilmenge von $\mathbb{N}_0$ hat ein minimales Element). Wegen obiger Überlegung ist $d \leq n$. Damit ist die Existenz von M_A gezeigt.

Zum Nachweis der Eindeutigkeit nehmen wir an, dass Q ein Polynom des gleichen Grades d mit Leitkoeffizient 1 und $Q(A) = 0$ ist. Da M_A und Q denselben Leitkoeffizienten haben, ist $R := M_A - Q$ ein Polynom vom Grad $\leq d - 1$ mit $R(A) = 0$. Ist $R \neq 0$ und c der Leitkoeffizient von R, so ist $c^{-1}R$ ein Polynom mit Leitkoeffizient 1, das A annulliert. Dies ist ein Widerspruch zur Minimalität von d, also ist $R = 0$, d.h. $Q = M_A$. □

7.6.7 **Lemma 5.6.8.** *Jedes Polynom S mit $S(A) = 0$ wird von M_A geteilt.*

Beweis. Polynomdivision ergibt Polynome Q, R mit $S = QM_A + R$, wobei der Grad von R echt kleiner als der von M_A ist. Da $S(A) = 0$ und $M_A(A) = 0$ gelten, ist auch $R(A) = 0$, was nach der Minimalität von M_A nur $R = 0$ zulässt. Damit folgt $S = QM_A$. □

Bemerkung 5.6.9. Nicht alle lineare Abbildungen haben annullierende Polynome. Ist z.B. $V = \mathcal{F}(\mathbb{N}, \mathbb{R})$ der unendlich-dimensionale Vektorraum der Folgen von reellen Zahlen, dann wird die Verschiebungsabbildung $\varphi : V \to V$ mit $\varphi(a_0, a_1, a_2, \ldots) = (0, a_0, a_1, \ldots)$ durch kein Polynom annulliert. Ist $e_n \in V$ die Folge, die genau eine 1 an der n-ten Stelle hat und sonst Nullen, so gilt für $P = a_0 + a_1 t + \ldots + a_n t^n$ die Beziehung

$$P(\varphi)(e_0) = a_0 e_0 + a_1 e_1 + \ldots + a_n e_n,$$

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:

minsim

Satz 5.6.10. *Ähnliche Matrizen haben das gleiche Minimalpolynom.*

Beweis. Seien $A, B \in M_n(\mathbb{K})$ ähnlich und $S \in \mathrm{GL}_n(\mathbb{K})$ mit $A = S^{-1}BS$. Damit ergibt sich $M_B(A) = M_B(S^{-1}BS) = S^{-1}M_B(B)S = 0$. Also annulliert M_B die Matrix A und ebenso annulliert M_A die Matrix B. Hieraus folgt, dass M_A ein Teiler von M_B ist und umgekehrt. Also haben beide denselben Grad. Aufgrund der Minimalität folgt nun $M_A = M_B$. □

5.6.11. [Berechnung des Minimalpolynoms] Wir wissen bereits, dass das Minimalpolynom M_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:

$$\chi_A(X) = (\lambda_1 - X)^{e_1} \cdots (\lambda_r - X)^{e_r},$$

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

$$M_A(X) = (t - \lambda_1)^{k_1} \cdots (t - \lambda_r)^{k_r}$$

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

$$(t - \lambda_1)^{k_1} \cdots (t - \lambda_r)^{k_r}$$

die Matrix A annullieren und hat $k_1, \ldots, k_r$ 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 5.6.12. *Sei A eine Matrix. Dann haben das charakteristische Polynom χ_A und das Minimalpolynom M_A die gleichen Nullstellen.*

Beweis. Aus dem Satz von Cayley–Hamilton und Lemma 5.6.8 folgt, dass M_A das charakteristische Polynom χ_A teilt. Insbesondere ist jede Nullstelle von M_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

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

Also ist $M_A(\lambda) = 0$. □

Beispiel 5.6.13. Sei

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

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

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

Hieraus folgt schnell $(A - 2E)^2 = 0$, und wegen $A - 2E \neq 0$ ist $M_A(X) = (t - 2)^2$.

Aufgabe 5.6.14. 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 $V_k := \mathrm{im}(A^{d-k})$ mit

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

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.

7.6.13 **Aufgabe 5.6.15.**

1. Sei $A \in M_n(\mathbb{K})$ mit $A^d = 0$. Dann ist $\chi_A(X) = (-1)^n t^n$. Hinweis: Der Fall $\mathbb{K} = \mathbb{C}$ ist etwas leichter.
2. Sei $A \in M_n(\mathbb{K})$ und $\lambda \in \mathbb{K}$ mit $(A - \lambda E)^d = 0$. Dann ist $\chi_A(X) = (\lambda - X)^n$.

Kapitel 6

Die Jordansche Normalform

chap11

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

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

Einige Spezialfälle von Jordanblöcken:

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

$$J_{2,0} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad J_{3,0} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Lemma 6.1.2. *Ein Jordanblock $J_{m,\lambda}$ mit $m > 1$ ist* nicht diagonalisierbar. *Sein Eigenraum zum Eigenwert λ ist eindimensional.*

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

Die Eigenvektoren x zum Eigenwert λ erfüllen die Gleichung

$$0 = (J - \lambda E)x = (x_2, x_3, \ldots, x_n, 0)^\top,$$

das bedeutet $x_2 = 0, x_3 = 0, \ldots, x_m = 0$ und $x_1 \in \mathbb{K}$ ist beliebig. Also ist die geometrische Vielfachheit d_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 $n \geq 2$. □

11.1.3 **Lemma 6.1.3.** *Das Minimalpolynom eines Jordanblocks $J_{m,\lambda}$ ist $M_J(X) = (t - \lambda)^m$.*

Beweis. Sei $J := J_{m,\lambda}$. Wir haben oben schon gesehen, dass das charakteristische Polynom $\chi_J(X) = (\lambda - X)^m$ ist. Da das Minimalpolynom ein Teiler von χ_J ist, hat es die Gestalt $(t - \lambda)^k$ für ein $k \in \{1, \ldots, m\}$.

Für die Standardbasis $e_1, \ldots, e_m$ von $\mathbb{K}^m$ gilt

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

Für $k < m$ ergibt sich insbesondere $(J - \lambda E)^k \neq 0$ und weiter $(J - \lambda E)^m = 0$ folgt. Folglich ist das Minimalpolynom $M_J(X) = (t - \lambda)^m$. □

Definition 6.1.4. 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 $J_1, J_2, \ldots, J_t$ ein Jordanblock ist.

Satz 6.1.5. [Hauptsatz über die Jordansche Normalform] *Sei V ein n-dimensionaler komplexer Vektorraum und $\varphi \in \mathrm{End}(V)$. Dann existiert eine Basis $B = (v_1, v_2, \ldots, v_n)$ von V, so dass die Matrix $[\varphi]_B$ von φ bzgl. B Jordansche Normalform besitzt.*

Solch eine Basis B heißt *Jordanbasis für φ*.

12.1.1 **Folgerung 6.1.6.** *Sei $A \in M_n(\mathbb{C})$. 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 6.1.7. Im Lichte der Folgerung 6.1.6 erhalten wir den folgenden Test für die Ähnlichkeit von Matrizen: 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.

Beispiel 6.1.8. 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:

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

Die folgenden Matrizen sind *nicht* in Jordanscher Normalform:

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

Bemerkung 6.1.9. 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 $e_i := e_{\lambda_i}$. Verschiedene Jordanblöcke können zu demselben Eigenwert λ_i gehören: **Die Zahl der Jordanblöcke zum Eigenwert λ_i ist die geometrische Vielfachheit $d_i := d_{\lambda_i}$ des Eigenwertes λ_i.**

Bemerkung 6.1.10. Aus Lemma 6.1.3 erhalten wir leicht, dass das Minimalpolynom einer Matrix A in Jordanscher Normalform, die aus Blöcken $J_{m_i,\lambda}$ besteht, gegeben ist durch

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

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 6.1.3 heißt, dass es von $(t - \lambda)^p$ geteilt wird.

Besteht A aus Jordanblöcken zu verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$ 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 6.1.3, dass das Minimalpolynom von A gegeben ist durch

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

6.2 Der Beweis des Hauptsatzes

6.2.1. [Einleitende Betrachtungen] Angenommen, $B = (v_1, \ldots, v_m, v_{m+1}, \ldots, v_n)$ ist eine Basis 12.1.2
des Vektorraums V, so dass die Matrix $A = [\varphi]_B$ von $\varphi \in \mathrm{End}(V)$ bzgl. B Jordansche Normalform besitzt, wobei die ersten m Basisvektoren $v_1, \ldots, v_m$ zum ersten Jordanblock

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

gehören. Dann ergibt sich

$$\begin{aligned} \varphi(v_1) &= \lambda v_1 & \qquad A v_1 &= \lambda v_1 \\ \varphi(v_2) &= v_1 + \lambda v_2 & A v_2 &= v_1 + \lambda v_2 \\ &\vdots & &\vdots \\ \varphi(v_m) &= v_{m-1} + \lambda v_m & A v_m &= v_{m-1} + \lambda v_m, \end{aligned}$$

oder anders geschrieben

$$\begin{aligned} (\varphi - \lambda\,\mathrm{id})v_1 &= 0 & (A-\lambda E)v_1 &= 0\\ (\varphi - \lambda\,\mathrm{id})v_2 &= v_1 & (A-\lambda E)v_2 &= v_1\\ &\vdots & &\vdots\\ (\varphi - \lambda\,\mathrm{id})v_m &= v_{m-1} & (A-\lambda E)v_m &= v_{m-1}. \end{aligned} \tag{$*$}$$

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

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

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

Lemma 6.2.3. *Bilden die Vektoren $v_1, \ldots, v_m$ 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 $k = \max\{i \colon \alpha_i \neq 0\}$. Beachte, dass $k \geq 2$ gelten muss. Dann ist

$$(\varphi - \lambda\mathrm{id})^{k-1} v_k = v_1,$$

ebenso

$$(\varphi - \lambda\mathrm{id})^{k-1} v_j = 0$$

für $j < k$, also

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

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

12.1.4 **Definition 6.2.4.** [Verallgemeinerte Eigenvektoren] Sei $\varphi \in \mathrm{End}(V)$. 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

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

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

$$v_1 := (\varphi - \lambda\,\mathrm{id})^{m-1} v, \qquad \ldots \qquad v_{m-1} := (\varphi - \lambda\,\mathrm{id})v, \qquad v_m := v$$

eine Jordankette, also linear unabhängig und damit $m \leq \dim V$. (Man kann insbesondere zeigen, dass $m \leq e_\lambda$ gilt, wobei e_λ die algebraische Vielfachheit von λ ist.)

Wir schreiben

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

für die Menge der verallgemeinerten Eigenvektoren von φ zum Eigenwert λ. 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 $L_A \colon \mathbb{K}^n \to \mathbb{K}^n, x \mapsto Ax$ und überträgt die Begriffe entsprechend.

Lemma 6.2.5. *Vertauschen φ und $\psi \in \mathrm{End}(V)$ miteinander, so sind die verallgemeinerten 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 $v \in V^\lambda(\varphi)$. Dann existiert ein $m \in \mathbb{N}$ mit $(\varphi - \lambda\,\mathrm{id})^m v = 0$. Da ψ mit φ vertauscht, vertauscht es auch mit $\varphi - \lambda\,\mathrm{id}$ und somit auch mit der Potenz $(\varphi - \lambda\,\mathrm{id})^m$. Damit erhalten wir

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

also $\psi(v) \in V^\lambda(\varphi)$. □

Wir werden nun den Hauptsatz beweisen, indem wir zuerst zeigen, dass für $\varphi \in \mathrm{End}(V)$, $\mathbb{K} = \mathbb{C}$ und $\dim V < \infty$ der Raum V die direkte Summe der verallgemeinerten Eigenräume ist. Zuerst benötigen wir noch ein paar Fakten über formale Polynome.

Definition 6.2.6. Eine Teilmenge $I \subseteq \mathbb{K}[X]$ heißt *Ideal*, wenn I ein Untervektorraum ist, und zusätzlich

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

gilt, d.h. für $f, g \in I$ und $h \in \mathbb{K}[X]$ sind $f + g, fh \in I$.

Beispiel 6.2.7. 1. Ist $A \in M_n(\mathbb{K})$, so ist

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

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 $fg \in I_A$.

Das Ideal I_A enthält das charakteristische Polynom χ_A, und für das Minimalpolynom M_A gilt

$$I_A = M_A \cdot \mathbb{K}[X],$$

denn wir haben in Lemma 7.6.8 gesehen, dass M_A jedes Polynom in I_A teilt.

2. Ist $\varphi \in \mathrm{End}(V)$, so ist

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

ein Ideal. Das sieht man genauso, wie in 1) für Matrizen.

11.2.8

Lemma 6.2.8. *Ist $I \subseteq \mathbb{K}[X]$ ein Ideal, so existiert ein Polynom $f \in I$ mit*

$$I = f \cdot \mathbb{K}[X].$$

Wir schreiben auch $I = (f) := f \cdot \mathbb{K}[X]$ für das von f erzeugte Ideal von $\mathbb{K}[X]$. Ideale, die von einem Element erzeugt werden, heißen Hauptideale. Das Lemma besagt also, dass jedes Ideal von $\mathbb{K}[X]$ ein Hauptideal ist. Man nennt $\mathbb{K}[X]$ daher auch einen *Hauptidealring*.

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

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

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

nicht leer, enthält also ein minimales Element d. Sei $f \in I$ mit $\deg(f) = d$. 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 $I = f \cdot \mathbb{K}[X]$. □

dirsum

Lemma 6.2.9. *Sei V ein Vektorraum und $E_1, \ldots, E_n \in \mathrm{End}(V)$ Elemente mit*

$$E_1 + \ldots + E_n = \mathrm{id}_V \qquad \textit{und} \qquad E_i E_j = \delta_{ij} E_i.$$

Dann ist V die direkte Summe der Unterräume $V_i := \mathrm{im}(E_i)$.

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

$$v = \mathrm{id}_V(v) = \sum_{j=1}^{m} E_i(v) \in V_1 + \ldots + V_m,$$

also

$$V = V_1 + \ldots + V_m.$$

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

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

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

12.1.5 **Satz 6.2.10.** (Zerlegung in verallgemeinerte Eigenräume) *Sei V ein endlichdimensionaler komplexer Vektorraum und $\varphi \in \mathrm{End}(V)$ mit den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_m$. 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. Sei zunächst $f \in \mathbb{C}[t]$ das Minimalpolynom von φ, sowie

$$f = (t - \lambda_1)^{k_1}(t - \lambda_2)^{k_2} \cdots (t - \lambda_m)^{k_m}$$

die Zerlegung in Linearfaktoren, die nach dem Fundamentalsatz der Algebra existiert. Sei $f_i := f/(t - \lambda_i)^{k_i}$. Wir betrachten das Ideal

$$I = (f_1) + \ldots + (f_m) \subseteq \mathbb{C}[t].$$

Dieses Ideal wird von einem Element g erzeugt (Lemma 6.2.8), das gemeinsamer Teiler aller Polynome f_i ist. Also hat g keine Nullstelle und ist somit konstant. Also ist $I = g \cdot \mathbb{C}[t] = \mathbb{C}[t]$. Wir finden daher Polynome $r_1, \ldots, r_m \in \mathbb{C}[t]$ mit

$$(*) \qquad 1 = r_1 f_1 + \ldots + r_m f_m.$$

Sei $E_i := (r_i f_i)(\varphi) \in \mathrm{End}(V)$. Wenden wir die Relation (*) auf den Endomorphismus φ an, so erhalten wir

$$E_1 + \ldots + E_m = \mathrm{id}_V .$$

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 $r_i f_i r_j f_j$. Wegen $f(\varphi) = 0$ ist daher

$$E_i E_j \;=\; 0.$$

Daraus folgt $E_i^2 = E_i\big(\sum_{j=1}^m E_j\big) = E_i \cdot \mathrm{id}_V = E_i$. Lemma 6.2.9 zeigt nun, dass V direkte Summe der Unterräume $V_i := \mathrm{im}(E_i) = E_i(V)$ ist:

$$V = V_1 \oplus \ldots \oplus V_m.$$

Wir zeigen nun, dass V_i der verallgemeinerte Eigenraum zum Eigenwert λ_i ist. Da φ mit E_i vertauscht, läßt φ die Unterräume V_i invariant. Damit läßt auch $f_i(\varphi)$ den Unterraum V_i invariant. Wegen

$$\mathrm{id}_{V_i} = E_i|_{V_i} = r_i(\varphi)f_i(\varphi)|_{V_i}$$

ist die Einschränkung von $f_i(\varphi)$ auf V_i sogar invertierbar. Mit $f(\varphi) = 0$ folgt daher

$$(\varphi - \lambda_i\,\mathrm{id})^{k_i}(V_i) = (\varphi - \lambda_i\,\mathrm{id})^{k_i}\big(f_i(\varphi)(V_i)\big) = f(\varphi)(V_i) = \{0\},$$

d.h. $V_i \subseteq V^{\lambda_i}(\varphi)$.

Wir zeigen jetzt noch, dass $V_i \supseteq V^{\lambda_i}(\varphi)$ ist. Sei also $v \in V^{\lambda_i}(\varphi)$, $(\varphi - \lambda_i\,\mathrm{id})^k(v) = 0$ sowie $v = \sum_{j=1}^m v_j$ mit $v_j \in V_j$. Dann ist

$$0 = (\varphi - \lambda_i\,\mathrm{id})^k(v) = \sum_{j=1}^m (\varphi - \lambda_i\,\mathrm{id})^k(v_j).$$

Aus der Invarianz aller Unterräume V_j unter φ folgt $(\varphi - \lambda_i\,\mathrm{id})^k(v_j) \in V_j$, und aus der Direktheit der Zerlegung erhalten wir damit $(\varphi - \lambda_i\,\mathrm{id})^k(v_j) = 0$, insbesondere $v_j \in V^{\lambda_i}(\varphi)$.

Ist $v_j \neq 0$, so existiert ein Eigenvektor $v_j' \in V^{\lambda_i}(\varphi) \cap V_j$. Hierzu setze man

$$v_j' = (\varphi - \lambda_i\,\mathrm{id})^l(v_j),$$

wobei l maximal ist mit der Eigenschaft, dass dieser Vektor nicht verschwindet. Nun existiert ein $d \in \mathbb{N}$ mit

$$(\varphi - \lambda_j\,\mathrm{id})^d(v_j') = (\lambda_i - \lambda_j)^d \cdot v_j' = 0,$$

also $\lambda_j = \lambda_i$, d.h. $j = i$. Hieraus folgt $v = v_i \in V_i$ und damit $V^{\lambda_i}(\varphi) = V_i$. □

11.2.10

Lemma 6.2.11. *Sei $\mathcal{C} = (v_1^1, \dots, v_{l_1}^1, v_2^2, \dots, v_{l_2}^2, \dots, v_1^s, \dots, v_{l_s}^s)$ eine Familie von s Jordanketten zum Eigenwert λ von φ, d.h*

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

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 $(\varphi - \lambda\,\mathrm{id})^{k-1}$ an, so erhalten wir eine lineare Relation zwischen den Eigenvektoren

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

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

reduc

Lemma 6.2.12. *Sei $\mathcal{C} = (v_1^1, \dots, v_{l_1}^1, v_2^2, \dots, v_{l_2}^2, \dots, v_1^s, \dots, v_{l_s}^s)$ eine* linear abhängige *Familie von Jordanketten. Dann gibt es eine Familie von Jordanketten $\mathcal{C}'$, so dass $\mathrm{span}(\mathcal{C}') = \mathrm{span}(\mathcal{C})$ ist und $\mathcal{C}'$ einen Vektor weniger als $\mathcal{C}$ enthält.*

Beweis. Nach dem vorigen Lemma sind die Eigenvektoren v_1^i linear abhängig, z.B. mit

$$\sum_{i=1}^s \alpha_i v_1^i = 0.$$

Wir wählen aus den Jordanketten $v_1^i, \dots, v_{l_i}^i$, für die $\alpha_i \neq 0$ ist, eine von minimaler Länge aus und zeigen, dass diese Jordankette durch eine von kürzerer Länge ersetzt werden kann. Wir nehmen o.B.d.A. an, dass $v_1^1, \dots, v_{l_1}^1$ die erste Jordankette mit dieser Eigenschaft ist, d.h. $\alpha_1 \neq 0$ und $l_1 \leq l_i$, falls $\alpha_i \neq 0$.

Für $l_1 = 1$ können wir die Kette v_1^1 einfach entfernen, da ihr alleiniges Element eine Linearkombination anderer Vektoren ist.

Für $l_1 \geq 2$ definieren wir

$$\widetilde{v}_j^1 = v_{j+1}^1 + \sum_{\alpha_i \neq 0} \frac{\alpha_i}{\alpha_1} v_{j+1}^i$$

für $j = 1, \ldots, l_1 - 1$, wobei die Summe über die i läuft, für die $\alpha_i \neq 0$ ist. Man beachte, dass wegen der Minimalitätsannahme v_{j+1}^i in diesem Fall existiert.

Die Vektoren $\widetilde{v}_1^1, \ldots, \widetilde{v}_{l_1-1}^1$ bilden nun eine Jordankette der Länge $l_1 - 1$, da $\widetilde{v}_0 = 0$ ist, und eine Ersetzung der Kette $v_1^1, \ldots, v_{l_1}^1$ durch $\widetilde{v}_{l_1-1}^1, \ldots, \widetilde{v}_1^1$ verändert nicht den aufgespannten Unterraum. Man erhält also eine kürzere Kette $\mathcal{C}'$, d.h. eine Kette, die weniger Vektoren enthält. □

Satz 6.2.13. *Für jeden Eigenwert λ von φ besitzt der verallgemeinerte Eigenraum $V^\lambda(\varphi)$ eine Jordanbasis.*

Beweis. Sei $u_1, \ldots, u_d$ eine Basis von $V^\lambda(\varphi)$. Für jedes der u_i bilden wir die zugehörige Jordankette $u_i, (\varphi - \lambda\,\mathrm{id})u_i, \ldots, (\varphi - \lambda\,\mathrm{id})^{k-1}u_i$, wobei k die Stufe des verallgemeinerten Eigenvektors u_i ist.

Sei $\mathcal{C}$ die Vereinigung aller Jordanketten, die bei den u_i begonnen haben. Beachte, dass obwohl $u_1, \ldots, u_d$ eine Basis von Eigenvektoren ist, $\mathcal{C}$ i.a. linear abhängig ist, da es mehr als d Elemente enthält. Mehrfaches Anwenden von Lemma 6.2.12 führt auf die gesuchte Jordanbasis. □

Beweis des Hauptsatzes über die Jordansche Normalform: Wir haben damit die Existenz einer Jordanbasis für jedes V^{λ_i} gezeigt. Da wir nach Satz 6.2.10 eine direkte Zerlegung $V = V^{\lambda_1} \oplus \cdots \oplus V^{\lambda_m}$ haben, erhalten wir eine Jordanbasis für V, indem wir einfach Jordanbasen für die verallgemeinerten Eigenräume kombinieren. Ist B eine Jordanbasis von V, so die Matrix $[\varphi]_B$ von φ bzgl. B in Jordanscher Normalform. Das folgt aus der Art, wie φ auf Basisvektoren wirkt. □

Fassen wir die entscheidenden Schritte, die auf eine Jordanbasis führen, zusammen.

1. Bestimme die Eigenwerte λ_i von A mit ihren algebraischen Vielfachheiten $e_i := e_{\lambda_i}$.

2. Bestimme für jeden Eigenwert λ_i eine Basis des verallgemeinerten Eigenraums V^{λ_i}.

   Dies tut man praktisch durch Lösen des homogenen Systems linearer Gleichungen

   $$(\varphi - \lambda\,\mathrm{id})^k v = 0$$

   und zwar schrittweise für $k = 1, \ldots,$ bis man e_i linear unabhängige verallgemeinerte Eigenvektoren gefunden hat.

3. Für jeden Eigenwert λ bilde die Jordanketten zu den verallgemeinerten Eigenvektoren bzgl. der Basis von V^λ, um ein (i.a. linear abhängiges) System von Jordanketten $\mathcal{C}$ zu erzeugen. Verringere die Größe dieses Systems durch Subtraktion von Linearkombinationen von längeren Ketten von kürzeren Ketten analog zum Beweis von Lemma 6.2.12, bis eine Basis erhalten wird.

4. 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.

5. Bilde die Jordansche Normalform, indem für jede Jordankette der Länge l und Eigenwert λ ein Jordanblock $J_{l,\lambda}$ genommen wird. Achte darauf, dass die Ordung der Jordanblöcke zur entsprechenden Ordnung der Jordanketten passt.

11.2.13

Beispiel 6.2.14. Sei

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

Das charakteristische Polynom von A ist $\chi_A(X) = (t-1)^5$, also ist $\lambda = 1$ der einzige Eigenwert mit algebraischer Vielfachheit $e = 5$.

In diesem Fall gilt $\mathbb{C}^5 = V^\lambda(A)$, also bildet z.B.

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, v_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

eine Basis von $V^\lambda(A)$. Wir hätten natürlich auch e_1, e_2, e_3, e_4, e_5 wählen können.

Es ist $(A-E)v_1 = w$ und $(A-E)v_2 = -w$ mit

$$w = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ -1 \end{pmatrix},$$

$(A-E)v_3 = (A-E)v_4 = v_5 = e_3$ und $(A-E)v_5 = 0$.

Die von v_1 und v_2 beginnenden Jordanketten sind also linear abhängig und wir können v_1, w durch

$$\widetilde{v}_1 = v_1 + \frac{1}{1} v_2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

ersetzen, was einen Eigenvektor liefert.

Weiterhin ist $(A-E)v_3 = (A-E)v_4 = v_5$, was ein Eigenvektor ist. Die Jordanketten, die von v_3 und v_4 beginnen, sind also linear abhängig und wir können v_3 durch

$$\widetilde{v}_3 = v_3 - \frac{1}{1} v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2 \end{pmatrix}$$

ersetzen, was ein Eigenvektor ist.

Wir haben folgendes System von Jordanketten erhalten: $\widetilde{v}_1, v_2 \to -w, \widetilde{v}_3, v_4 \to v_5$.

Die vier Eigenvektoren sind linear abhängig: $-w - v_5 - \frac{1}{2}\widetilde{v}_3 + \frac{1}{2}\widetilde{v}_1 = 0$, wir entfernen $\widetilde{v}_1$, um ein linear unabhängiges System von Jordanketten: $-w, v_2, v_5, v_4, \widetilde{v}_3$ zu erhalten, d.h. die Transformationsmatrix

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

Beachte, dass die Jordanketten in S eingesetzt werden, indem mit dem Eigenvektor begonnen wird!

Jetzt gilt tatsächlich

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

Bemerkung 6.2.15. Alternativ kann mach auch, um eine Jordanbasis zu erhalten, rückwärts von den Eigenvektoren für einen Eigenwert λ arbeiten. D.h. man startet mit einer Basis $v_1^1, \ldots, v_1^t$ des Eigenraums $V_\lambda(A)$ zum Eigenwert λ, die man wie folgt konstruiert:

Sei $(A - \lambda E)^k = 0$ aber $(A - \lambda E)^{k-1} \neq 0$. Dann definieren wir

$$F_m := \operatorname{im}(A - \lambda E)^{k-m} \cap V_\lambda(A), \quad m = 0, \ldots, k.$$

Dann ist

$$F_0 = \{0\} \subseteq F_1 = \operatorname{im}(A - \lambda E)^k \subseteq F_2 \subseteq \ldots \subseteq F_k = V_\lambda(A).$$

Zuerst bestimmt man nun eine Basis $v_1^1, \ldots, v_1^{d_1}$ von F_1, dann ergänzt man diese zu einer Basis $v_1^{d_1+1}, \ldots, v_1^{d_2}$ von F_2 usw., und erhält schließlich eine Basis $v_1^1, \ldots, v_1^{d_k}$ des Eigenraums $V_\lambda(A)$.

Für $d_{m-1} < j \leq d_m$ zeigt die Konstruktion der Basis, die Existenz von Vektoren v_{k+1-m}^j mit $(A - \lambda E)^{k-m} v_{k+1-m}^j = v_1^j$. Definieren wir nun

$$v_l^j := (A - \lambda E)^{k+1-m-l} v_{k+1-m}^j, \quad 1 \leq l \leq k+1-m,$$

so erhalten wir Jordanketten

$$v_{k+1-m}^j, \ldots, v_1^j$$

der Länge $k+1-m$, $m = 1, \ldots, k$. Da die Vektoren v_1^j linear unabhängig sind, gilt dies nach Lemma 6.2.11 auch für das System aller so konstruierten Vektoren v_i^j. In der Tat bildet dieses System sogar automatisch eine Jordanbasis von $\mathbb{C}^n$.

In Beispiel 6.2.14 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 $A - \lambda E$ ist, d.h. das Ende einer Jordankette der Länge > 1. Allgemeiner treten solche Beispiele wie folgt auf: ist v ein Eigenvektor, so dass $(A - \lambda E)x = v$ keine Lösungen hat und v_1 ein Eigenvektor, der eine Jordankette der Länge > 1 terminiert, d.h. $v_1 = (A - \lambda E)v_2$, dann ist $v_1 + v$ ein anderer Eigenvektor, so dass $(A - \lambda E)x = (v_1 + v)$ keine Lösung hat.

6.2.16. [Eindeutigkeit der Jordanschen Normalform] 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.

Dazu bilden wir für jeden Eigenwert λ die (aufsteigende) Folge von ganzen Zahlen

$$\begin{array}{l} \dim(\ker(A - \lambda E)) \\ \dim(\ker(A - \lambda E)^2) \\ \dim(\ker(A - \lambda E)^3) \\ \dots \end{array}$$

Es ist offensichtlich, dass ähnliche Matrizen dieselbe Folge erzeugen, andererseits bestimmt diese Folge die Größe und Zahl der Jordanblöcke, was man am leichtesten durch ein Beispiel sieht: Es sei für eine Matrix A

$$\begin{array}{l} \dim(\ker(A - \lambda E)) = 5 = d_\lambda, \\ \dim(\ker(A - \lambda E)^2) = 9, \\ \dim(\ker(A - \lambda E)^3) = 11, \\ \dim(\ker(A - \lambda E)^4) = 12 = e_\lambda \end{array}$$

Ersetzen wir A durch eine zu A ähnliche Matrix, insbesondere durch eine Jordansche Normalform von A, so erhalten wir dieselbe Sequenz von Zahlen und daraus ersehen wir, dass wir in unserer Jordanbasis 5 Eigenvektoren, $4 = 9 - 5$ verallgemeinerte Eigenvektoren der Stufe zwei, $2 = 11 - 9$ verallgemeinerte Eigenvektoren der Stufe drei und $1 = 12 - 11$ verallgemeinerte Eigenvektoren der höchsten Stufe vier haben. Der einzige Weg um das zu erreichen, besteht aus einer Jordankette der Länge 4, einer der Länge 3, zwei der Länge 2 und einem einzelnen Eigenvektor.

Allgemeiner ist die Zahl der Jordanketten der Länge k gleich

$$\dim(\ker(A - \lambda E)^k) - \dim(\ker(A - \lambda E)^{k-1}).$$

6.2.17. [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 $D = \lambda E$ 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

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

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 6.2.14, so gilt

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

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

Beispiel 6.2.18. Man erinnere sich, dass sich die Lösung einer linearen Rekursion erster Ordnung

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

mit $f : \mathbb{N} \to \mathbb{C}^n, g \in \mathbb{C}^n, A \in M_n(\mathbb{C})$ zu $f(n) = A^n g$ 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:

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

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

$$f(k+1) = (A+2E)f(k), f(0) = g$$

wobei A die Matrix aus Beispiel 6.2.14 ist. Wegen $S^{-1}AS = J$ gilt auch $S^{-1}(A+2E)S = J+2E$, woraus wir die Jordansche Normalform von $A+2E$ ohne weitere Berechnung erhalten.

Aus dem Obigen erkennen wir

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

woraus wir folgende Lösung in geschlossener Form erhalten:

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

oder

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

Damit können wir Matrixpotenzen benutzen, um eine geschlossene Form für $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$ zu erhalten und damit Systeme linearer Differenzialgleichungen lösen. Sie werden darüber mehr in späteren Vorlesungen erfahren.

Kapitel 7

Euklidische und unitäre Vektorräume

chap8

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. Wir betrachten zunächst das Standardskalarprodukt auf $\mathbb{R}^n$ und $\mathbb{C}^n$. Danach werden wir eine allgemeinere Definition eines Skalarproduktes auf einem reellen oder komplexen Vektorraum einführen.

7.1 $\mathbb{R}^n$ als euklidischer Vektorraum

sec8.1

Definition 7.1.1. [Das Standardskalarprodukt] Das *Standardskalarprodukt* zweier Vektoren $x, y \in$ 8.1.1
$\mathbb{R}^n$ wurde definiert als

$$\langle x, y\rangle = \langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \rangle := x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^\top y.$$

Damit ist das Skalarprodukt zweier Vektoren in $\mathbb{R}^n$ eine reelle Zahl. Der Vektorraum $\mathbb{R}^n$ ausgestattet mit diesem Skalarprodukt heißt *euklidischer Vektorraum* $\mathbb{R}^n$.

Definition 7.1.2. [Länge und Orthogonalität] Die *Länge* eines Vektors $x \in \mathbb{R}^n$ wurde definiert 8.1.2
als

$$\|x\| := \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Jeder Vektor der Länge 1 heißt *Einheitsvektor*. Jeder nichtverschwindende Vektor x kann *normiert* werden: $\frac{1}{\|x\|}x$ ist ein Einheitsvektor.

Zwei Vektoren $x, y \in \mathbb{R}^n$ heißen *orthogonal*, i.Z. $x \perp y$, wenn $\langle x, y\rangle = 0$ gilt. Den Winkel zwischen zwei Vektoren im $\mathbb{R}^n$ werden wir später definieren.

Aus der Definition des Skalarproduktes kann man direkt das Folgende ableiten:

7.1.3. [Regeln für das Skalarprodukt] Das Skalarprodukt ist bilinear ($\mathrm{B}_{1,2}$), symmetrisch (S) und 8.1.4
positiv definit (P), d.h. für alle $\alpha \in \mathbb{R}$ und alle $x, x', y, y' \in \mathbb{R}^n$ gilt:

$$\begin{array}{llcllcl}
(\mathrm{B}_1) & \langle x+x',y\rangle & = & \langle x,y\rangle+\langle x',y\rangle, & \langle \alpha x,y\rangle & = & \alpha\langle x,y\rangle. \\
(\mathrm{B}_2) & \langle x,y+y'\rangle & = & \langle x,y\rangle+\langle x,y'\rangle, & \langle x,\alpha y\rangle & = & \alpha\langle x,y\rangle. \\
(\mathrm{S}) & \langle x,y\rangle & = & \langle y,x\rangle. & & & \\
(\mathrm{P}) & \langle x,x\rangle & \geq & 0, & \langle x,x\rangle = 0 & \iff & x=0.
\end{array}$$

7.2 $\mathbb{C}^n$ als unitärer Vektorraum

sec8.2

Wir möchten ein Skalarprodukt auf $\mathbb{C}^n$ definieren, das uns dieselben Begriffe von Länge und Orthogonalität im komplexen Fall zu erklären erlaubt. 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

$$|z| = |x+iy| = \sqrt{x^2+y^2} = \sqrt{(x+iy)(x-iy)} = \sqrt{z\overline{z}}.$$

Wir führen daher die folgende Schreibweise ein:

$$\text{Für jeden Vektor } z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^n \text{ schreiben wir } \overline{z} = \begin{pmatrix} \overline{z_1} \\ \overline{z_2} \\ \vdots \\ \overline{z_n} \end{pmatrix}$$

für den *komplex konjugierten Vektor*.

8.2.1 **Definition 7.2.1.** [Das Standardskalarprodukt auf $\mathbb{C}^n$] Das *Standard-(Hermitesche) Skalarprodukt* zweier Vektoren $z, w \in \mathbb{C}^n$ wird definiert als

$$\langle z,w\rangle = \langle \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \rangle := z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n} = z^\top \overline{w}.$$

Damit ist das Skalarprodukt zweier Vektoren in $\mathbb{C}^n$ eine komplexe Zahl. Der Vektorraum $\mathbb{C}^n$, ausgestattet mit diesem Skalarprodukt, heißt *unitärer* Vektorraum $\mathbb{C}^n$.

8.2.2 **Definition 7.2.2.** [Länge und Orthogonalität] Die *Länge* oder *hermitesche Norm* eines Vektors $z \in \mathbb{C}^n$ ist definiert durch

$$\begin{aligned} \|z\| &:= \sqrt{\langle z,z\rangle} = \sqrt{z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n}} \\ &= \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}. \end{aligned}$$

Jeder Vektor der Länge eins heißt *Einheitsvektor*. Jeder nichtverschwindende Vektor z kann *normiert* werden: $\frac{1}{\|z\|}z$ ist ein Einheitsvektor.

Zwei Vektoren $z, w \in \mathbb{C}^n$ heißen *orthogonal*, i.Z. $z \perp w$, wenn $\langle z,w\rangle = 0$ gilt.

Aus der Definition des Skalarproduktes kann man direkt das Folgende ableiten:

8.2.4 **7.2.3.** [Regeln für das Skalarprodukt] Das Standardskalarprodukt auf $\mathbb{C}^n$ ist semi-bilinear (B_1, B_2'), hermitesch (H) und positiv definit (P), d.h. für alle $\alpha \in \mathbb{C}$ und alle z, z', w, $w' \in \mathbb{C}^n$ gilt:

$$\begin{array}{llcllcl}
(\mathrm{B}_1) & \langle z+z',w\rangle & = & \langle z,w\rangle+\langle z',w\rangle, & \langle \alpha z,w\rangle & = & \alpha\langle z,w\rangle. \\
(\mathrm{B}_2') & \langle z,w+w'\rangle & = & \langle z,w\rangle+\langle z,w'\rangle, & \langle z,\alpha w\rangle & = & \overline{\alpha}\langle z,w\rangle. \\
(\mathrm{H}) & \langle z,w\rangle & = & \overline{\langle w,z\rangle}. & & & \\
(\mathrm{P}) & \langle z,z\rangle & \geq & 0, & \langle z,z\rangle = 0 & \iff & z=0.
\end{array}$$

Beachte, dass dieses Skalarprodukt wegen der zweiten Gleichung in (B_2') nicht bilinear ist; beschränken wir uns auf reelle Zahlen α, sind diese Eigenschaften genau die Eigenschaften ($\mathrm{B}_{1,2}$), (S), (P) aus 7.1.3.

Die Eigenschaft B_2' ist redundant. Sie folgt direkt aus B_1 und H (Nachweis!).

7.3 Euklidische und unitäre Vektorräume sec8.3

In diesem Abschnitt verallgemeinern wir das Konzept des euklidischen bzw. unitären Raums von $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ auf allgemeinere Vektorräume.

Definition 7.3.1. [Euklidische und unitäre Vektorräume] Ein *euklidischer* Vektorraum ist ein 8.3.1
reeller Vektorraum V ausgestattet mit einer symmetrischen, positiv definiten Bilinearform, d.h. einer Abbildung

$$(v, w) \mapsto \langle v, w\rangle \;:\; V \times V \to \mathbb{R},$$

die die Eigenschaften ($\mathrm{B}_{1,2}$), (S) und (P) aus 7.1.3 erfüllt.

Ein *unitärer* Vektorraum ist ein komplexer Vektorraum V mit einer hermiteschen, positiv definiten Sesquilinearform, d.h. einer Abbildung

$$(v, w) \mapsto \langle v, w\rangle \;:\; V \times V \to \mathbb{C},$$

die die Eigenschaften (B_1), (B_2'), (H) und (P) aus 7.2.3 erfüllt.

Beispiel 7.3.2. Das Standardskalarprodukt definiert auf $\mathbb{R}^n$ die Struktur eines euklidischen Raums.

Das hermitesche Standardskalarprodukt definiert auf $\mathbb{C}^n$ die Struktur eines unitären Raums.

Im folgenden werden wir, sofern nichts anderes gesagt wird, auf dem $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ jeweils nur das Standardskalarprodukt betrachten.

Bemerkung 7.3.3. 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 Standardskalarprodukt 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 7.3.4. 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

$$\mathbb{C} \times V \to V, \quad (\lambda, v) \mapsto \lambda v$$

zu einer Abbildung

$$\mathbb{R} \times V \to V, \quad (\lambda, v) \mapsto \lambda v$$

ein. Weiter ist

$$(v, w) := \mathrm{Re}\langle v, w\rangle = \frac{1}{2}(\langle v, w\rangle + \langle w, v\rangle) = \frac{1}{2}(\langle v, w\rangle + \overline{\langle v, w\rangle})$$

ein reellwertiges Skalarprodukt mit dem $V_{\mathbb{R}}$ ein euklidischer Raum wird. (Nachweis als Übung!)

Definition 7.3.5. Für Elemente v, w eines beliebigen euklidischen oder unitären Vektorraumes definieren wir die *Länge* (oder die *Norm*) und *Orthogonalität* genau wie in 7.1.2 und 7.2.2:

$$\|v\| := \sqrt{\langle v, v\rangle} \quad , \quad v \perp w \quad :\Longleftrightarrow \quad \langle v, w\rangle = 0$$

Vektoren der Länge 1 heißen *Einheitsvektoren*.

Im Folgenden sei V ein euklidischer oder unitärer Raum. Für die Beweise der folgenden Aussagen müssen wir nicht zwischen beiden Fällen unterscheiden. Handelt es sich um einen euklidischen Raum, so gilt natürlich auch $\langle v, \alpha w\rangle = \overline{\alpha}\langle v, w\rangle$ für alle $\alpha \in \mathbb{R}$, da $\overline{\alpha} = \alpha$ ist.

Bemerkung 7.3.6. 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. 8.3.6a

In einem euklidischen Vektorraum haben wir die *Polarisierungsidentität*:

$$\langle x, y\rangle = \frac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2\big)$$

und in einem unitären Vektorraum gilt die etwas kompliziertere *Polarisierungsidentität*:

$$\langle x, y\rangle = \frac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\big).$$

Beide Identitäten verifiziert man leicht durch Nachrechnen.

8.2.5 **Satz 7.3.7.** [Satz von Pythagoras] *Aus $x \perp y$ folgt*

$$\boxed{\|x+y\|^2 = \|x\|^2 + \|y\|^2.}$$

Beweis. Sei $x \perp y$. Definitionsgemäß gilt $\langle x, y\rangle = 0$. Aus (H) folgt dann auch $\langle y, x\rangle = \overline{\langle x, y\rangle} = 0$. Unter Anwendung von (B) folgern wir

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2.$$

□

8.2.6 **Satz 7.3.8.** [Die Cauchy-Schwarzsche Ungleichung] *Für alle Vektoren $x, y \in V$ gilt*

$$\boxed{|\langle x, y\rangle| \leq \|x\| \cdot \|y\|.}$$

Beweis. Es genügt, das Folgende zu zeigen:

$$|\langle x, y\rangle|^2 \leq \|x\|^2\|y\|^2 \quad \text{d.h.} \quad \langle x, y\rangle\overline{\langle x, y\rangle} = \langle x, y\rangle\langle y, x\rangle \leq \langle x, x\rangle\langle y, y\rangle.$$

(Wurzelziehen auf beiden Seiten ergibt die gewünschte Ungleichung.) Im Fall $y = 0$ verschwinden beide Seiten und die Ungleichung gilt trivialerweise. Betrachten wir den Fall $y \neq 0$. In diesem Fall gilt $\langle y, y\rangle > 0$ nach (P) und damit können wir die komplexe Zahl $\gamma := \frac{\langle x,y\rangle}{\langle y,y\rangle}$ definieren. Es folgt

$$\begin{aligned}
0 &\leq \langle x-\gamma y, x-\gamma y\rangle \quad \text{nach (P)}\\
&= \langle x, x\rangle - \langle x, \gamma y\rangle - \langle \gamma y, x\rangle + \langle \gamma y, \gamma y\rangle \quad \text{nach (B)}\\
&= \langle x, x\rangle - \overline{\gamma}\langle x, y\rangle - \gamma\langle y, x\rangle + \overline{\gamma}\gamma\langle y, y\rangle \quad \text{nach (B)}\\
&= \langle x, x\rangle - \overline{\gamma}\langle x, y\rangle - \gamma\langle y, x\rangle + \overline{\gamma}\langle x, y\rangle \quad \text{nach der Definition von } \gamma\\
&= \langle x, x\rangle - \gamma\langle y, x\rangle.
\end{aligned}$$

Daraus folgern wir $\gamma\langle y, x\rangle \leq \langle x, x\rangle$. Ersetzen wir γ durch dessen Wert, so folgt $\frac{\langle x,y\rangle}{\langle y,y\rangle}\langle y, x\rangle \leq \langle x, x\rangle$. Daraus folgt direkt $\langle x, y\rangle\langle y, x\rangle \leq \langle x, x\rangle\langle y, y\rangle$. □

Definition 7.3.9. Sei V ein euklidischer oder unitärer Raum. Der *Winkel* $\gamma \in [0, \pi[$ zwischen zwei nichtverschwindenen Vektoren v und w wird durch

$$\cos\gamma = \frac{\mathrm{Re}\langle v, w\rangle}{\|v\| \cdot \|w\|}$$

definiert. Nach der Cauchy-Schwarzschen Ungleichung wissen wir, dass $-1 \leq \frac{\mathrm{Re}\langle v,w\rangle}{\|v\|\cdot\|w\|} \leq 1$ gilt; es existiert also ein eindeutiges $\gamma \in [0, \pi[$ mit $\cos\gamma = \frac{\mathrm{Re}\langle v,w\rangle}{\|v\|\cdot\|w\|}$.

Da $\mathbb{C}^n$ mit dem Standardskalarprodukt ein unitärer Raum ist (7.2.3), erhalten wir durch Spezialisieren der allgemeinen Cauchy-Schwarzschen Ungleichung:

Folgerung 7.3.10. *Es gilt für alle Vektoren $z, w \in \mathbb{C}^n$:*

$$|z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}| \leq \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}\ \sqrt{|w_1|^2 + |w_2|^2 + \cdots + |w_n|^2}.$$

8.2.7 **Folgerung 7.3.11.** [Die Dreiecksungleichung] *Für $x, y \in V$ gilt*

$$\boxed{\|x + y\| \leq \|x\| + \|y\|.}$$

Beweis.

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle && \text{nach (B)} \\
&\leq \|x\|^2 + 2\|x\|\ \|y\| + \|y\|^2 && \text{nach der Cauchy-Schwarzschen Ungleichung} \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

Wurzelziehen auf beiden Seiten ergibt die Dreiecksungleichung. □

Aufgabe 7.3.12.

a) Zeige, dass in der Dreiecksungleichung Gleichheit genau dann gilt, wenn $x = \alpha y$ für eine reelle Zahl $\alpha > 0$ ist.

b) Zeige, dass die Cauchy-Schwarzsche Ungleichung genau dann mit Gleichheit erfüllt ist, wenn x und y linear abhängig sind.

Aufgabe 7.3.13. 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) $d(x, y) = d(y, x)$.

(M2) $d(x, y) \geq 0$ mit Gleichheit nur für $x = y$.

(M3) $d(x, z) \leq d(x, y) + d(y, z)$.

Aufgabe 7.3.14. 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 7.3.15. * Sei V ein euklidischer Raum und $\varphi\colon V \to V$ 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.

7.4 Orthonormalbasen

Definition 7.4.1. [Orthonormalbasen] Wir nennen ein System $v_1, v_2, \ldots, v_m$ 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: 8.3.2

$$\langle v_i, v_j \rangle = 0, \quad \text{falls} \quad i \neq j.$$

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 $v_1, v_2, \ldots, v_n$, das eine Basis von V ist, heißt *Orthonormalbasis* (ONB).

Lemma 7.4.2. *Jedes Orthogonalsystem $v_1, v_2, \ldots, v_m$ 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 i:

$$0 = \langle \sum_{j=1}^m \lambda_j v_j, v_i \rangle = \sum_{j=1}^m \lambda_j \langle v_j, v_i \rangle = \lambda_i \langle v_i, v_i \rangle$$

Aus $\langle v_i, v_i \rangle \neq 0$ folgern wir $\lambda_i = 0$ für jedes i. □

Beispiel 7.4.3.

1. Ist $\dim V = n$, so ist ein Orthonormalsystem aus n Vektoren eine Orthonormalbasis. Die Standardbasis in $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ ist eine Orthonormalbasis bzgl. des Standardskalarprodukts.

2. Im $\mathbb{R}^2$ bilden für jedes $\gamma \in \mathbb{R}$ die Vektoren

$$v_1 = \begin{pmatrix} \cos\gamma \\ \sin\gamma \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -\sin\gamma \\ \cos\gamma \end{pmatrix}$$

ebenso eine Orthonormalbasis.

8.3.3 **Bemerkung 7.4.4.** [Koordinaten bzgl. einer Orthonormalbasis] Sei $v_1, v_2, \ldots, v_n$ eine Orthonormalbasis von V. Dann hat ein beliebiger Vektor $v \in V$ eine Darstellung als Linearkombination

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = \sum_{j=1}^m \lambda_j v_j$$

Bildung des Skalarproduktes mit v_i auf beiden Seiten ergibt

$$\langle v, v_i \rangle = \langle \sum_{j=1}^m \lambda_j v_j, v_i \rangle = \sum_{j=1}^m \lambda_j \langle v_j, v_i \rangle = \sum_{j=1}^m \lambda_j \delta_{ij} = \lambda_i.$$

Also ist die i-te Koordinate von v bzgl. der Orthonormalbasis $v_1, v_2, \ldots, v_n$ gleich $\lambda_i = \langle v, v_i \rangle$ und die Darstellung wird zu

$$\boxed{v = \langle v, v_1 \rangle v_1 + \langle v, v_2 \rangle v_2 + \cdots + \langle v, v_n \rangle v_n.}$$

Beispiel 7.4.5. 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 $v = \langle v_1, v\rangle v_1 + \langle v_2, v\rangle v_2 + \langle v_3, v\rangle v_3 = \frac{4}{3}v_1 - \frac{10}{3}v_2 - \frac{10}{3}v_3$.

7.4.6. [Das Gram-Schmidtsche Orthogonalisierungsverfahren] Wir betrachten mit $b_1, b_2, \ldots, b_m$ 8.3.5
ein System von m linear unabhängigen Vektoren in einem euklidischen oder unitären Vektorraum V. Diese spannen einen m-dimensionalen linearen Teilraum U von V auf. Wie wir unten zeigen werden, ergibt die folgende Prozedur ein Orthogonalsystem $u_1, u_2, \ldots, u_m$ und nach Normierung eine Orthonormalbasis $v_1, v_2, \ldots, v_m$ 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:

$$\begin{aligned}
u_1 &:= b_1,\\
u_2 &:= b_2 - \frac{\langle b_2, u_1\rangle}{\langle u_1, u_1\rangle}u_1,\\
u_3 &:= b_3 - \frac{\langle b_3, u_1\rangle}{\langle u_1, u_1\rangle}u_1 - \frac{\langle b_3, u_2\rangle}{\langle u_2, u_2\rangle}u_2,\\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots\\
u_m &:= b_m - \sum_{i=1}^{m-1}\frac{\langle b_m, u_i\rangle}{\langle u_i, u_i\rangle}u_i.
\end{aligned}$$

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. $\operatorname{span}(v_1, \ldots, v_{m-1}, v_m) = \operatorname{span}(b_1, \ldots, b_{m-1}, b_m)$
3. $v_1, \ldots, v_{m-1}, v_m$ 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

$$b_m = \sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i \in \operatorname{span}(v_1, \dots, v_{m-1}) = \operatorname{span}(b_1, \dots, b_{m-1})$$

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 $\operatorname{span}(v_1, \dots, v_{m-1}) = \operatorname{span}(b_1, \dots, b_{m-1})$ ist; weiter gilt

$$v_m := \frac{u_m}{\|u_m\|} = \frac{1}{\|u_m\|}(b_m - \underbrace{\sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i}_{\in \operatorname{span}(b_1,\dots,b_{m-1})}) \in \operatorname{span}(b_1, \dots, b_{m-1}, b_m)$$

und

$$b_m = \|u_m\| v_m + \sum_{i=1}^{m-1} \langle b_m, v_i \rangle v_i \in \operatorname{span}(v_1, \dots, v_{m-1}, v_m).$$

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:

$$\begin{aligned}
\langle v_i, v_m \rangle = \langle v_i, \frac{u_m}{\|u_m\|} \rangle &= \frac{1}{\|u_m\|} \langle v_i, b_m - \sum_{j=1}^{m-1} \langle b_m, v_j \rangle v_j \rangle \\
&= \frac{1}{\|u_m\|} \left(\langle v_i, b_m \rangle - \sum_{j=1}^{m-1} \langle v_j, b_m \rangle \underbrace{\langle v_i, v_j \rangle}_{\delta_{ij}} \right) \\
&= \frac{1}{\|u_m\|} (\langle v_i, b_m \rangle - \langle v_i, b_m \rangle) = 0.
\end{aligned}$$

□

Bemerkung 7.4.7. 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 Schnritte 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

$$b_{k+1} = \langle b_{k+1}, v_1 \rangle v_1 + \ldots + \langle b_{k+1}, v_k \rangle v_k,$$

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:

8.3.6

Satz 7.4.8. *Jeder endlichdimensionale lineare Teilraum eines euklidischen oder unitären Vektorraumes hat eine Orthonormalbasis.*

8.3.7

Beispiel 7.4.9. 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

$$b_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \quad , \quad b_2 = \begin{pmatrix} 3 \\ 1 \\ -1 \\ -3 \end{pmatrix} \quad , \quad b_3 = \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

Wir benutzen nun das Gram-Schmidtsche Orthogonalisierungsverfahren, um eine orthonormale Basis von U zu finden:

$$u_1 = b_1, \quad \|u_1\| = 2, \quad v_1 = \frac{u_1}{\|u_1\|} = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix},$$

$$u_2 = b_2 - \langle b_2, v_1\rangle v_1 = b_2 - 2v_1 = \begin{pmatrix} 2 \\ 2 \\ -2 \\ -2 \end{pmatrix}, \quad \|u_2\| = 4, \quad v_2 = \frac{u_2}{\|u_2\|} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

$$u_3 = b_3 - \langle b_3, v_1\rangle v_1 - \langle b_3, v_2\rangle v_2 = b_3 - 2v_1 - 2v_2 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \qquad \|u_3\| = 2,$$

$$v_3 = \frac{u_3}{\|u_3\|} = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

Definition 7.4.10. Eine lineare Abbildung $\varphi\colon V \to W$ zwischen euklidischen (unitären) Räumen heißt *Isometrie*, wenn

$$\langle \varphi(x), \varphi(y)\rangle = \langle x, y\rangle$$

für alle $x, y \in V$ gilt.

Aufgabe 7.4.11. Zeige: Für eine lineare Abbildung $\varphi\colon V \to W$ sind äquivalent:

1. φ ist eine Isometrie, d.h. $\langle \varphi(x), \varphi(y)\rangle = \langle x, y\rangle$ für alle $x, y \in V$.
2. $\|\varphi(x)\| = \|x\|$ für alle $x \in V$.
3. $\|\varphi(x) - \varphi(y)\| = \|x - y\|$ für alle $x, y \in V$.

Hinweis: Die Polarisierungsidentität 7.3.6.

Satz 7.4.12. *Sei V ein euklidscher oder unitärer $\mathbb{K}$-Vektorraum der Dimension n. Dann ist V isometrisch isomorph zu $\mathbb{K}^n$ mit dem Standardskalarprodukt, d.h. es existiert eine bijektive Isometrie $\kappa\colon V \to \mathbb{K}^n$.*

Beweis. Wir haben oben gesehen, dass V eine Orthonormalbasis $B = \{v_1, v_2, \dots, v_n\}$ besitzt. Wir berechnen das Skalarprodukt zweier Vektoren v und w, die durch die Koordinaten

$$v = \sum_{i=1}^n \alpha_i v_i, \qquad w = \sum_{j=1}^n \beta_j v_j$$

bzgl. der Basis B dargestellt werden:

$$\begin{aligned}\langle v, w\rangle &= \langle \sum_{i=1}^n \alpha_i v_i, \sum_{j=1}^n \beta_j v_j\rangle = \sum_{i=1}^n \alpha_i \cdot \langle v_i, \sum_{j=1}^n \beta_j v_j\rangle \\ &= \sum_{i=1}^n \alpha_i \cdot \sum_{j=1}^n \overline{\beta_j}\langle v_i, v_j\rangle = \sum_{i,j=1}^n \alpha_i \overline{\beta_j}\delta_{ij} = \sum_{i=1}^n \alpha_i \overline{\beta_i}. \qquad .\end{aligned}$$

Wir stellen fest, dass wir denselben Ausdruck für das Standardskalarprodukt auf $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ erhalten.

Wir betrachten jetzt den Vektorraumisomorphismus

$$\kappa_B\colon V \to \mathbb{K}^n, \quad v \mapsto \begin{pmatrix}\langle v, v_1\rangle \\ \langle v, v_2\rangle \\ \vdots \\ \langle v, v_n\rangle\end{pmatrix},$$

der jeden Vektor $v \in V$ auf seinen Koordinatenvektor bzgl. der Orthonormalbasis B abbildet. Diese Abbildung erhält das Skalarprodukt, wie wir oben gesehen haben, d.h. für alle $v, w \in V$ gilt

$$\langle \kappa_B(v), \kappa_B(w)\rangle = \langle v, w\rangle.$$

Also ist κ_B eine Isometrie. □

8.3.8 **Satz 7.4.13.** [Die Q-R-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 ist.

Beweis. Wegen $\operatorname{rank}(B) = m$ 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

$$(\sum_{j=1}^{m} q_{ij} r_{jk})_{i=1,\dots,n}$$

besitzt. Da q_{ij} die i-te Komponente der j-ten Spalte v_j on Q ist, ist die k-te Spalte von QR gegeben durch

$$\begin{aligned} \sum_{j=1}^{m} r_{jk} v_j = \sum_{j=1}^{k} r_{jk} v_j &= r_{1k} v_1 + \dots + r_{k-1,k} v_{k-1} + r_{kk} v_k \\ &= \langle b_k, v_1\rangle v_1 + \dots + \langle b_k, v_{k-1}\rangle v_{k-1} + \|u_k\| v_k = b_k. \end{aligned}$$

Also ist $B = QR$. □

Beispiel 7.4.14. Wir betrachten die Matrix B, wobei die Spalten b_1, b_2, b_3 die Vektoren aus Beispiel 7.4.9 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 7.4.9 erhalten wir

$$\underbrace{\begin{pmatrix} 1 & 3 & 3 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -3 & -1 \end{pmatrix}}_{B} = \underbrace{\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \end{pmatrix}}_{Q} \cdot \underbrace{\begin{pmatrix} 2 & 2 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 2 \end{pmatrix}}_{R}$$

Definition 7.4.15. [Orthogonale lineare Teilräume] Sei V ein euklidischer oder unitärer Vektorraum. Für jede Teilmenge $M \subseteq V$ definieren wir den *Orthogonalraum* von M durch 8.3.9

$$M^{\perp} := \{v \in V : (\forall a \in M)\langle a, v\rangle = 0\}.$$

Lemma 7.4.16. *$M^\perp$ 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

$$\langle a, v + w\rangle \underset{\text{(B2)}}{=} \langle a, v\rangle + \langle a, w\rangle = 0$$

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 $\langle a, \lambda v\rangle = \overline{\lambda}\langle a, v\rangle = 0$ und damit $\lambda v \in M^\perp$.

□

Aufgabe 7.4.17. Die Zuordnung $M \mapsto M^\perp$ hat 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. $A^\perp = ((A^\perp)^\perp)^\perp$.

Bemerkung 7.4.18. [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

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \qquad \text{und} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

kann die Gleichung $\langle a, v\rangle = 0$ geschrieben werden als

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

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

$$\langle a, v\rangle = 0, \qquad \langle b, v\rangle = 0, \qquad \langle c, v\rangle = 0, \qquad \dots$$

von homogenen linearen Gleichungen.

8.3.25

Lemma 7.4.19. *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.

(i) Offensichtlich ist $U^\perp \subseteq \{a_1, a_2, \dots, a_m\}^\perp$.

(ii) Für die umgekehrte Inklusion betrachten wir ein $v \in \{a_1, a_2, \dots, a_m\}^\perp$. Dann ist $\langle a_1, v\rangle = 0$, $\langle a_2, v\rangle = 0, \dots, \langle a_m, v\rangle = 0$. Sei a ein Element von U. Da der Teilraum U durch $a_1, a_2, \dots,$ a_m aufgespannt wird, kann a als Linearkombination

$$a = \lambda_1a_1 + \lambda_2a_2 + \cdots + \lambda_ma_m$$

dargestellt werden. Also folgt

$$\begin{aligned}\langle a, v\rangle &= \langle \lambda_1a_1 + \lambda_2a_2 + \cdots + \lambda_ma_m, v\rangle \\ &= \lambda_1\langle a_1, v\rangle + \lambda_2\langle a_2, v\rangle + \cdots + \lambda_m\langle a_m, v\rangle = 0.\end{aligned}$$

Damit ergibt sich $\langle a, v\rangle = 0$ für jedes $a \in U$, also $v \in U^\perp$.

□

Satz 7.4.20. [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 $\pi : V \longrightarrow V$ durch* 8.3.10

$$\boxed{\pi(v) = \textstyle\sum_{i=1}^m \langle v, v_i \rangle v_i}$$

für alle $v \in V$.

Dann gelten die folgenden Eigenschaften:

(1) *π ist eine lineare Abbildung.*

(2) *$\pi(v) \in U$ für alle $v \in V$.*

(3) *$\pi(u) = u$ für alle $u \in U$.*

(4) *$\pi \circ \pi = \pi$.*

(5) *$\operatorname{im} \pi = U$ und $\ker \pi = U^\perp$.*

(6) *$v - \pi(v) \in U^\perp$ für alle $v \in V$.*

(7) *$\|v - \pi(v)\| \leq \|v - u\|$ 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 (7) folgt insbesondere, dass die Abbildung π nicht von der Wahl der ONB in U abhängt.

Beweis.

(1) Für jedes $w \in V$ ist die Abbildung

$$w^* \colon V \to \mathbb{K}, \quad v \mapsto \langle v, w \rangle$$

linear. Hieraus schliesst man leicht, dass auch $\pi = \sum_{i=1}^m v_i^* \cdot v_i$ linear ist (Nachweis!).

(2) In der Tat ist $\pi(v)$ eine Linearkombination der Basisvektoren von U.

(3) Ist $u = \displaystyle\sum_{j=1}^m \lambda_j v_j$ ein Element von U, so folgt

$$\pi(u) = \sum_{i=1}^m \langle \sum_{j=1}^m \lambda_j v_j, v_i \rangle v_i = \sum_{i=1}^m \sum_{j=1}^m \lambda_j \underbrace{\langle v_j, v_i \rangle}_{\delta_{ij}} v_i = \sum_{i=1}^m \lambda_i v_i = u.$$

(4) Nach (2) ist $\pi(v) \in U$ und damit $\pi(\pi(v)) = \pi(v)$ wegen (3).

(5) Die Behauptung für das Bild folgt aus (2) und (3). Man beachte für den Kern:

$$\begin{aligned} \pi(v) = 0 &\iff \langle v, v_i \rangle = 0 \qquad \text{für alle } i = 1, \dots, m \\ &\iff v \in \{v_1, \dots, v_m\}^\perp \underset{(7.4.19)}{=} U^\perp. \end{aligned}$$

(6) Für $v \in V$ ergibt sich

$$\pi(v - \pi(v)) \underset{(1)}{=} \pi(v) - \pi(\pi(v)) \underset{(4)}{=} \pi(v) - \pi(v) = 0$$

und damit $v - \pi(v) \in \ker \pi \underset{(5)}{=} U^\perp$.

(7) 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 (6). 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 $u = \pi(v)$. □

Aufgabe 7.4.21. Zeige, dass $\pi' : V \to V$ definiert durch $\pi'(v) = v - \pi(v)$ die orthogonale Projektion auf $U^\perp$ ist. Beachte, dass $\pi' = \mathrm{id}_V - \pi$ gilt. Vergleiche mit dem Gram-Schmidt-Verfahren.

8.3.11 **Proposition 7.4.22.** *Für jeden endlichdimensionalen linearen Teilraum U von V gilt*

$$V = U \oplus U^\perp.$$

Insbesondere gilt für endlichdimensionales V:

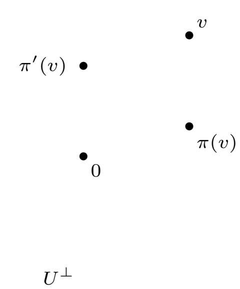
$$\dim U^\perp = \dim V - \dim U.$$

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 $v \in V$ die Aussage $v = \pi(v) + (v - \pi(v)) \in U + U^\perp$ gilt, wobei π die Orthogonalprojektion auf U ist.

□



Folgerung 7.4.23. *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 $\pi\colon V \to U$. 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

$$(U^\perp)^\perp = U + \big((U^\perp)^\perp \cap U^\perp\big) = U.$$

□

Bemerkung 7.4.24. Sei V ein endlichdimensionaler euklidischer Raum und $H \subseteq V$ eine Hyperebene. Dann ist $\dim H^\top = \dim V - \dim H = 1$, also $H^\top = \mathbb{R}n_H$ mit einem Einheitsvektor $n_H \in V$, den wir einen *Normalenvektor von H* nennen. Damit erhalten wir die Darstellung

$$H = (H^\perp)^\perp = (\mathbb{R}n_H)^\perp = n_H^\perp = \{v \in V\colon \langle v, n_H\rangle = 0\}.$$

Eine affine Hyperebene der Gestalt $p + H$ lässt sich daher auch beschreiben als

$$p + H = \{v \in V\colon \langle v, n_H\rangle = \langle p, n_H\rangle\}.$$

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.

7.4.25. [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 ein 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 $\|\pi_i(s)\|^2$ geschieht, wobei $\pi_i : \mathbb{C}^2 \to U_i$ die Projektion auf U_i ist. Nach der Polarisierung ist der Zustand des Partikels repräsentiert durch $s' := \pi_i(s)/||\pi_i(s)||$, wobei i die Nummer des Teilraums ist, zu dem die Polarisierung geführt hat. Man erinnere sich, dass $i = 1$ oder $i = 2$ i.A. beide möglich sind; ein Partikel i tritt mit der Wahrscheinlichkeit $||\pi_i(s)||^2$ 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 Unterräumen 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 $V = U_1 \oplus \cdots \oplus U_l$ beschrieben. Ist der Ausgangszustand durch den Einheitsvektor s gegeben und

$$s = s_1 + \ldots + s_l$$

die Zerlegung von s in Komponenten $s_j \in U_j$, so folgt aus wiederholtem Anwenden des Satzes von Pythagoras die Beziehung

$$1 = \|s\|^2 = \|s_1\|^2 + \ldots + \|s_l\|^2.$$

Die Wahrscheinlichkeit, nach dem Experiment den Endzustand i zu messen, ist nun $\|s_i\|^2$. Führt man eine Messung durch, ob sich das System im Zustand i befindet, so ist der neue Zustand repräsentiert durch den Einheitsvektor $\frac{1}{\|s_i\|}s_i$, wenn $s_i \neq 0$ gilt. Ist $s_i = 0$, so kann sich das System gar nicht im Zustand i 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.

8.3.12 **Beispiel 7.4.26.** [Der Hilbertraum ℓ^2] Sei $\ell^2 := \ell^2(\mathbb{N}, \mathbb{C})$ die Menge aller Folgen

$$x = (x_i)_{i=1,2,\ldots} = (x_1, x_2, x_3, \ldots)$$

von komplexen Zahlen mit

$$\sum_{i=1}^{\infty} |x_i|^2 < \infty.$$

Wir definieren für zwei Folgen $x = (x_i)$ und $y = (y_i)$ in ℓ^2

$$x + y := (x_i + y_i) \qquad \text{und} \qquad \lambda x = (\lambda x_i) \qquad \text{für } \lambda \in \mathbb{C}.$$

Aufgabe 7.4.27. Zeige $\sum_{i=1}^{\infty} |x_i + y_i|^2 < +\infty$, $\sum_{i=1}^{\infty} |\lambda x_i|^2 < \infty$, und dass $\sum_{i=1}^{\infty} x_i \overline{y_i}$ absolut konvergent ist.

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

$$\langle x, y \rangle := \sum_{i=1}^{\infty} x_i \overline{y_i}$$

(Da die unendliche Reihe absolut konvergiert, beschreibt die Reihe eine komplexe Zahl.) Man zeigt leicht, dass die Eigenschaften (B_1), (B_2'), (S) und (H) eines Skalarproduktes erfüllt sind. Damit wird ℓ^2 zu einem unitären Vektorraum. Die Norm von $x = (x_i) \in \ell^2$ ist $\|x\| = \sqrt{\sum_{i=1}^{\infty} |x_i|^2}$. Die Vektoren

$$\begin{aligned}
e_1 &= (1, 0, 0, 0, \ldots) \\
e_2 &= (0, 1, 0, 0, \ldots) \\
e_3 &= (0, 0, 1, 0, \ldots) \\
&\vdots
\end{aligned}$$

bilden ein unendliches Orthonormalsystem für ℓ^2.

Beispiel 7.4.28. [Der Vektorraum $C([0,1]) := C([0,1],\mathbb{R})$ der stetigen Funtionen $[0,1] \to \mathbb{R}$ mit der L^2-Norm] Sei $C([0,1])$ die Menge aller stetigen Funktionen $f : [0,1] \to \mathbb{R}$. Bzgl. der gewöhnlichen Addition und der skalaren Multiplikation 8.3.13

$$(f+g)(x) := f(x) + g(x), \qquad (\lambda f)(x) := \lambda \cdot f(x), \quad \lambda \in \mathbb{R},$$

wird $C([0,1])$ zu einem reellen Vektorraum. Wir definieren für $f, g \in C([0,1])$ ein Skalarprodukt

$$\langle f, g\rangle := \int_0^1 f(x)g(x)\,dx.$$

Aufgabe 7.4.29. Die Eigenschaften $(B_{1,2}), (S), (P)$ eines Skalarproduktes sind erfüllt.

Ausgestattet mit diesem Skalarprodukt wird $C([0,1])$ ein euklidischer Vektorraum. Die euklidische Norm einer Funktion f ist

$$\|f\|_2 = \sqrt{\int_0^1 f(x)^2 dx} = \left(\int_0^1 f(x)^2 dx\right)^{1/2}.$$

Wie in jedem euklidischen Vektorraum gilt die Cauchy-Schwarzsche Ungleichung 7.3.8:

$$|\langle f, g\rangle| \le \|f\|_2 \cdot \|g\|_2 \qquad \text{d.h.}$$

$$\left|\int_0^1 f(x)g(x)dx\right| \le \left(\int_0^1 f(x)^2 dx\right)^{1/2} \left(\int_0^1 g(x)^2 dx\right)^{1/2}.$$

Ebenso gilt die Dreiecksungleichung 7.3.11:

$$\|f+g\|_2 \le \|f\|_2 + \|g\|_2 \qquad \text{d.h.}$$

$$\left(\int_0^1 (f(x)+g(x))^2 dx\right)^{1/2} \le \left(\int_0^1 f(x)^2 dx\right)^{1/2} + \left(\int_0^1 g(x)^2 dx\right)^{1/2}.$$

Beispiele von orthogonalen Funktionen in $C([0,1])$ erhält man aus den folgenden *Orthogonalitätsrelationen* für Sinus- und Kosinusfunktion:

$$\int_0^1 \sin(2\pi mx)\sin(2\pi nx)\ dx = \begin{cases} 0 & \text{für } n \ne m \\ \frac{1}{2} & \text{für } n = m > 0, \end{cases}$$

$$\int_0^1 \sin(2\pi mx)\cos(2\pi nx)\ dx = \quad 0 \quad \text{für alle } n, m,$$

$$\int_0^1 \cos(2\pi mx)\cos(2\pi nx)\ dx = \begin{cases} 0 & \text{für } n \ne m \\ \frac{1}{2} & \text{für } n = m > 0 \\ 1 & \text{für } n = m = 0. \end{cases}$$

Betrachten wir die Funktionen $v_0, v_1, v_2, \ldots, w_1, w_2, \ldots$, die für $x \in [0,1]$ durch

$$v_k(x) := \cos(2\pi kx), \qquad w_k(x) := \sin(2\pi kx), \qquad \text{für } k = 0, 1, 2, \ldots$$

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 7.4.30. Beweisen Sie die Orthogonalitätsrelationen.

Wir betrachten nun endlich viele von diesen Funktionen, z.B.

$$v_0, v_1, \ldots, v_n, \qquad w_1, \ldots, w_n.$$

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

$$T = \alpha_0 v_0 + \sum_{k=1}^{n} (\alpha_k v_k + \beta_k w_k), \qquad \text{d.h.}$$

$$T(x) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \cos(2\pi kx) + \beta_k \sin(2\pi kx) \ .$$

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 7.4.20(7) 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 7.4.20 ist $\pi_n(f)$ gegeben durch

$$\pi_n(f) = \langle f, \widetilde{v}_0 \rangle \widetilde{v}_0 + \sum_{k=1}^{n} (\langle f, \widetilde{v}_k \rangle \widetilde{v}_k + \langle f, \widetilde{w}_k \rangle \widetilde{w}_k) = \frac{\langle f, v_0 \rangle}{\|v_0\|^2} + \ldots,$$

wobei $\widetilde{v}_0, \widetilde{v}_1, \widetilde{v}_2, \ldots, \widetilde{w}_1, \widetilde{w}_2, \ldots$ durch Normierung aus $v_0, v_1, v_2, \ldots, w_1, w_2, \ldots$ 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

$$\alpha_k = 2 \int_0^1 f(x) \cos(2\pi kx) dx \ , \quad k = 0, 1, 2, \ldots$$

$$\beta_k = 2 \int_0^1 f(x) \sin(2\pi kx) dx \ , \quad k = 1, 2, 3, \ldots$$

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*.

7.5 Orthogonale und unitäre Abbildungen

sec8.4

In diesem Abschnitt sei V ein euklidischer oder unitärer Vektorraum. Wir werden hier alle Überlegungen parallel für den reellen (euklidischen) und komplexen (unitären) Fall anstellen. Natürlich werden $\mathbb{R}^n$ und $\mathbb{C}^n$ mit dem Standardskalarprodukt (siehe die Abschnitte 7.1 und 7.2) die wichtigsten Beispiele sein. Wir wollen lineare Abbildungen $\varphi\colon V \to V$ betrachten, die das Skalarprodukt erhalten, so wie diese Abbildungen darstellende Matrizen.

8.4.1 **Definition 7.5.1.** [Orthogonale [unitäre] Abbildung] Eine lineare Abbildung $\varphi\colon V \to V$ heißt *orthogonal [unitär]*, wenn φ bijektiv ist und isometrisch, d.h.,

$$\langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle \quad \text{für alle} \quad v, w \in V$$

gilt.

Bemerkung 7.5.2. Orthogonale [unitäre] Abbildungen φ erhalten Länge und Orthogonalität, d.h.

$$\|\varphi(v)\| = \|v\|, \qquad v \perp w \Rightarrow \varphi(v) \perp \varphi(w).$$

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 orthogonale Abbildungen Winkel, da sich diese durch das Skalarprodukt ausdrücken lassen.

Beispiele 7.5.3. (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 7.5.4. Für endlichdimensionale V kann man die Definition abschwächen: Bijektivität ist eine Konsequenz der anderen Eigenschaften. Ist nämlich $\varphi\colon V \to V$ linear und erhält das Skalarprodukt, dann erhält φ die Länge. Also haben nichtverschwindende Vektoren ein nichtverschwindendes Bild, d.h. $\ker(\varphi) = \{0\}$. Also ist φ injektiv. Für endlichdimensionale V folgt daraus, dass φ bijektiv ist (vgl. 5.1.5).

Für unendlichdimensionale euklidische Vektorräume gibt es lineare Abbildungen, die das Skalarprodukt erhalten und nicht bijektiv sind. Im Raum ℓ^2 (siehe 7.4.26) ist der Shiftoperator $(x_1, x_2, x_3, \dots) \mapsto (0, x_1, x_2, x_3, \dots)$ hierfür ein Beispiel.

Definition 7.5.5. [Die orthogonale und die unitäre Gruppe] Wir bezeichnen mit 8.4.3

$$O(V) := \{\varphi \in \mathrm{Aut}(V)\colon (\forall v, w \in V)\, \langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle\}$$

die Menge aller orthogonalen Abbildungen eines euklidischen Vektorraums, die sogenannte *orthogonale Gruppe von V* und mit

$$U(V) := \{\varphi \in \mathrm{Aut}(V)\colon (\forall v, w \in V)\, \langle \varphi(v), \varphi(w) \rangle = \langle v, w \rangle\}$$

die Menge aller unitären Abbildungen eines unitären Vektorraums V, die sogenannte *unitäre Gruppe von V*.

Bemerkung 7.5.6. Wir behaupten, dass $O(V)$ und $U(V)$ Gruppen sind. Dazu zeigen wir, dass diese Mengen Untergruppen der Gruppe $\mathrm{Aut}(V) = \mathrm{GL}(V)$ aller linearer Automorphismen von V sind. Das tun wir für den euklidischen Fall (die Beweise und Aussagen sind im unitären Fall gleich):

1. $\mathrm{id}_V \in O(V)$, da die identische Abbildung offensichtlich orthogonal ist.
2. $\varphi, \psi \in O(V) \Rightarrow \varphi \circ \psi \in O(V)$: Erhalten φ und ψ das Skalarprodukt, so folgt für alle $v, w \in V$:

$$\langle (\varphi \circ \psi)(v), (\varphi \circ \psi)(w) \rangle = \langle \varphi(\psi(v)), \varphi(\psi(w)) \rangle = \langle \psi(v), \psi(w) \rangle = \langle v, w \rangle.$$

3. $\varphi \in O(V) \Rightarrow \varphi^{-1} \in O(V)$: 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.

8.4.2

Proposition 7.5.7. [Orthogonalitätskriterium] *Sei $v_1, v_2, \dots, v_n$ eine Orthonormalbasis von V. Dann ist eine lineare Abbildung $\varphi\colon V \to V$ genau dann orthogonal [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 orthogonal [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. Es folgt in diesem Fall

$$\langle \varphi(v_i), \varphi(v_j)\rangle = \delta_{ij}.$$

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 7.4.4:

$$\begin{aligned}\langle \varphi(v), \varphi(w)\rangle &= \langle \varphi(\sum_{i=1}^n \alpha_i v_i), \varphi(\sum_{j=1}^n \beta_j v_j)\rangle = \langle \sum_{i=1}^n \alpha_i \varphi(v_i), \sum_{j=1}^n \beta_j \varphi(v_j)\rangle \\ &= \sum_{i,j} \alpha_i \overline{\beta_j} \langle \varphi(v_i), \varphi(v_j)\rangle = \sum_{i,j} \alpha_i \overline{\beta_j} \delta_{ij} = \sum_{i=1}^n \alpha_i \overline{\beta_i} = \langle v, w\rangle.\end{aligned}$$

□

Definition 7.5.8. [Orthogonale Matrizen] Wir nennen $Q \in M_n(\mathbb{R})$ eine *orthogonale* Matrix, wenn

$$QQ^\top = Q^\top Q = E_n$$

gilt, d.h. Q ist invertierbar mit $Q^{-1} = Q^\top$. Wir schreiben

$$O_n(\mathbb{R}) := \{Q \in \mathrm{GL}_n(\mathbb{R})\colon Q^\top = Q^{-1}\}$$

für die Menge der orthogonalen Matrizen.

Für eine komplexe Matrix $A = (a_{ij})$ benutzen wir folgende Schreibweisen:

$$\overline{A} := (\overline{a_{ij}}) \quad , \quad A^* := \overline{A}^\top = \overline{A^\top} \qquad \text{(die Transponierte von } \overline{A}\text{)}$$

Beispiel 7.5.9.

$$A = \begin{pmatrix} 1+i & 1 \\ i & 2-i \end{pmatrix}, \quad \overline{A} = \begin{pmatrix} 1-i & 1 \\ -i & 2+i \end{pmatrix}, \quad A^* = \begin{pmatrix} 1-i & -i \\ 1 & 2+i \end{pmatrix}.$$

Definition 7.5.10. Wir nennen $Q \in M_n(\mathbb{C})$ eine *unitäre* Matrix, wenn

$$QQ^* = Q^*Q = E_n$$

gilt, d.h. Q ist invertierbar mit $Q^{-1} = Q^*$. Wir schreiben

$$U_n(\mathbb{C}) := \{Q \in \mathrm{GL}_n(\mathbb{C})\colon Q^* = Q^{-1}\}$$

für die Menge der unitäre Matrizen.

Beachte, dass die orthogonalen Matrizen genau diejenigen unitären Matrizen sind, bei denen alle Einträge reell sind (Nachweis).

Lemma 7.5.11. *$O_n(\mathbb{R})$ ist eine Untergruppe von $\mathrm{GL}_n(\mathbb{R})$ und $U_n(\mathbb{C})$ ist eine Untergruppe von $\mathrm{GL}_n(\mathbb{C})$.*

Beweis. Wir führen den Nachweis im reellen Fall. Den komplexen Fall behandelt man analog. Zuerst bachten wir: $E_n^\top = E_n = E_n^{-1}$, also ist die Einheitsmatrix orthogonal. Sind $A, B \in O_n(\mathbb{R})$, also $A^{-1} = A^\top$ und $B^{-1} = B^\top$, so ist auch

$$(AB)^{-1} = B^{-1}A^{-1} = B^\top A^\top = (AB)^\top,$$

und damit $AB \in O_n(\mathbb{R})$. Weiter folgt aus der Relation $A^{-1} = A^\top$ durch Transponieren

$$(A^{-1})^\top = (A^\top)^\top = A = (A^{-1})^{-1},$$

also $A^{-1} \in O_n(\mathbb{R})$. □

Definition 7.5.12. [Die orthogonalen und unitären Matrixgruppen] Wir haben schon gesehen, dass $O_n(\mathbb{R})$ und $U_n(\mathbb{C})$ Untergruppen von $\mathrm{GL}_n(\mathbb{K})$ sind. Da 8.4.8

$$\mathrm{SL}_n(\mathbb{K}) := \{g \in \mathrm{GL}_n(\mathbb{K})\colon \det g = 1\}$$

für jeden Körper eine Untergruppe von $\mathrm{GL}_n(\mathbb{K})$ ist (Nachweis als Übung), erhalten wir weitere Untergruppen

$$\mathrm{SO}_n(\mathbb{R}) := O_n(\mathbb{R}) \cap \mathrm{SL}_n(\mathbb{R}) \quad \text{und} \quad \mathrm{SU}_n(\mathbb{C}) := U_n(\mathbb{C}) \cap \mathrm{SL}_n(\mathbb{C}).$$

Diese Untergruppen heißen die *spezielle orthogonale* und die *spezielle unitäre Gruppe*.

Satz 7.5.13. [Charakterisierung von orthogonalen [unitären] Matrizen] *Für $Q \in M_n(\mathbb{K})$, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ sind die folgenden Eigenschaften äquivalent:* 8.4.6

(i) *Die lineare Abbildung $\varphi_Q\colon \mathbb{K}^n \to \mathbb{K}^n, x \mapsto Qx$ ist orthogonal [unitär] bzgl. des Standardskalarprodukts, d.h. es gilt:*
$$\langle Qv, Qw\rangle = \langle v, w\rangle \quad \textit{für alle } v, w.$$

(ii) *Die Spalten $s_1, \ldots, s_n$ von Q bilden ein Orthonormalsystem, d.h.*

$$\langle s_i, s_j\rangle = \delta_{ij} = \begin{cases} 1 & \textit{falls } i = j \\ 0 & \textit{falls } i \neq j \end{cases}$$

(iii) $Q^\top Q = E_n$ $[Q^*Q = E_n]$.

(iv) *Q ist invertierbar und es gilt* $Q^{-1} = Q^\top$ $[Q^{-1} = Q^*]$.

(v) *Die Zeilen von Q bilden ein Orthonormalsystem.*

(vi) $QQ^\top = E_n$ $[QQ^* = E_n]$.

Beweis. (i) ⇔ (ii) Betrachte die Standardbasis $e_1, e_2, \ldots, e_n$ in $\mathbb{K}^n$, die eine Orthonormalbasis ist. Wir wissen aus 7.5.7, dass φ_Q genau dann orthogonal [unitär] ist, wenn die Vektoren $\varphi_Q(e_1) = Qe_1$, $\varphi_Q(e_2) = Qe_2$, …, $\varphi_Q(e_n) = Qe_n$, also die Spalten von Q, ebenfalls eine Orthonormalbasis bilden.

(ii) ⇔ (iii) Wir betrachten $Q^\top Q = (\gamma_{ij})$ mit:

$$\begin{aligned}\gamma_{ij} &= (i - \text{te Zeile von } Q^\top) \cdot (j - \text{te Spalte von}\, Q) \\ &= (i - \text{te Spalte von } \ Q) \cdot (j - \text{te Spalte von}\, Q)\end{aligned}$$

Es gilt also genau dann $Q^\top Q = E_n = (\delta_{ij})$, wenn

$$(i - \text{te Spalte von}\, Q) \cdot (j - \text{te Spalte von}\, Q) = \delta_{ij}$$

d.h. genau dann, wenn die Spalten von Q ein Orthonormalsystem bilden.

(iii) $\Leftrightarrow$ (iv) Es ist klar, dass (iii) aus (iv) folgt. Ist $\mathbb{K} = \mathbb{R}$ und $Q^\top Q = E_n$, so ist Q nach 5.4.20 invertierbar und $Q^{-1} = Q^\top$, d.h. Q ist eine orthogonale Matrix. Ist $\mathbb{K} = \mathbb{C}$ und $Q^*Q = E_n$, so ist Q invertierbar und $Q^{-1} = Q^*$.

(v) $\Leftrightarrow$ (vi) zeigt man analog zu (ii) $\Leftrightarrow$ (iii).

(iv) $\Leftrightarrow$ (vi) zeigt man analog zu (iii) $\Leftrightarrow$ (iv).

□

8.4.4

Beispiel 7.5.14. 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 einen Orthonormalbasis bilden. Die inverse Matrix ist gegeben durch

$$Q^{-1} = Q^\top = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ -2 & 2 & -1 \\ 2 & 1 & -2 \end{pmatrix}.$$

8.4.5

Folgerung 7.5.15. *Ein Endomorphismus φ eines endlichdimensionalen euklidischen [unitären] Vektorraumes V ist genau dann orthogonal [unitär], wenn seine Matrix Q bzgl. einer beliebigen Orthonormalbasis $v_1, v_2, \ldots, v_n$ eine orthogonale [unitäre] Matrix ist.*

Beweis. Dies folgt aus 7.5.7, der Charakterisierung orthogonaler [unitärer] Matrizen, und aus der Tatsache, dass die Spalten der Matrix Q von φ bzgl. der Basis $B = (v_1, \ldots, v_n)$ die Koordinatenvektoren $[\varphi(v_1)]_B, \ldots, [\varphi(v_n)]_B$ sind. □

8.4.7 **Satz 7.5.16.** [Eigenschaften orthogonaler [unitärer] Matrizen Q]

(i) $|\det(Q)| = 1$ *(für 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 $\mathbb{R}^n$ $[\mathbb{C}^n]$, 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) $\det(Q^*) = \det(\overline{Q^\top}) = \det(\overline{Q}) = \overline{\det(Q)}$. Aus $Q^*Q = E_n$ folgt daher

$$1 = \det(E_n) = \det(Q^*Q) = \det(Q^*) \cdot \det(Q) = \overline{\det(Q)} \cdot \det(Q) = |\det(Q)|^2$$

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

$$\langle v, v\rangle = \langle Qv, Qv\rangle = \langle \lambda v, \lambda v\rangle = \overline{\lambda}\lambda\langle v, v\rangle.$$

Aus $\langle v, v\rangle \neq 0$ ergibt sich $\overline{\lambda}\lambda = 1$, d.h. $|\lambda|^2 = 1$, $|\lambda| = 1$.

(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

$$\langle v_1, v_2\rangle = \langle Qv_1, Qv_2\rangle = \langle \lambda_1 v_1, \lambda_2 v_2\rangle = \overline{\lambda_1}\lambda_2\langle v_1, v_2\rangle.$$

Wir folgern $\overline{\lambda_1}\lambda_2 = 1$, 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 orthogonal [unitär] ist, schließen wir $\langle Qv, Qu\rangle = 0$ für alle $u \in U$. Da Q orthogonal [unitär] ist und da U nach Annahme Q-invariant ist, induziert Q eine orthogonale [unitäre] Abbildung von U auf sich selbst. Die Matrix Q induziert also insbesondere eine bijektive Abbildung von U auf sich. Wir folgern $\langle Qv, w\rangle = 0$ für alle $w \in U$, also $Qv \in U^\perp$.

□

7.5.17. [Orthogonale 2×2-Matrizen Q] Es sei Q eine orthogonale 2×2-Matrix. Da die erste Spalte q_1 von Q ein Einheitsvektor ist, können wir sie für ein $\gamma \in [0, 2\pi[$ schreiben als 8.4.9

$$q_1 = \begin{pmatrix} \cos\gamma \\ \sin\gamma \end{pmatrix}.$$

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 $\det Q = -\cos^2\gamma - \sin^2\gamma = -1$.

- $\det Q = 1$: Dann beschreibt Q eine Drehung um den Ursprung um den Winkel γ ($0 \leq \gamma \leq 2\pi$):
$$Q = D(\gamma) := \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix}$$
- $\det Q = -1$: Dann kann Q als eine Spiegelung R an der x_1-Achse mit anschließender Drehung aufgefasst werden:
$$Q = \begin{pmatrix} \cos\gamma & \sin\gamma \\ \sin\gamma & -\cos\gamma \end{pmatrix} = D(\gamma) \cdot R = D(\gamma) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
Dies ist in der Tat eine Spiegelung an der Geraden g, die den Winkel $\frac{\gamma}{2}$ mit der x_1-Achse einschließt, denn wir haben
$$Q\begin{pmatrix} \cos\frac{\gamma}{2} \\ \sin\frac{\gamma}{2} \end{pmatrix} = \begin{pmatrix} \cos\frac{\gamma}{2} \\ \sin\frac{\gamma}{2} \end{pmatrix}, \quad Q\begin{pmatrix} -\sin\frac{\gamma}{2} \\ \cos\frac{\gamma}{2} \end{pmatrix} = \begin{pmatrix} \sin\frac{\gamma}{2} \\ -\cos\frac{\gamma}{2} \end{pmatrix},$$
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.

$\det Q = +1$

e_2 Qe_2 Qe_1 γ e_1

$\det Q = -1$

e_2 Qe_1 g γ $\frac{1}{2}\gamma$ e_1 Qe_2

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.

8.4.10 **7.5.18.** [Orthogonale 3×3-Matrizen] Sei $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ eine orthogonale Abbildung und Q die Matrix von φ bzgl. der Standardbasis e_1, e_2, e_3.

Wir zeigen Folgendes:

1. Ist $\det Q = +1$, dann ist φ 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 (siehe 5.8.7)
   $$\cos\gamma = \frac{1}{2}(\mathrm{tr}(Q) - 1).$$

g u 0 w v

2. Ist $\det Q = -1$, dann ist φ 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 $\lambda = -1$.

Wählt man eine Orthogonalbasis u, v, w, wie in der Zeichnung angedeutet, so lautet die Matrix Q' von φ bzgl. dieser neuen Basis

1. $\det Q = +1$: $$Q' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}$$

2. $\det Q = -1$: $$Q' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}$$

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 $\det Q = 1$ oder -1 nach 7.5.16 (i) ist.

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 (dies ist ein Satz aus der Analysis). Aus 7.5.16(ii) folgt $\lambda_1 = \pm 1$.

1. Fall: $\det Q = 1$. Wir behaupten: 1 ist Eigenwert von Q. Seien $\lambda_1, \lambda_2, \lambda_3$ die komplexen Eigenwerte von Q, wobei λ_1 der reelle Eigenwert sei. Nach dem oben Gesagten gilt $\lambda_1 = 1$ oder $\lambda_1 = -1$. Ist λ_2 nicht reell, so ist $\lambda_3 = \overline{\lambda_2}$ 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 $\lambda_3 = \overline{\lambda_2}$. 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 Eigenvektor u zum Eigenwert 1. Wir dürfen sogar annehmen, dass u normiert ist. Sei $U = \{u\}^\perp$ der zu u orthogonale lineare Teilraum. Nach 7.5.16(iv) ist U invariant unter φ.

2. Fall: $\det Q = -1$.

Mittels einer ähnlichen Argumentation wie im vorigen Fall zeigt man, dass -1 ein Eigenwert ist. Wir wählen dann einen normierten Eigenvektor u zum Eigenwert -1. Wieder ist $U := \{u\}^\perp$ invariant unter φ.

Wir wissen aus 7.5.17, dass auf dem zweidimensionalen Raum U die Abbildung φ eine Drehung oder eine Spiegelung induziert. Im ersten Fall ist φ eine Drehung um g (dann ist $\det Q = +1$) und im zweiten Fall eine Spiegelung an g (dann ist $\det Q = -1$). □

Aufgabe 7.5.19. Sei V ein euklidischer Vektorraum. Zeige: Ist $v \in V$ ein Einheitsvektor, so wird durch

$$s_v\colon V \to V, \quad w \mapsto w - 2\langle w, v\rangle v$$

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 7.5.20. * Sei V ein endlichdimensionaler euklidischer Vektorraum. Zeige: Jede orthogonale Abbildung ist ein Produkt von Spiegelungen, d.h., zu jedem $\varphi \in O(V)$ existieren $v_1, \ldots, v_n \in V$ 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 $\neq 0$.

7.5.21. [Die QR-Zerlegung einer invertierbaren Matrix B] Sei B eine invertierbare $n \times n$-Matrix 8.4.11
mit reellen [komplexen] Koeffizienten. Dann sind die n Spalten von B linear unabhängig und wir können die QR-Zerlegung von B wie in 7.4.13 konstruieren und erhalten

$$B = QR$$

Dabei ist Q eine orthogonale [unitäre] Matrix und R eine obere Dreicksmatrix mit positiven reellen Einträgen auf der Diagonalen.

Aufgabe 7.5.22. Zeige, dass die Zerlegung einer invertierbaren Matrix B in ein Produkt aus einer orthogonalen [unitären] Matrix Q und einer oberen Dreiecksmatrix R mit positiven reellen Einträgen eindeutig ist.

Bemerkung 7.5.23. 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 $Q^{-1} = Q^\top$ $[Q^{-1} = Q^*]$ können wir das System in der Form

$$Rx = Q^\top b \ \ [Rx = Q^* b]$$

schreiben. Das letzte Gleichungssystem kann leicht gelöst werden, da R obere Dreiecksgestalt besitzt.

7.6 Normale Matrizen

sec9.1

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

$$A' := Q^{-1} A Q \quad \text{eine Diagonalmatrix ist,}$$

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.

Wir definieren für Matrizen A mit komplexen Einträgen 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 in Kapitel 6 für die Cramersche Regel definiert wurde.

9.1.0 **Definition 7.6.1.** [Die Adjungierte] Man erinnere sich an die folgenden Schreibweisen: Für $A = (a_{ij}) \in M_n(\mathbb{C})$ schreiben wir

$$\begin{aligned} \overline{A} &= (\overline{a_{ij}}) \\ A^* &= \overline{A}^\top = \overline{A^\top} \end{aligned}$$

und nennen A^* die *Adjungierte von* A. Für die Adjungierte gelten folgende Regeln:

$$\begin{aligned} (A+B)^* &= A^* + B^*, & (\lambda A)^* &= \overline{\lambda} A^* \qquad (\text{für } \lambda \in \mathbb{C}), \\ (AB)^* &= B^* A^*, & A^{**} &= A. \end{aligned}$$

Die beiden ersten und die vierte Regel sind offensichtlich; wir berechnen zum Nachweis der dritten:

$$(AB)^* = \overline{AB}^\top = (\overline{A}\,\overline{B})^\top = \overline{B}^\top \overline{A}^\top = B^* A^*.$$

Sind alle Einträge a_{ij} von A reell, so gilt $\overline{A} = A$ und $A^* = A^\top$.

9.1.2a **Bemerkung 7.6.2.** [Spezialfälle] Für eine komplexe $n \times n$-Matrix A gilt:

$$\begin{aligned} A \text{ ist reell } &\iff A = \overline{A}, \\ A \text{ ist rein imaginär } &\iff A = -\overline{A}. \end{aligned}$$

Wir definieren weiterhin:

$$\begin{aligned} A \text{ ist symmetrisch} :&\iff A = A^\top \\ A \text{ ist schiefsymmetrisch} :&\iff A = -A^\top \\ A \text{ ist hermitesch} :&\iff A = A^* \\ A \text{ ist schiefhermitesch} :&\iff A = -A^* \end{aligned}$$

Hermitesche Matrizen heißen auch *selbstadjungiert*.

Bemerkung 7.6.3. Man beachte, dass die rellen (schief-)hermiteschen Matrizen genau die (schief-) symmetrischen sind. Einige Beispiele für obige Klassen:

$$\begin{array}{cccc} \text{symmetrisch} & \text{schiefsymmetrisch} & \text{hermitesch} & \text{schiefhermitesch} \\ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 2 \end{pmatrix} & \begin{pmatrix} 0 & 1 & -2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{pmatrix} & \begin{pmatrix} 1 & -i & 1-i \\ i & 2 & 2+i \\ 1+i & 2-i & -2 \end{pmatrix} & \begin{pmatrix} i & i & 1-i \\ i & 2i & 2+i \\ -1-i & -2+i & -2i \end{pmatrix} \end{array}$$

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:

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*).$$

9.1.1

Lemma 7.6.4. *Für* $A \in M_n(\mathbb{C})$ *gilt* $\boxed{\langle Av, w\rangle = \langle v, A^* w\rangle}$ *für alle* $v, w \in \mathbb{C}^n$.

Beweis. $\langle Av, w\rangle = (Av)^\top \overline{w} = v^\top A^\top \overline{w} = v^\top \overline{A^* w} = \langle v, A^* w\rangle$. □

Bemerkung 7.6.5. 1. Wegen $A^{**} = A$ folgt aus dem oben Gesagten $\langle v, Aw\rangle = \langle v, A^{**}w\rangle = \langle A^* v, w\rangle$.

2. Ist A eine *reelle* Matrix, so gilt $A^* = A^\top$ und damit:

$$\langle Av, w\rangle = \langle v, A^\top w\rangle, \quad \langle A^\top v, w\rangle = \langle v, Aw\rangle \quad \text{für alle } v, w \in \mathbb{R}^n.$$

Wir kommen nun zum zentralen Begriff dieses Kapitels:

Definition 7.6.6. [Normale Matrizen] Eine Matrix $A \in M_n(\mathbb{C})$ heißt *normal*, wenn sie mit ihrer Adjungierten vertauscht, d.h. 9.1.2

$$AA^* = A^* A.$$

Beispiele 7.6.7. 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 $Q^* Q = E_n = QQ^*$ 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.

9.1.3

Lemma 7.6.8. *Ist A normal, so gilt:*

(a) *$A - \lambda E$ ist normal für jedes $\lambda \in \mathbb{C}$,*

(b) *$Q^* AQ$ ist normal für jede unitäre Matrix Q.*

Beweis. (a) folgt aus folgender Rechnung:

$$(A-\lambda E)(A-\lambda E)^* = (A-\lambda E)(A^* - \overline{\lambda}E) = AA^* - \lambda A^* - \overline{\lambda}A + \lambda\overline{\lambda}E,$$
$$(A-\lambda E)^*(A-\lambda E) = (A^* - \overline{\lambda}E)(A-\lambda E) = A^*A - \overline{\lambda}A - \lambda A^* + \overline{\lambda}\lambda E.$$

Beide Ausdrücke sind gleich, da $AA^* = A^*A$ gilt (A ist normal).

(b) folgt aus

$$(Q^*AQ)(Q^*AQ)^* = (Q^*A\underbrace{Q)(Q^*}_{E}A^*Q^{**}) = Q^*AA^*Q,$$
$$(Q^*AQ)^*(Q^*AQ) = (Q^*A^*Q^{**})(Q^*AQ) = Q^*A^*\underbrace{QQ^*}_{E}AQ = Q^*A^*AQ.$$

Wegen $AA^* = A^*A$ sind die Matrizen Q^*AA^*Q und Q^*A^*AQ gleich. □

9.1.4

Lemma 7.6.9. *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 $\overline{\lambda}$ ein Eigenwert von A^* und v ist ein Eigenvektor von A^* bzgl. $\overline{\lambda}$.*

Beweis. Da A normal ist, gilt:

$$\begin{aligned}\|Av-\lambda v\|^2 =& \langle (A-\lambda E)v, (A-\lambda E)v\rangle \underset{7.6.4}{=} \langle (A-\lambda E)^*(A-\lambda E)v, v\rangle \\ \underset{7.6.8}{=}& \langle (A-\lambda E)(A-\lambda E)^*v, v\rangle \underset{7.6.4}{=} \langle (A-\lambda E)^*v, (A-\lambda E)^*v\rangle = \|A^*v - \overline{\lambda}v\|^2.\end{aligned}$$

Der Vektor v ist genau dann ein Eigenvektor von A zum Eigenwert λ, wenn

$$Av = \lambda v \iff (A-\lambda E)v = 0 \iff \langle (A-\lambda E)v, (A-\lambda E)v\rangle = \|(A-\lambda E)v\|^2 = 0.$$

Ebenso ist v genau dann ein Eigenvektor von A^* zum Eigenwert $\overline{\lambda}$, wenn

$$A^*v = \overline{\lambda}v \iff (A^* - \overline{\lambda}E)v = 0 \iff \langle (A-\lambda E)^*v, (A-\lambda E)^*v\rangle = \|(A-\lambda E)^*v\|^2 = 0.$$

Da nach der ersten Rechnung die beiden Ausdrücke für normale Matrizen A gleich sind, erhalten wir das gewünschte Ergebnis. □

9.1.4a

Satz 7.6.10. *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

$$\lambda \cdot \langle v, w\rangle \underset{(\mathrm{B}_2)}{=} \langle \lambda v, w\rangle = \langle Av, w\rangle \underset{(7.6.4)}{=} \langle v, A^*w\rangle \underset{(7.6.9)}{=} \langle v, \overline{\mu}w\rangle \underset{(\mathrm{B'}_1)}{=} \mu \cdot \langle v, w\rangle.$$

Aus $\mu \neq \lambda$ ergibt sich, dass $\langle v, w\rangle$ verschwindet. Also sind v und w orthogonal. □

9.1.5

Lemma 7.6.11. *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:

$$\langle v, Aw\rangle \underset{7.6.4}{=} \langle A^*v, w\rangle \underset{7.6.9}{=} \langle \overline{\lambda}v, w\rangle \underset{(\mathrm{B'}_1)}{=} \overline{\lambda} \cdot \langle v, w\rangle = 0.$$

□

Diagonalisierungssätze

9.2

Nun können wir unsere wichtigsten Sätze angeben.

9.2.1 **Satz 7.6.12.** [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. Wegem dem Fundamentalsatz der Algebra besitzt das charakteristische Polynom $\chi_A(X)$ eine Nullstelle, also A einen Eigenwert λ_1. Sei v_1 ein zugehöriger Eigenvektor. Wir normieren v_1, so dass $\|v_1\| = 1$. Ist $n = 1$, so sind wir fertig. Ist dies nicht der Fall, so nehmen wir an, wir hätten bereits ein Orthonormalsystem $(v_1, \ldots, v_k)$ von Eigenvektoren von A konstruiert. Nach 7.6.11 ist der $(n-k)$-dimensionale Unterraum

$$V_k := \{v_1, \ldots, v_k\}^\perp = v_1^\perp \cap \ldots \cap v_k^\perp$$

invariant unter A. Ist $k = n$, so sind wir fertig. Ist $k < n$, so hat die lineare Abbildung

$$V_k \to V_k, v \mapsto Av$$

einen normierten Eigenvektor v_{k+1}. Dann bilden $(v_1, \ldots, v_{k+1})$ ein Orthonormalsystem. Dieses Verfahren bricht ab, wenn wir $k = n$ erreicht haben. Wir erhalten also schließlich ein Orthonormalsystem $v_1, \ldots, v_n$ von Eigenvektoren von A.

(ii) $\Rightarrow$ (iii). Sei $B' = (v_1, v_2, \ldots, v_n)$ eine Orthonormalbasis von $\mathbb{C}^n$ aus Eigenvektoren von A. Sei Q die Matrix, deren Spalten die Vektoren $v_1, \ldots, v_n$ sind. Dann ist Q unitär (Satz 8.5.13) 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 7.6.8(b), dass A ebenfalls normal ist. □

Wir wenden nun diesen Satz auf die speziellen Klassen normaler Matrizen an, die in den Beispielen von 7.6.6 betrachtet wurden. Wir beginnen mit hermiteschen (selbstadjungierten) Matrizen:

9.2.2

Lemma 7.6.13. *Alle Eigenwerte einer hermiteschen Matrix sind reell. Insbesondere sind alle komplexen Eigenwerte einer reellen symmetrischen Matrix reell.*

Beweis. Sei $A = A^*$ und $v \in \mathbb{C}^n$ ein Eigenvektor zum Eigenwert $\lambda \in \mathbb{C}$ von A. Dann ist $Av = \lambda v$. Aus 7.6.9 folgt $A^*v = \overline{\lambda}v$ und aus $A^* = A$ folgern wir $Av = \overline{\lambda}v$ und damit $\lambda v = \overline{\lambda}v$, was $\lambda = \overline{\lambda}$ impliziert. Dies führt auf $\lambda \in \mathbb{R}$. □

Satz 7.6.14. [Hauptsatz über hermitesche Matrizen] *Für $A \in M_n(\mathbb{C})$ sind die folgenden Eigen-* 9.2.3
schaften 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 es gilt $A^*A = A^2 = AA^*$. Aus Lemma 7.6.13 wissen wir, dass alle Eigenwerte von A reell sind.

(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. □ □

9.2.4 **Satz 7.6.15.** [Hauptsatz über reelle symmetrische Matrizen] *Für eine $n \times n$-Matrix A mit* reellen *Einträgen 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, so dass*

$$A' := Q^{-1}AQ$$

eine (reelle) 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 Lemma 7.6.13 alle Eigenwerte von A reell, so dass die Diagonalisierbarkeit von A über $\mathbb{R}$ aus Bemerkung 7.3.10 folgt.

Wählen wir zu jedem Eigenwert eine Orthonormalbasis des zugehörigen Eigenraums, so folgt aus der Orthogonalität der Eigenräume (Satz 7.6.10) die Existenz einer Orthonormalbasis von $\mathbb{R}^n$, die aus Eigenvektoren von A besteht.

(ii) $\Rightarrow$ (iii): Genau wie in 7.6.12 für normale Matrizen.

(iii) $\Rightarrow$ (i): Angenommen, Q ist eine orthogonale Matrix und $A' = Q^\top AQ$ ist eine Diagonalmatrix, dann gilt $A = QA'Q^\top$ und damit $A^\top = (QA'Q^\top)^\top = Q^{\top\top}A'^\top Q^\top = QA'Q^\top = A$, d.h. A ist symmetrisch. □

Bemerkung 7.6.16. Reelle symmetrische Matrizen sind normal, aber eine komplexe symmetrische Matrix muss nicht normal sein:

$$A = \begin{pmatrix} i & 1 \\ 1 & 0 \end{pmatrix}, \quad A^* = \begin{pmatrix} -i & 1 \\ 1 & 0 \end{pmatrix}; \qquad AA^* = \begin{pmatrix} 2 & i \\ -i & 1 \end{pmatrix} \neq \begin{pmatrix} 2 & -i \\ i & 1 \end{pmatrix} = A^*A.$$

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 $\lambda_2 = 0$.

9.2.5

Beispiel 7.6.17. 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*:

$$\det(A - \lambda E) = \begin{vmatrix} \frac{2}{9} - \lambda & \frac{2}{9} & \frac{10}{9} \\ \frac{2}{9} & \frac{11}{9} - \lambda & -\frac{8}{9} \\ \frac{10}{9} & -\frac{8}{9} & \frac{5}{9} - \lambda \end{vmatrix} = -\lambda^3 + 2\lambda^2 + \lambda - 2 = -(\lambda - 1)(\lambda - 2)(\lambda + 1).$$

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 (nach 7.6.10). 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 7.6.18. [Hauptsatz über unitäre Matrizen] *Für $A \in M_n(\mathbb{C})$ sind die folgenden Eigenschaften äquivalent:* 9.2.6

(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 $A^*A = A^{-1}A = E = AA^{-1} = AA^*$ folgt, dass jede unitäre Matrix normal ist. Aus Satz 8.5.16 wissen wir, dass alle Eigenwerte von A reell sind.

(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

$$\overline{\lambda} = \lambda^{-1}$$

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

$$Q^{-1}AQ = A' = \begin{pmatrix} e^{i\gamma_1} & & & 0 \\ & e^{i\gamma_2} & & \\ & & \ddots & \\ 0 & & & e^{i\gamma_n} \end{pmatrix}.$$

9.2.7 **7.6.19.** [Normalform für orthogonale Matrizen] Nun sei A eine (reelle) orthogonale $n \times n$-Matrix. Wir behaupten, es gibt eine orthogonale Matrix Q mit

$$Q^{-1}AQ = A' = \begin{pmatrix} \overbrace{\boxed{\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix}}}^{p} & & \\ & \overbrace{\boxed{\begin{matrix} -1 & & 0 \\ & \ddots & \\ 0 & & -1 \end{matrix}}}^{q} & \\ 0 & & \overbrace{\begin{matrix} \boxed{D_1} & & \\ & \ddots & \\ & & \boxed{D_r} \end{matrix}}^{2r} \end{pmatrix}$$

wobei gilt:

- p ist die Vielfachheit des Eigenwertes $+1$ von A;
- q ist die Vielfachheit des Eigenwertes -1 von A;
- r ist die Anzahl von Paaren komplex konjugierter Eigenwerte $e^{i\gamma_j}$, $e^{-i\gamma_j}$ von A, wobei $\gamma_j \in]0, \pi[$.

Jedes D_j ist eine Matrix der Form

$$D_j = \begin{pmatrix} \cos\gamma_j & -\sin\gamma_j \\ \sin\gamma_j & \cos\gamma_j \end{pmatrix}$$

Zum Nachweis dieser Behauptung argumentieren wir wie folgt:

Eine orthogonale Matrix ist ein Spezialfall einer unitären Matrix. Wir gehen deshalb zunächst vor wie für unitäre Matrizen in 7.6.18:

- Bestimme die komplexen Eigenwerte von A:

  Sei p die Vielfachheit des Eigenwertes 1 und sei q die Vielfachheit des Eigenwertes -1. (Ist $+1$ oder -1 kein Eigenwert von A, sei $p = 0$ oder entsprechend $q = 0$.)

  Die anderen Eigenwerte haben die Form $e^{i\gamma} = \cos\gamma + i\sin\gamma$ 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 $\overline{\lambda} = a - ib$ ebenso eine Wurzel von $\chi_A(X)$ (vgl. H11) . Also ist für jeden Eigenwert von A mit der Form

$$e^{i\gamma_j} = \cos\gamma_j + \sin\gamma_j$$

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:

$$\underbrace{1,\ldots,1}_{p}, \underbrace{-1,\ldots,-1}_{q}, \underbrace{e^{i\gamma_1}, e^{-i\gamma_1},\ldots,e^{i\gamma_r}, e^{-i\gamma_r}}_{2r}.$$

nach ihrer Vielfachheit.

- Da A insbesondere unitär ist, existiert in $\mathbb{C}^n$ eine Orthonormalbasis von Eigenvektoren zu diesen Eigenwerten:

$$v_1,\ldots,v_p,v_{p+1},\ldots,v_{p+q},w_1,\overline{w_1},\ldots,w_r,\overline{w_r}$$

Man beachte, dass die Eigenvektoren $v_1, \ldots, v_p, v_{p+1}, \ldots, v_{p+q}$ zu den reellen Eigenwerten ± 1 reell gewählt werden können, denn wenn man den Gauuß-Jordan Algorithmus auf reelle Matrizen anwendet, liefert er eine reelle Basis des Lösungsraums. Man kann sogar für die komplex konjugierten Eigenwerte die Eigenvektoren komplex konjugiert wählen: Aus $Aw = \lambda w$ folgt $\overline{Aw} = \overline{\lambda w}$, da aber die Matrix A reell ist, folgern wir $A\overline{w} = \overline{\lambda}\overline{w}$.

In dem zweidimensionalen komplexen Vektorraum, der von w_j und $\overline{w_j}$ aufgespannt wird, bilden wir die Vektoren

$$u_j := \frac{1}{\sqrt{2}}(w_j + \overline{w_j})\,,\; u_j' = \frac{i}{\sqrt{2}}(w_j - \overline{w_j})$$

Natürlich sind u_j und u_j' *reelle* Vektoren, sie haben die Länge 1 und sind orthogonal:

$$\langle u_j, u_j'\rangle = \frac{-i}{2}(\underbrace{\langle w_j, w_j\rangle}_{1} + \underbrace{\langle \overline{w_j}, w_j\rangle}_{0} - \underbrace{\langle w_j, \overline{w_j}\rangle}_{0} - \underbrace{\langle \overline{w_j}, \overline{w_j}\rangle}_{1}) = 0.$$

Bzgl. u_j, u_j' gehört zur Wirkung von A die Matrix (vgl. 8.5.17)

$$D_j = \begin{pmatrix} \cos\gamma_j & -\sin\gamma_j \\ \sin\gamma_j & \cos\gamma_j \end{pmatrix},$$

denn

$$Au_j = \frac{1}{\sqrt{2}}(\lambda_j w_j + \overline{\lambda_j}\overline{w_j}) = (\operatorname{Re}\lambda_j)u_j + (\operatorname{Im}\lambda_j)u_j'$$

und

$$Au_j' = \frac{i}{\sqrt{2}}(\lambda_j w_j - \overline{\lambda_j}\overline{w_j}) = (-\operatorname{Im}\lambda_j)u_j + (\operatorname{Re}\lambda_j)u_j'.$$

7.7 Die Adjungierte eines Endomorphismus

9.3

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 sei V ein euklidischer [unitärer] Vektorraum und $\varphi : V \to V$ ein Endomorphismus.

7.7.1. [Die Adjungierte eines Endomorphismus] Eine Abbildung $\varphi^* : V \to V$ heißt *adjungiert zu* φ, wenn 9.3.1

$$\langle \varphi^*(v), w\rangle = \langle v, \varphi(w)\rangle \qquad \text{für alle } v, w \in V$$

gilt. Wie wir in 7.6.4 gesehen haben, ist die Adjungierte einer linearen Abbildung $\varphi : \mathbb{C}^n \to \mathbb{C}^n$, die durch eine Matrix A gegeben ist, diejenige lineare Abbildung, die durch die adjungierte Matrix A^* gegeben ist.

Es ist i.A. nicht offensichtlich, dass eine gegebene lineare Abbildung $\varphi : V \to V$ eine Adjungierte besitzt. Besitzt sie eine Adjungierte, so ist nicht offensichtlich, ob die Adjungierte linear und eindeutig ist:

9.3.2 **Satz 7.7.2.** [Eindeutigkeit und Linearität der Adjungierten] *Hat ein Endomorphismus $\varphi : V \to V$ eine Adjungierte φ^*, dann ist diese Adjungierte φ^* linear und durch φ eindeutig bestimmt.*

Beweis.

(i) *Eindeutigkeit:* Angenommen, φ^* und φ' sind zwei Adjungierte von φ, d.h. es gilt

$$\langle \varphi^*(v), w\rangle = \langle v, \varphi(w)\rangle \qquad \text{und} \qquad \langle \varphi'(v), w\rangle = \langle v, \varphi(w)\rangle$$

für alle $v, w \in V$. Dann folgern wir $\langle \varphi^*(v), w\rangle - \langle \varphi'(v), w\rangle = 0$, d.h.

$$\langle \varphi^*(v) - \varphi'(v), w\rangle = 0.$$

Da dies für alle $w \in V$ gilt, können wir $w = \varphi^*(v) - \varphi'(v)$ wählen und erhalten

$$\langle \varphi^*(v) - \varphi'(v), \varphi^*(v) - \varphi'(v)\rangle = 0$$

Es folgt $\varphi^*(v) - \varphi'(v) = 0$, d.h. $\varphi^*(v) = \varphi'(v)$ für alle $v \in V$. Damit gilt $\varphi' = \varphi^*$.

(ii) *Linearität:* Wir müssen $\varphi^*(u + v) = \varphi^*(u) + \varphi^*(v)$ und $\varphi^*(\lambda v) = \lambda\varphi^*(v)$ zeigen.
Es gilt für alle $u, v, w \in V$

$$\begin{aligned}\langle \varphi^*(u + v), w\rangle &= \langle u + v, \varphi(w)\rangle = \langle u, \varphi(w)\rangle + \langle v, \varphi(w)\rangle \\ &= \langle \varphi^*(u), w\rangle + \langle \varphi^*(v), w\rangle \\ &= \langle \varphi^*(u) + \varphi^*(v), w\rangle\end{aligned}$$

Da dies für alle $w \in V$ gilt, folgern wir $\varphi^*(u + v) = \varphi^*(u) + \varphi^*(v)$. Analog zeigt man $\varphi^*(\lambda v) = \lambda\varphi^*(v)$. □

Auf unendlichdimensionalen Vektorräumen V hat nicht jeder Endomorphismus eine Adjungierte. Auf endlichdimensionalen Räumen existiert die Adjungierte immer:

9.3.3 **Satz 7.7.3.** [Existenz der Adjungierten] *Sei V ein endlich dimensionaler euklidischer* [*unitärer*] *Vektorraum. Dann hat jeder Endomorphismus φ von V eine Adjungierte φ^*.*

Ist $B = (v_1, \dots, v_n)$ eine Orthonormalbasis von V und $A = [\varphi]_B$ die Matrix von φ bzgl. B, so ist A^* die Matrix $[\varphi^*]_B$ von φ^* bzgl. B.

Beweis. Sei $[v]_B \in \mathbb{K}^n$ der Koordinatenvektor von v bzgl. der Basis B und A die Matrix von φ bzgl. B.

Wir definieren φ^* als die lineare Abbildung, deren Matrix die Adjungierte A^* (bzgl. der Basis B) ist. Dann gilt

$$\begin{aligned}\langle \varphi^*(v), w\rangle &= \langle A^*[v]_B\,,\ [w]_B\rangle = \langle [v]_B\,,\ A[w]_B\rangle \\ &= \langle v, \varphi(w)\rangle \qquad\qquad \text{für alle } v, w \in V.\end{aligned}$$

Damit ist φ^* adjungiert zu φ. □

9.3.4 **Definition 7.7.4.** [Normale Endomorphismen] Ein Endomorphismus φ von V heißt *normal*, wenn ein $\varphi^* \in \mathrm{End}(V)$ existiert mit $\varphi \circ \varphi^* = \varphi^* \circ \varphi$.

Satz 7.7.5. *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 7.7.3 genau dann normal, wenn dessen Matrix A bzgl. einer Orthonormalbasis B normal ist. Die Behauptung folgt damit aus 7.6.12. □

7.7.6. [Selbstadjungierte Endomorphismen] Ein Endomorphismus φ eines euklidischen [unitären] 9.3.5
endlichdimensionalen Vektorraums heißt *selbstadjungiert*, falls $\varphi = \varphi^*$.

Nach 7.7.3 ist φ genau dann selbstadjungiert, wenn die Matrix A von φ bzgl. irgendeiner Orthonormalbasis symmetrisch [hermitesch] ist.

Aus 7.6.13 folgt, dass alle Eigenwerte eines selbstadjungierten Endomorphismus reell sind.

Wir folgern aus 7.6.15, dass ein selbstadjungierter Endomorphismus eines endlichdimensionalen euklidischen Vektorraumes eine Orthonormalbasis von Eigenvektoren hat und durch eine (reelle) Diagonalmatrix (bzgl. dieser Basis) dargestellt werden kann.

Kapitel 8

Quotientenraum und Dualraum

chap10a

8.1 Quotientenräume

[BK, 3.4]

8.2 Der Dualraum

[BK, 3.5]

Kapitel 9

Bilinearformen und Quadriken

In diesem Kapitel werden wir die Diagonalisierungssätze aus den vorigen Kapiteln auf quadratische Formen anwenden. Zuerst diskutieren wir Bilinearformen $V \times V \to \mathbb{K}$ und werden sehen, dass Bilinearformen sich, ähnlich wie lineare Abbildung, auch 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. chap10

9.1 Bilinearformen

sec10.1

Sei V ein Vektorraum über einem beliebigen Körper $\mathbb{K}$.

9.1.1. [Bilinearformen] Eine *Bilinearform* auf V ist eine Abbildung 10.1.1

$$F : V \times V \to \mathbb{K}$$

die bilinear ist, d.h. für alle $v, v', w, w' \in V$ und alle $\lambda \in \mathbb{K}$ gilt:

$$F(v + v', w) = F(v, w) + F(v', w), \quad F(\lambda v, w) = \lambda F(v, w). \tag{B1}$$

$$F(v, w + w') = F(v, w) + F(v, w'), \quad F(v, \lambda w) = \lambda F(v, w). \tag{B2}$$

Beispiele 9.1.2.

a) Das Standardskalarprodukt auf $\mathbb{R}^n$, allgemeiner das Skalarprodukt auf einem euklidischen Vektorraum ist nach Definition eine Bilinearform.

   Das (hermitesche) Standardskalarprodukt auf $\mathbb{C}^n$ ist hingegen *keine* Bilinearform.

   Für jeden Körper $\mathbb{K}$ erhalten wir auf $\mathbb{K}^n$ durch

$$F(x, y) := \sum_{j=1}^{n} x_j y_j$$

   eine Bilinearform.

b) Definiere für $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ in $\mathbb{R}^2$

$$F(x, y) = x_1 y_1 - x_1 y_2 + 2x_2 y_1 - x_2 y_2$$

   Unter Benutzung der Matrix $A = \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix}$ können wir dies in folgender Form schreiben:

$$F(x, y) = x^\top A y = (x_1 \ x_2) \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

(Rechnen Sie nach, dass F in der Tat eine Bilinearform auf $\mathbb{R}^2$ ist!)

c) Sei allgemeiner $A = (a_{ij})$ eine $n \times n$-Matrix mit Einträgen $a_{ij} \in \mathbb{K}$. Für alle $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ in $\mathbb{K}^n$ sei

$$\boxed{F(x,y) = x^\top A y = \sum_{i,j=1}^{n} a_{ij} x_i y_j.}$$

Dann ist F eine Bilinearform auf $\mathbb{K}^n$. Dies erklärt auch den Namen *Bilinearform*: Der F definierende Ausdruck ist linear in den x_i's und analog in den y_j's.

Wir werden sehen, dass jede Bilinearform auf $\mathbb{K}^n$ durch eine Matrix A in folgender Weise dargestellt werden kann:

10.1.2 **Definition 9.1.3.** [Matrix einer Bilinearform] Sei $F(x, y)$ eine Bilinearform auf einem endlichdimensionalen Vektorraum V, und sei $B = (v_1, \ldots, v_n)$ eine Basis von V. Sei

$$a_{ij} := F(v_i, v_j) \quad \text{für} \quad i, j = 1, \ldots, n.$$

Die Matrix $A := [F]_B := (a_{ij}) = (F(v_i, v_j))$ heißt *Matrix der Bilinearform F bzgl. der Basis B.*

Bemerkung 9.1.4. Die Bilinearform F ist durch ihre Matrix A bestimmt:
Sind $x = \sum_{i=1}^n x_i v_i$ und $y = \sum_{j=1}^n y_j v_j$ beliebige Vektoren in V, so gilt

$$\begin{aligned} F(x,y) = F(\sum_i x_i v_i\,, \sum_j y_j v_j) &\underset{(B1)}{=} \sum_i x_i F(v_i\,, \sum_j y_j v_j) \\ &\underset{(B2)}{=} \sum_i x_i \cdot \sum_j y_j \cdot F(v_i, v_j) = \sum_{i,j=1}^{n} x_i y_j \cdot a_{ij} = \sum_{i,j=1}^{n} a_{ij} x_i y_j \end{aligned}$$

Also ist $F(x, y) = [x]_B{}^\top\ [F]_B\ [y]_B$

10.1.3 **Satz 9.1.5.** [Basiswechsel] *Sei F eine Bilinearform auf V. Wir betrachten zwei Basen $B = (v_1, \ldots, v_n)$ und $B' = (v'_1, \ldots, v'_n)$ von V. Sei $A = (F(v_i, v_j))$ die Matrix von F bzgl. B und sei $A' = (F(v'_i, v'_j))$ die Matrix von F bzgl. B'. Dann gilt*

$$\boxed{A' = S^\top A S}$$

wobei $S = (s_{ij})$ die Transformationsmatrix wie in 5.6.1 *ist, d.h. die j-te Spalte von S ist der Koordinatenvektor $[v'_j]_B$.*

Beweis.

$$a'_{ij} = F(v'_i, v'_j) = F(\sum_{k=1}^{n} s_{ki} v_k\,, \sum_{l=1}^{n} s_{lj} v_l) = \sum_{k=1}^{n} \sum_{l=1}^{n} s_{ki} s_{lj} \underbrace{F(v_k, v_l)}_{a_{kl}} = \sum_{k,l=1}^{n} s_{ki} a_{kl} s_{lj}.$$

Alternativer Beweis. Aus 9.1.3 folgt

$$F(x, y) = [x]_{B'}{}^\top\ [F]_{B'}\ [y]_{B'} \qquad \text{und} \tag{1}$$
$$F(x, y) = [x]_B{}^\top\ [F]_B\ [y]_B \tag{2}$$

Da $[x]_B = S\ [x]_{B'}$ und $[y]_B = S\ [y]_{B'}$ nach 5.6.1 gilt, ergibt die letzte Gleichung

$$F(x, y) = [x]_B{}^\top\ [F]_B\ [y]_B = (S\ [x]_{B'})^\top [F]_B (S\ [y]_{B'}) = [x]_{B'}{}^\top\ S^\top [F]_B S\ [y]_{B'}$$

Vergleicht man diese Beziehung mit (1), folgt direkt $[F]_{B'} = S^\top [F]_B S$. □

Bemerkung 9.1.6. Man beachte, dass die Transformationsformel für die Matrix einer Bilinearform verschieden ist von der Transformationsformel für die Matrix einer linearen Abbildung. Nur wenn $S^{-1} = S^\top$ gilt, sind die beiden Transformationsformeln gleich. Der letzte Fall liegt für orthogonale Matrizen S vor, d.h. für Koordinatentransformationen zwischen Orthonormalbasen B und B'. Wir werden später sehen, wie sich dieser Sachverhalt bei der Hauptachsentransformation anwenden läßt.

10.1.4

Beispiel 9.1.7. Sei $V = \mathbb{R}^2$, $B = (e_1, e_2)$ die Standardbasis und

$$\begin{aligned} F(x,y) &= 2x_1y_2 + x_2y_1 && \text{für } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \ y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2 \\ &= x^\top Ay && \text{mit } A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

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 $F(x,y) = (x_1' \ x_2') S^\top AS \begin{pmatrix} y_1' \\ y_2' \end{pmatrix}$ und

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

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 9.1.8. [Symmetrische Bilinearformen] Eine Bilinearform $F(x,y)$ auf einem Vektorraum V heißt *symmetrisch*, wenn 10.1.5

$$F(x,y) = F(y,x) \quad \text{für alle} \quad x, y \in V$$

gilt.

Bemerkung 9.1.9. Eine Bilinearform F auf einem endlichdimensionalren Vektorraum V ist genau dann symmetrisch, wenn ihre Matrix A bzgl. einer beliebigen Basis B symmetrisch ist.

Ist F symmetrisch, so folgt

$$a_{ij} = F(v_i, v_j) = F(v_j, v_i) = a_{ji}.$$

Ist umgekehrt A symmetrisch, so gilt

$$F(y,x) = [y]_B{}^\top \ A \ [x]_B = ([y]_B \ A \ [x]_B)^\top = [x]_B{}^\top \ A^\top \ [y]_B = [x]_B{}^\top \ A \ [y]_B = F(x,y).$$

9.2 Quadratische Formen

sec10.2

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 10.2.1

$$Q\colon V \to \mathbb{K}$$

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

$$F\colon V \times V \to \mathbb{K}, \quad (x, y) \mapsto \frac{1}{2}(Q(x+y) - Q(x) - Q(y))$$

bilinear ist. Man beachte, dass F definitionsgemäß symmetrisch ist, und dass sich Q durch

$$F(x, x) = \frac{1}{2}(Q(2x) - 2Q(x)) = \frac{1}{2}(4Q(x) - 2Q(x)) = Q(x)$$

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

$$Q : V \to \mathbb{K}, \quad Q(x) = F(x, x), \quad x \in V$$

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

$$\begin{aligned}
&\tfrac{1}{2}(Q(x+y) - Q(x) - Q(y)) \\
=&\tfrac{1}{2}(F(x+y, x+y) - F(x, x) - F(y, y)) \\
=&\tfrac{1}{2}(F(x, x) + F(x, y) + F(y, x) + F(y, y) - F(x, x) - F(y, y)) \\
=&\tfrac{1}{2}(F(x, y) + \underbrace{F(y, x)}_{=F(x,y)}) = F(x, y).
\end{aligned}$$

Also ist Q eine quadratische Form und F die zugehörige Bilinearform. Wir nennen Q die zu F *assoziierte quadratische Form*. Wir erhalten so einen 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 Audrücke $a_{ij}x_ix_j + a_{ji}x_jx_i$ zu $2a_{ij}x_ix_j$ für alle $i < j$ zusammengefasst. Das können wir tun, da $a_{ij} = a_{ji}$ aufgrund der Symmetrie von A gilt.

Umgekehrt sei $Q : \mathbb{K}^n \to \mathbb{K}$ 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

$$A = \begin{pmatrix} a_1 & \frac{1}{2}b_{12} & \dots & \frac{1}{2}b_{1n} \\ \frac{1}{2}b_{12} & a_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{2}b_{1n} & \dots & \dots & a_n \end{pmatrix}$$

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

sec10.3

In diesem Abschnitt betrachten wir nur den euklidischen Vektorraum $V = \mathbb{R}^n$ mit dem Standardskalarprodukt.

10.3.1

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*

$$B' = (v'_1, \dots, v'_n)$$

von $\mathbb{R}^n$, so dass die Matrix A' von F bzgl. der Basis B' eine Diagonalmatrix ist:

$$A' = \begin{pmatrix} \lambda_1 & & & \text{\huge 0} \\ & \lambda_2 & & \\ & & \ddots & \\ \text{\huge 0} & & & \lambda_n \end{pmatrix},$$

d.h. für $x = x'_1 v'_1 + \dots + x'_n v'_n$, $y = y'_1 v'_1 + \dots + y'_n v'_n$ gilt

$$F(x, y) = \lambda_1 x'_1 y'_1 + \dots + \lambda_n x'_n y'_n,$$
$$Q(x) = \lambda_1 {x'_1}^2 + \dots + \lambda_n {x'_n}^2.$$

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 Basis B zur Basis B' heißt* Hauptachsentransformation.

Beweis. Sei A die Matrix von F bzgl. der Basis B. Da F symmetrisch ist, ist die Matrix A eine reelle symmetrische Matrix (vgl. 9.1.8). Aufgrund des Hauptsatzes für reelle symmetrische Matrizen (Satz 9.2.4) 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 9.1.5 ist $A' := Q^\top AQ$ die Matrix von F bzgl. der Basis B'. □

Definition 9.3.2. [Quadratische Hyperflächen] Für eine quadratische Form Q auf $\mathbb{R}^n$ heißt 10.3.2

$$\{x \in \mathbb{R}^n : Q(x) = 1\}$$

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

$$\sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_ix_j = 1 \tag{2}$$

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 9.3.1 gibt es eine Orthonormalbasis $B' = (v_1', \ldots, v_n')$, so dass die Gleichung (2) auf die Form

$$\lambda_1 {x_1'}^2 + \cdots + \lambda_n {x_n'}^2 = 1 \tag{3}$$

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 $v_1', \ldots, v_n'$ aufgespannt werden, heißen *Hauptachsen* der Quadrik. Die Transformation der Gleichung von der Form (2) in die Form (3) heißt *Hauptachsentransformation*.

10.3.3 **9.3.3.** [Hauptachsentransformation] Wir nehmen an, eine Quadrik im $\mathbb{R}^n$ sei durch eine Gleichung der Form

$$\sum_{i=1}^{n} a_i x_i^2 + \sum_{1\le i<j\le n} b_{ij}x_ix_j = 1 \tag{Q}$$

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

$$A = \begin{pmatrix} a_1 & \frac{1}{2}b_{12} & \cdots & \frac{1}{2}b_{1n} \\ \frac{1}{2}b_{12} & a_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}b_{1n} & \cdots & \cdots & a_n \end{pmatrix}.$$

Diese Matrix ist symmetrisch und reell.

(2) Gleichung (Q) kann nun in folgender Form geschrieben werden:

$$x^\top Ax = 1. \tag{Q_1}$$

(3) Berechne die Eigenwerte $\lambda_1, \ldots, \lambda_n$ von A, jeden mit seiner algebraischen Vielfachheit. (Aus Lemma 9.2.2 wissen wir, dass alle Eigenwerte von A reell sind.)

(4) Bestimme eine Orthonormalbasis $B' = (v_1', \ldots, v_n')$ von Eigenvektoren von A in $\mathbb{R}^n$. Die Existenz einer solchen Basis folgt aus dem Hauptsatz über symmetrische Matrizen (Satz 9.2.4).

(5) Sei $S = (v_1', \ldots, v_n')$ 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 Suadrik bzgl. der neuen Koordinaten beschreibt: Seien $x'_1, \ldots, x'_n$ 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

$$x^\top Ax = (x'_1 \ldots x'_n) S^\top AS \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = (x'_1 \ldots x'_n) A' \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}.$$

Bzgl. der neuen Basis B' wird die Quadrik durch die folgende Gleichung beschrieben:

$$\lambda_1 {x'_1}^2 + \lambda_2 {x'_2}^2 + \cdots + \lambda_n {x'_n}^2 = 1. \qquad (Q')$$

9.3.1 Klassifikation der quadratischen Formen

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. Insbsondere 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:

$$Q(x) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2.$$

Für eine Klassifikation genügt es, quadratische Formen dieses Typs zu betrachten. Ist einer oder mehrere der Koeffizienten $\lambda_i = 0$, dann heißt die quadratische Form *degeneriert* order *entartet*. Ist $\lambda_i \neq 0$ für alle i, dann heißt sie *nichtdegeneriert* oder *nicht entartet*. Zur Klassifikation betrachten wir die Quadriken mit der Gleichung

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 = 1.$$

9.3.4. [Fall $\boxed{n = 2}$: Quadratische Kurven] 10.3.5

Nach Hauptachsentransformation erhält die quadratische Gleichung die Form:

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 = 1. \qquad (Q)$$

Wir müssen folgende Fälle betrachten:

(a) $\underline{\lambda_1 > 0 \text{ und } \lambda_2 > 0}$:
Sei $a = \frac{1}{\sqrt{\lambda_1}}$, $b = \frac{1}{\sqrt{\lambda_2}}$. 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) $\underline{\lambda_1 > 0\,,\ \lambda_2 < 0}$:
Sei $a := \frac{1}{\sqrt{\lambda_1}}$, $b := \frac{1}{\sqrt{-\lambda_2}}$. 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 $x_1 = a\sqrt{\frac{x_2^2}{b^2}+1}$ und somit

$$\lim_{x_2\to\infty}\frac{x_1}{x_2} = \lim_{x_2\to\infty} a\sqrt{\frac{1}{b^2}+\frac{1}{x_2^2}} = \frac{a}{b}.$$

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) $\underline{\lambda_1 < 0\,,\ \lambda_2 < 0:}$ Die Gleichung (Q) hat keine reelle Lösung.

(d) $\underline{\lambda_1 > 0\,,\ \lambda_2 = 0:}$

Hier vereinfacht sich (Q) zu $\lambda_1 x_1^2 = 1$, also gilt

$$\boxed{x_1 = \pm\tfrac{1}{\sqrt{\lambda_1}} = \pm a}$$

In diesem degenerierten Fall besteht die quadratische Kurve aus zwei parallelen Geraden (in der Zeichnung punktiert gezeichnet).

10.3.7

Beispiel 9.3.5. Wir betrachten zunächst die quadratische Gleichung:

$$\boxed{x_1^2 + 6x_1x_2 + x_2^2 = 1.} \qquad (Q)$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$.

Charakteristische Gleichung: $\det(A-\lambda E) = \begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 9 = 0$.

Eigenwerte: $\lambda_1 = 4$, $\lambda_2 = -2$.

Eigenvektoren (normiert):

$$v_1' = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad v_2' = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

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:

$$\boxed{4{x_1'}^2 - 2{x_2'}^2 = 1} \qquad (Q')$$

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

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}, \qquad \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = S^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

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 9.3.6. Wir betrachten nun die quadratische Gleichung

$$\boxed{7x_1^2 + 6\sqrt{3}x_1x_2 + 13x_2^2 = 16.} \tag{Q}$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} 7 & 3\sqrt{3} \\ 3\sqrt{3} & 13 \end{pmatrix}$
Charakteristisches Polynom:

$$\det(A - XE) = \begin{vmatrix} 7-t & 3\sqrt{3} \\ 3\sqrt{3} & 13-t \end{vmatrix} = (7-t)(13-t) - 27 = t^2 - 20t + 64.$$

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:

$$v_1' = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}, \qquad v_2' = \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}.$$

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 $4{x_1'}^2 + 16{x_2'}^2 = 16$ oder

$$\boxed{\tfrac{{x_1'}^2}{4} + {x_2'}^2 = 1.} \tag{Q'}$$

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:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}, \qquad \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = S^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

oder

$$\begin{aligned} x_1 &= \frac{1}{2}(\sqrt{3}\,x_1' - x_2'), & x_1' &= \frac{1}{2}(\sqrt{3}\,x_1 + x_2), \\ x_2 &= \frac{1}{2}(x_1' + \sqrt{3}\,x_2'), & x_2' &= \frac{1}{2}(-x_1 + \sqrt{3}\,x_2). \end{aligned}$$

9.3.7. [Fall $\boxed{n=3}$: Quadratische Flächen] (nichtentartete Fälle) 10.3.6
Wir betrachten die quadratische Gleichung

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^3 = 1. \tag{Q}$$

(a) $\underline{\lambda_1 > 0,\ \lambda_2 > 0,\ \lambda_3 > 0}$:
Sei $a := \frac{1}{\sqrt{\lambda_1}}$, $b := \frac{1}{\sqrt{\lambda_2}}$, $c := \frac{1}{\sqrt{\lambda_3}}$. Dann wird Gleichung (Q) zu

$$\boxed{\tfrac{x_1^2}{a^2} + \tfrac{x_2^2}{b^2} + \tfrac{x_3^2}{c^2} = 1}$$

Dies ist die Gleichung eines *Ellipsoids* mit den Halbachsen a, b, c.

(b) $\underline{\lambda_1 > 0,\ \lambda_2 > 0,\ \lambda_3 < 0}$:
Nun sei $a := \frac{1}{\sqrt{\lambda_1}}$, $b := \frac{1}{\sqrt{\lambda_2}}$, $c := \frac{1}{\sqrt{-\lambda_3}}$. Dann wird (Q) zu

$$\boxed{\tfrac{x_1^2}{a^2} + \tfrac{x_2^2}{b^2} - \tfrac{x_3^2}{c^2} = 1.}$$

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) $\underline{\lambda_1 < 0,\ \lambda_2 < 0,\ \lambda_3 > 0}$:
Sei $a := \frac{1}{\sqrt{-\lambda_1}}$, $b := \frac{1}{\sqrt{-\lambda_2}}$, $c := \frac{1}{\sqrt{\lambda_3}}$. (Q) wird dann zu

$$\boxed{-\tfrac{x_1^2}{a^2} - \tfrac{x_2^2}{b^2} + \tfrac{x_3^2}{c^2} = 1}$$

Dies ist ein *zweischaliges Hyperboloid*.

(d) $\underline{\lambda_1 < 0,\ \lambda_2 < 0,\ \lambda_3 < 0}$: Die Gleichung (Q) hat keine reelle Lösung.

Aufgabe 9.3.8. Untersuchen Sie die degenerierten Fälle, bei denen ein oder zwei der Koeffizienten λ_i gleich 0 sind.

Bemerkung 9.3.9. 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 9.3.10. Im Fall $a = b$ ergeben sich Fächen, 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 des Eigenwertes eine Ellipse oder eine Hyperbel. Die Schnittmengen der Fläche mit horizontalen Ebenen (parallel zur $x_1 - x_2$-Ebene) sind Kreise.

Bemerkung 9.3.11. [Quadriken mit linearen Termen] Wir betrachten nun allgemeiner eine Gleichung der Form

$$\sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{1 \le i < j \le n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i = c, \tag{9.1}$$

die sich in Matrixschreibweise liest als

$$x^\top A x + b^\top x = c \tag{9.2}$$

mit einer symmetrischen Matrix A und einem Vektor b.

Für den Fall $n = 1$ erhalten wir durch quadratische Ergänzung

$$ax^2 + bx = a\big(x + \frac{b}{2a}\big)^2 - \frac{b^2}{4a}.$$

Dies können wir auf den allgemeinen Fall übertragen sofern A invertierbar ist: Für $v := \frac{1}{2}A^{-1}b$ erhalten wir

$$(x+v)^\top A(x+v) = x^\top Ax+v^\top Ax+x^\top Av+v^\top Av = x^\top Ax+2x^\top Av+v^\top Av = x^\top Ax+x^\top b+v^\top Av.$$

Also ergibt sich

$$x^\top Ax + b^\top x = (x + v)^\top A(x + v) - v^\top Av,$$

so dass (10.2) mit $x' := x + v$ äquivalent ist zu

$$x'^\top Ax' = c' := c + v^\top Av.$$

Die Fläche ist also eine der vorigen Flächen, allerdings um $-v$ *verschoben*.

Achtung: Das obige Verfahren funktioniert nur, wenn A invertierbar ist. Im Allgemeinen kann man die linearen Terme nicht wegtransformieren. Ein Beispiel ist die Parabel mit der Gleichung $x_2^2 - x_1 = 0$.

10.3.8

Beispiel 9.3.12. Wir betrachten die quadratische Gleichung

$$\boxed{x_1x_2 + x_1x_3 - x_2x_3 = 1.} \qquad (Q)$$

Zugehörende 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:

$$\begin{aligned} \det(A - XE) &= \begin{vmatrix} -t & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -t & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -t \end{vmatrix} = -t^3 - \frac{1}{8} - \frac{1}{8} + \frac{t}{4} + \frac{t}{4} + \frac{t}{4} \\ &= -t^3 + \frac{3}{4}t - \frac{1}{4} = (t+1)(-t^2 + t - \frac{1}{4}) = -(t+1)\Big(t - \frac{1}{2}\Big)^2. \end{aligned}$$

(dabei haben wir die Wurzel -1 geraten, um $t + 1$ 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

$$\begin{array}{ccc} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{array} \quad \rightsquigarrow \quad \begin{array}{ccc} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{4} & -\frac{3}{4} \\ 0 & -\frac{3}{4} & \frac{3}{4} \end{array} \quad \rightsquigarrow \quad \begin{array}{ccc} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{4} & -\frac{3}{4} \\ 0 & 0 & 0 \end{array}$$

  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

$$\begin{array}{ccc} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array} \quad \rightsquigarrow \quad \begin{array}{ccc} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}$$

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

$$\boxed{-{x_1'}^2 + \tfrac{1}{2}{x_2'}^2 + \tfrac{1}{2}{x_3'}^2 = 1}. \qquad (Q')$$

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 9.3.13. 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 9.3.14. Wir betrachten nun die Gleichung

$$\boxed{-7x_1^2 - x_2^2 - x_3^2 + 8x_1x_2 + 8x_1x_3 + 16x_2x_3 = 9.} \qquad (Q)$$

Zugehörige Matrix der quadratischen Form: $A = \begin{pmatrix} -7 & 4 & 4 \\ 4 & -1 & 8 \\ 4 & 8 & -1 \end{pmatrix}$

Charakteristisches Polynom:

$$\begin{aligned} \det(A - \lambda E) = \begin{vmatrix} -7-t & 4 & 4 \\ 4 & -1-t & 8 \\ 4 & 8 & -1-t \end{vmatrix} &= (-7-t)(-1-t)(-1-t) + \\ &+ 128 + 128 + 16(1+t) + 16(1+t) + 64(7+t) \\ &= -(7+t)(1+t)^2 + 96t + 736 = -t^3 - 9t^2 + 81t + 729 \\ &= -(t-9)(t+9)^2 \end{aligned}$$

Eigenwerte: $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = -9$.
Nach Koordinatentransformation lautet die Gleichung $9{x'_1}^2 - 9{x'_2}^2 - 9{x'_3}^2 = 9$ oder

$$\boxed{{x'_1}^2 - {x'_2}^2 - {x'_3}^2 = 1.} \tag{Q'}$$

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 9.3.15. Wir betrachten abschließend

$$\boxed{7x_1^2 + 6x_2^2 + 5x_3^2 - 4x_1x_2 - 4x_2x_3 = 3.} \tag{Q}$$

Zugehörige Matrix der Bilinearform: $A = \begin{pmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{pmatrix}$

Charakteristisches Polynom:

$$\begin{aligned} \det(A - XE) &= \begin{vmatrix} 7-t & -2 & 0 \\ -2 & 6-t & -2 \\ 0 & -2 & 5-t \end{vmatrix} = (7-t)(6-t)(5-t) - 4(7-t) - 4(5-t) \\ &= (7-t)(6-t)(5-t) - 8(6-t) = (6-t)((7-t)(5-t) - 8) \\ &= (6-t)(t^2 - 12t + 27) = (6-t)(t-3)(t-9). \end{aligned}$$

Eigenwerte: $\lambda_1 = 3$, $\lambda_2 = 6$, $\lambda_3 = 9$.
Die Gleichung nach Hauptachsentransformation ist $3{x'_1}^2 + 6{x'_2}^2 + 9{x'_3}^2 = 3$ oder

$$\boxed{{x'_1}^2 + 2{x'_2}^2 + 3{x'_3}^2 = 1.} \tag{Q'}$$

Dies ist die Gleichung eines Ellipsoids mit den Hauptachsen

$$a = 1\,, \qquad b = \frac{1}{\sqrt{2}}\,, \qquad c = \frac{1}{\sqrt{3}}.$$

Die Hauptachsen erhält man durch die Berechnung (normierter) Eigenvektoren zu den Eigenwerten λ_1, λ_2, λ_3:

$$v'_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \qquad v'_2 = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}, \qquad v'_3 = \frac{1}{3}\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}.$$

Die Transformationsmatrix ist $Q = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix}$.

9.4 Definitheit von Formen und von Matrizen

sec10.4

Sei V ein reeller Vektorraum, $F\colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf V und $Q\colon V \to \mathbb{R}$ die zugehörige quadratische Form, also $Q(v) = F(v,v)$. Insbesondere ist dann $Q(0) = F(0,0) = 0$.

10.4.1

Definition 9.4.1. 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

$$F(x, y) := x^\top Ay \quad \text{für alle } x, y \in \mathbb{R}^n$$

und die zugehörige quadratische Form:

$$Q(x) := x^\top Ax \quad \text{für alle } x \in \mathbb{R}^n.$$

Wir nennen eine Matrix A *positiv* [*negativ*] *(semi-)definit* oder *indefinit*, wenn die zugehörige quadratische Form diese Eigenschaften hat.

Bemerkung 9.4.2. In Kapitel 8 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

$$F(x, y) = x^\top Ay = \overbrace{\langle x, Ay\rangle}^{\text{(Standardskalarprodukt)}} \qquad \text{für alle } x, y \in \mathbb{R}^n$$

ein Skalarprodukt auf $\mathbb{R}^n$, das vom Standardskalarprodukt verschieden ist, falls $A \neq E$ ist. Umgekehrt ist jedes Skalarprodukt auf $\mathbb{R}^n$ auf diese Weise durch eine positiv definite symmetrische Matrix definiert.

Aufgabe 9.4.3. 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.

10.4.2 **Beispiele 9.4.4.** Sei $V = \mathbb{R}^2$ und $Q(x) = \lambda_1 x_1^2 + \lambda_2 x_2^2$.

(a) $\underline{\lambda_1 > 0, \lambda_2 > 0}$: Q ist *positiv definit.*

Die durch die Gleichung

$$x_3 = Q(x)$$

gegebene Fäche 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) $\underline{\lambda_1 < 0, \lambda_2 < 0}$: Q ist *negativ definit.*

Die durch die Gleichung

$$x_3 = Q(x)$$

gegebene Fläche ist wieder ein elliptisches Paraboloid (nach unten geöffnet).

(c) $\underline{\lambda_1 > 0, \lambda_2 < 0}$: 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) $\underline{\lambda_1 > 0,\ \lambda_2 = 0}$: 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) $\underline{\lambda_1 < 0,\ \lambda_2 = 0}$: Q ist *negativ semidefinit.*
Der parabolische Zylinder ist nun nach unten offen.

9.4.5. [Quadratische Formen auf $\mathbb{R}^2$] Wir betrachten auf $\mathbb{R}^2$ eine beliebige quadratische Form 10.4.3

$$\boxed{Q(x) = ax_1^2 + 2bx_1x_2 + cx_2^2} \tag{Q}$$

(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 9.4.4 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

$$S^\top AS = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

gilt und damit $\det A = \det(S^\top AS) = \lambda_1 \cdot \lambda_2$ ist.

Wir folgern:

- Ist $\underline{\det A < 0}$, so gilt $\lambda_1 \cdot \lambda_2 < 0$, d.h. Q ist indefinit.
- Ist $\underline{\det A > 0}$, 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 9.4.6. *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:

10.4.4 **Satz 9.4.7.** [Kriterium für positive Definitheit] *Sei* $A = (a_{ij})$ *eine symmetrische relle* $n \times n$*-Matrix. Dann sind die folgenden Eigenschaften äquivalent*

(a) *A ist positiv definit;*

(b) *$\lambda_i > 0$ für jeden Eigenwert λ_i von A;*

(c) $\det(A_r) > 0$ *für* $r = 1, 2, \ldots, n$*, wobei die Matrizen*

$$A_r := \begin{pmatrix} a_{11} & \ldots & a_{1r} \\ \vdots & & \vdots \\ a_{r1} & \ldots & a_{rr} \end{pmatrix}$$

die Hauptminoren *von A sind.*

Analog sind die folgenden Eigenschaften äquivalent:

(a') *A (und Q) sind negativ definit;*

(b') *$\lambda_i < 0$ für alle Eigenwerte λ_i von A;*

(c') $(-1)^r \det(A_r) > 0$ *für* $r = 1, 2, \ldots, n$.

Wir können die Hauptminoren schematisch wie folgt darstellen:

$$A = r\left\{ \overbrace{\begin{matrix} \quad & \quad & \quad \\ & A_r & \\ & & \end{matrix}}^{r} \right. \quad \ddots$$

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:

$$Q(x) := x^\top Ax = \lambda_1 {x_1'}^2 + \cdots + \lambda_n {x_n'}^2.$$

Dabei sind die $\lambda_1, \lambda_2, \ldots \lambda_n$ die Eigenwerte von A. In dieser Form erkennt man direkt, dass $Q(x)$ genau dann positiv [negativ] definit ist, wenn alle λ_i positiv [negativ] sind.

(a) $\Longrightarrow$ (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:

$$\begin{aligned} \det A_r &= \det(Q A'_r Q^{-1}) = \det Q \cdot \det A'_r \cdot \det Q^{-1} = \det A'_r \\ &= \mu_1 \cdot \mu_2 \ldots \cdot \mu_r > 0. \end{aligned}$$

Damit gilt $\det A_r > 0$ für alle Hauptminoren.

(c) $\Longrightarrow$ (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 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.

Es sei nun W die (direkte) Summe der Eigenräume zu den nichtpositiven (≤ 0) Eigenwerten. Eingeschränkt auf W ist die durch A definierte Form negativ semidefinit. Dies folgt aus der Implikation b)$\Rightarrow$a), da die Einschränkung von A auf W nur nichtpositive Eigenwerte hat. Ist $x \in U \cap W$ und $x \neq 0$, so folgt $x^\top Ax > 0$ und $x^\top Ax \leq 0$, was nicht möglich ist, also ist $U \cap W = \{0\}$.

Da U und W Unterräume von $\mathbb{R}^n$ sind, gilt $\dim(U + W) \leq n$. Es gilt andererseits wegen $U \cap W = \{0\}$ die Beziehung $\dim(U + W) = \dim(U) + \dim(W) = n - 1 + s$, wobei s die Zahl der nichtpositiven Eigenwerte mit deren algebraischer (= geometrischer) Vielfachheit ist. Also ist $s = 0$ oder $s = 1$. Die Determinante von A ist das Produkt der Eigenwerte (mit deren algebraischer Vielfachheit). Wegen $\det(A) = \det(A_n) > 0$ gilt $s = 0$, also sind alle Eigenwerte von A strikt positiv. $\square$

10.4.5

Beispiel 9.4.8. (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.

10.4.6 **Satz 9.4.9.** [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

$$S^\top AS = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A sind.

Da A positiv definit ist, gilt $\lambda_i > 0$ für alle i nach 9.4.7. Wir können die Matrix

$$D = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix} \quad \text{bilden mit } D \cdot D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

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 $B^\top = (DS^\top)^\top = S^{\top\top} D^\top = SD$. 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. $\square$

10.4.7 **Beispiel 9.4.10.** 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 9.4.8 vor). Nach 9.4.9 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.)

9.4.11. [Anwendung auf Lösen linearer Gleichungssysteme] Wir betrachten ein System linearer 10.4.8
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 9.4.12. Sei A wie in 9.4.10 und $b = \begin{pmatrix} -2 \\ 4 \\ 2 \\ -5 \end{pmatrix}$.

Um (L) $Ax = b$ zu lösen, lösen wir nacheinander

$$\begin{array}{lrll}
R^\top y = b, \text{ d.h} & 2y_1 = -2 & y_1 = -1 & (1) \\
 & -y_1 + y_2 = 4 & y_2 = 3 & \\
 & 2y_1 - y_2 + y_3 = 2 & y_3 = 7 & \\
 & y_1 - 2y_2 - 2y_3 + 2y_4 = -5 & y_4 = 8 & \\
Rx = y, \text{ d.h.} & 2y_4 = 8 & y_4 = 4 & (2) \\
 & y_3 - 2y_4 = 7 & y_3 = 15 & \\
 & y_2 - y_3 - 2y_4 = 3 & y_2 = 26 & \\
 & 2y_1 - y_2 + 2y_3 + y_4 = -1 & y_1 = -9/2 &
\end{array}$$

Dieses Verfahren funktioniert nur, wenn die Koeffizientenmatrix A symmetrisch und positiv definit ist. Zur allgemeineren Verwendbarkeit benutzen wir folgenden Trick:

Angenommen, A sei invertierbar. Nach der Übung in 9.4.1 wissen wir, dass $B := A^\top A$ positiv definit und symmetrisch ist. Um

$$Ax = b \tag{L}$$

zu lösen, sei $B = A^\top A$ und $c = A^\top b$ und man betrachte das äquivalente System

$$A^\top Ax = A^\top b \qquad \text{oder} \qquad Bx = c, \tag{L'}$$

welches dieselbe Lösung hat. Auf dieses System wenden wir obige Prozedur an.

Natürlich kann (L) mit dem Gauß-Jordan-Algorithmus gelöst werden. Sie werden in der Vorlesung Numerische Mathematik Fälle kennenlernen, bei denen der Gauß-Jordan-Algorithmus aufgrund großer Rundungsfehler nicht praktikabel ist. In solchen Fällen ist die Choleskyzerlegung oft wesentlich besser.

Kapitel 10

Tensorprodukte

Kapitel 11

Konvexgeometrie

[BK,6]

11.1 Polyeder

11.2 Extremalprobleme