Lineare Algebra

Inhalt

Literatur

[1] G. Fischer, "Lineare Algebra," Vieweg (2005).

Teil I
Grundlegende Methoden und Begriffe

1 Lineare Gleichungssysteme und Matrizen

Wir geben hier zunächst eine elementare Einführung in lineare Gleichungssysteme (LGS) und das Standardverfahren zur Bestimmung der Lösung. Eine systematische Untersuchung der Menge der Lösungen erfolgt später, wenn Vektorräume und linearen Abbildungen zwischen ihnen bereitgestellt sind.

Definition 1.1 Es sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Ein *lineares Gleichungssystem* über $\mathbb{K}$ aus m Gleichungen mit n Unbekannten ist ein System der Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

wobei die Koeffizienten $(a_{ij}) = \{a_{11}, \ldots, a_{mn}\} \in \mathbb{K}$ und die $b_1, \ldots, b_m \in \mathbb{K}$ fest vorgegebene Zahlen sind und das n-Tupel $(x_1, \ldots, x_n) \in \mathbb{K}^n$ eine zu bestimmende Lösung des Systems ist.

Wichtig ist, daß keine Exponenten wie x_1^2, x_2^{-1}, x_n^3 usw. der x_i auftreten.

Beispiel 1.2 Gegeben sei das System

$$\begin{array}{rcll} x_1 + x_3 - x_5 = 2 & \Leftrightarrow & 1 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 + 0 \cdot x_4 + (-1) \cdot x_5 & = 2 \\ 3x_2 - x_4 = 6 & \Leftrightarrow & 0 \cdot x_1 + 3 \cdot x_2 + 0 \cdot x_3 + (-1) \cdot x_4 + 0 \cdot x_5 & = 6 \\ x_2 + \frac{1}{4}x_3 = -\frac{1}{4}x_5 & \Leftrightarrow & 0 \cdot x_1 + 1 \cdot x_2 + \frac{1}{4} \cdot x_3 + 0 \cdot x_4 + \frac{1}{4} \cdot x_5 & = 0 \\ x_1 + x_2 + x_3 + x_4 = x_5 & \Leftrightarrow & 1 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 + 1 \cdot x_4 + (-1) \cdot x_5 & = 0 \end{array}$$

◁

Offenbar gilt:

Satz 1.3 (elementare Zeilenumformungen) *Gegeben sei ein lineares Gleichungssystem*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Die Menge der Lösungen $(x_1, \ldots, x_n) \in \mathbb{K}^n$ des Systems bleibt unverändert bei folgenden Typen von Änderungen des Systems:

Typ I *Vertauschen zweier Gleichungen.*

Typ II *Ersetzen der j-ten Gleichung durch die Summe aus der i-ten und j-ten Gleichung und Beibehalten der i-ten Gleichung.*

Typ III *Multiplikation der i-ten Gleichung mit $\lambda \in \mathbb{K}^*$ (d.h. $\lambda \neq 0$).*

Die Änderungen vom Typ II und III werden oft zusammen ausgeführt:

Typ IV *Ersetzen der j-ten Gleichung durch die Summe aus dem λ-fachen der i-ten Gleichung und der j-ten Gleichung und Beibehalten der i-ten Gleichung.*

Alle diese elemenaren Zeilenumformungen lassen sich wieder rückgängig machen. Z.B. Typ II durch

i) Multiplikation der i-ten Gleichung mit (-1)

ii) Addition der neuen i-ten Gleichung zur j-ten

iii) Multiplikation der i-ten Gleichung mit (-1)

und Typ III durch

i) Multiplikation der i-ten Gleichung mit $\frac{1}{\lambda}$.

Die Lösungsstrategie besteht darin, durch elementaren Zeilenumformungen das System so umzuformen, daß zumindest eine der Variablen allein steht und abgelesen werden kann. Das Verfahren wird dann für die verbliebenen $n-1$ Variablen wiederholt, u.s.w. Wir sehen uns das Beispiel 1.2 an:

Beispiel 1.4 I_{ij} bedeutet Vertauschen der Gleichungen i und j und $\mathrm{IV}_{ij}(\lambda)$ bedeutet Addition der λ-fachen i-ten Gleichung zur j-ten:

$$\begin{aligned}
1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + 0\cdot x_4 + (-1)\cdot x_5 &= 2\\
0\cdot x_1 + 3\cdot x_2 + 0\cdot x_3 + (-1)\cdot x_4 + 0\cdot x_5 &= 6\\
0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + 0\cdot x_4 + \tfrac{1}{4}\cdot x_5 &= 0\\
1\cdot x_1 + 1\cdot x_2 + 1\cdot x_3 + 1\cdot x_4 + (-1)\cdot x_5 &= 0
\end{aligned}$$

$$\mathrm{I}_{23},\ \mathrm{IV}_{14}(-1):\quad \begin{aligned}
1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + 0\cdot x_4 + (-1)\cdot x_5 &= 2\\
0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + 0\cdot x_4 + \tfrac{1}{4}\cdot x_5 &= 0\\
0\cdot x_1 + 3\cdot x_2 + 0\cdot x_3 + (-1)\cdot x_4 + 0\cdot x_5 &= 6\\
0\cdot x_1 + 1\cdot x_2 + 0\cdot x_3 + 1\cdot x_4 + 0\cdot x_5 &= -2
\end{aligned}$$

$$\mathrm{IV}_{23}(-3),\ \mathrm{IV}_{24}(-1):\quad \begin{aligned}
1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + 0\cdot x_4 + (-1)\cdot x_5 &= 2\\
0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + 0\cdot x_4 + \tfrac{1}{4}\cdot x_5 &= 0\\
0\cdot x_1 + 0\cdot x_2 + (-\tfrac{3}{4})\cdot x_3 + (-1)\cdot x_4 + (-\tfrac{3}{4})\cdot x_5 &= 6\\
0\cdot x_1 + 0\cdot x_2 + (-\tfrac{1}{4})\cdot x_3 + 1\cdot x_4 + (-\tfrac{1}{4})\cdot x_5 &= -2
\end{aligned}$$

$$\mathrm{IV}_{34}(-\tfrac{1}{3}):\quad \begin{aligned}
1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + 0\cdot x_4 + (-1)\cdot x_5 &= 2\\
0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + 0\cdot x_4 + \tfrac{1}{4}\cdot x_5 &= 0\\
0\cdot x_1 + 0\cdot x_2 + (-\tfrac{3}{4})\cdot x_3 + (-1)\cdot x_4 + (-\tfrac{3}{4})\cdot x_5 &= 6\\
0\cdot x_1 + 0\cdot x_2 + 0\cdot x_3 + \tfrac{4}{3}\cdot x_4 + 0\cdot x_5 &= -4
\end{aligned}$$

Daraus lesen wir $x_4 = -3$ ab. Schrittweise bestimmen sich die anderen Variablen zu $x_3 + x_5 = -4$, dann $x_2 = 1$ und schließlich $x_1 = 6 + 2x_5$. Die Variable x_5 ist dabei beliebig. $\triangleleft$

Das ist eine typische Situation: in Lösungen linearer Gleichungssysteme können manche Variablen beliebig gewählt werden, andere sind eindeutig bestimmt, oder es kann auch gar keine Lösung geben. Wir werden im Verlauf des Semesters Methoden kennenlernen, mit denen die Menge der Lösungen eines LGS charakterisiert werden kann. Zunächst führen wir eine Matrixschreibweise ein, mit der lineare Gleichungssysteme und die elementaren Zeilenumformungen sich übersichtlicher schreiben lassen.

Definition 1.5 Ein Schema der Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

mit $a_{ij} \in \mathbb{K}$ für alle $1 \leq i \leq m$ und $1 \leq j \leq n$ heißt $m \times n$-*Matrix (über* $\mathbb{K}$). Die Menge aller $m \times n$-Matrizen über $\mathbb{K}$ wird mit $M(m \times n, \mathbb{K})$ bezeichnet.

Man schreibt auch $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ oder, wenn die Größe $m \times n$ klar ist, auch $A = (a_{ij})$, mit dem Eintrag a_{ij} auf der Kreuzung der i-ten Zeile mit der j-ten Spalte. Der erste Index ist also der Zeilenindex, der zweite Index der Spaltenindex.

Wir identifizieren $M(n \times 1, \mathbb{K})$ mit $\mathbb{K}^n$, d.h. wir schreiben Elemente aus $\mathbb{K}^n$ in der Regel als Spalten $x = (x_i)_{i=1,\dots,n} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Sind $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{K}^n$ und ist $\lambda \in \mathbb{K}$, so erklärt man die Addition und die Multiplikation mit Skalaren durch

$$x + y := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \qquad \text{und} \qquad \lambda x := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Schließlich erklären wir eine Matrixmultiplikation $\cdot : M(m \times n, \mathbb{K}) \times \mathbb{K}^n \to \mathbb{K}^m$ wie folgt:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix}.$$

Die Größe der Matrizen muß dabei passen, d.h. ein Produkt $M(m \times n, \mathbb{K}) \times \mathbb{K}^l$ ist für $l \neq n$ nicht erklärt! Ist also $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ und $x = (x_j)_{j=1,\dots,n}$, so ist

$A \cdot x = b = (b_i)_{i=1,\dots,m}$ mit $b_i = \sum_{j=1}^{n} a_{ij} x_j$.

Beispiel 1.6 Sei $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 8 & 7 & 6 & 5 \\ 9 & 0 & 1 & 2 \end{pmatrix} \in M(3 \times 4, \mathbb{R})$ und $x = \begin{pmatrix} 1 \\ 2 \\ -1 \\ -2 \end{pmatrix} \in \mathbb{R}^4$, so ist

$$A \cdot x = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot (-1) + 4 \cdot (-2) \\ 8 \cdot 1 + 7 \cdot 2 + 6 \cdot (-1) + 5 \cdot (-2) \\ 9 \cdot 1 + 0 \cdot 2 + 1 \cdot (-1) + 2 \cdot (-2) \end{pmatrix} = \begin{pmatrix} -6 \\ 6 \\ 4 \end{pmatrix} \in \mathbb{R}^3 .$$

$\triangleleft$

Damit läßt sich ein lineares Gleichungssystem aus m Gleichungen mit n Unbekannten schreiben als $A \cdot x = b$, wobei $A \in M(m \times n, \mathbb{K})$ und $b \in \mathbb{K}^m$ gegeben sind und die Lösung $x \in \mathbb{K}^n$ gesucht ist. Wir können nun die elementaren Zeilenumformungen des LGS $A \cdot x = b$ in Matrixschreibweise formulieren:

- Typ I Vetauschen zweier Zeilen von A und der entsprechenden Einträge von b.
- Typ II Ersetzen der j-ten Zeile von A durch die Summe aus der i-ten und j-ten Zeile und Beibehalten der i-ten Zeile, sowie Ersetzen des j-ten Eintrags von b durch die Summe aus i-tem und j-tem Eintrag von b und Beibehalten des i-ten Eintrags von b.
- Typ III Multiplikation der i-ten Zeile von A und des i-ten Eintrags von b mit $\lambda \in \mathbb{K}^*$.
- Typ IV Ersetzen der j-ten Zeile von A durch die Summe aus dem λ-fachen der i-ten Zeile und der j-ten Zeile und Beibehalten der i-ten Zeile, und entsprechend für die Einträge von b.

Man sieht dabei, daß die elementaren Zeilenumformungen an der Lösung $x = (x_j)$ nichts ändern. Deshalb kann x gefahrlos weggelassen werden und die Zeilenumformungen einheitlich für die *erweiterte Koeffizientenmatrix*

$$(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

durchgeführt werden.

Beispiel 1.7 Wir wiederholen die Schritte aus Beispiel 1.4:

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 3 & 0 & -1 & 0 & 6 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 1 & 1 & 1 & 1 & -1 & 0 \end{array}\right) \xrightarrow[\mathrm{IV}_{14}(-1)]{\mathrm{I}_{23}} \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 3 & 0 & -1 & 0 & 6 \\ 0 & 1 & 0 & 1 & 0 & -2 \end{array}\right)$$

$$\xrightarrow[\mathrm{IV}_{24}(-1)]{\mathrm{IV}_{24}(-3)} \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 3 \\ 0 & 0 & -\frac{3}{4} & -1 & -\frac{3}{4} & 6 \\ 0 & 0 & -\frac{1}{4} & 1 & -\frac{1}{4} & -2 \end{array}\right) \xrightarrow{\mathrm{IV}_{34}(-\frac{1}{3})} \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & -\frac{3}{4} & -1 & -\frac{3}{4} & 6 \\ 0 & 0 & 0 & \frac{4}{3} & 0 & -4 \end{array}\right)$$

Eine solche Matrix heißt in *Zeilenstufenform.* Durch Zeilenumformung vom Typ III kann erreicht werden, daß die erste von Null verschiedene Zahl jeder Zeile zu 1 wird:

$$\xrightarrow[\mathrm{III}_4(\frac{3}{4})]{\mathrm{III}_3(-\frac{4}{3})} \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & \frac{4}{3} & 1 & -8 \\ 0 & 0 & 0 & 1 & 0 & -3 \end{array}\right)$$

Schließlich kann mit Typ IV erreicht werden, daß alle Zahlen oberhalb der ersten 1 jeder Zeile zu Null werden:

$$\xrightarrow{\mathrm{IV}_{43}(-\frac{4}{3})} \left(\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -1 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 & 0 & -3 \end{array}\right) \xrightarrow[\mathrm{IV}_{31}(-1)]{\mathrm{IV}_{32}(-\frac{1}{4})} \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -2 & 6 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 & 0 & -3 \end{array}\right)$$

Die Lösung ist damit $x_2 = 1$ und $x_4 = -3$ sowie $x_3 = -4 - x_5$ und $x_1 = 6 + 2x_5$ mit $x_5 \in \mathbb{K}$ beliebig. $\triangleleft$

Eine nützliche Definition ist also

Definition 1.8 Eine Matrix $A \in M(m \times n, \mathbb{K})$ heißt in *Zeilenstufenform*, falls gilt:

i) Es gibt eine Zahl $r \leq m$, so daß die Zeilen mit Index $i = 1, \ldots, r$ nicht identisch Null sind, während die Zeilen mit Index $i = r+1, \ldots, m$ identisch Null sind.

ii) Bezeichne $j_i := \min\{j \; : \; a_{ij} \neq 0\}$ den kleinsten Spaltenindex der Zeilen $i = 1, \ldots, r$ mit Eintrag $\neq 0$, so gilt $j_1 < j_2 < \cdots < j_r$.

Beispiel 1.9 Ein Beispiel für eine Matrix in Zeilenstufenform ist also:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 9 & 8 & 7 & 6 & 5 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\triangleleft$

Satz 1.10 *Jede Matrix $A \in M(m \times n, \mathbb{K})$ läßt sich durch endlich viele elementare Zeilenumfomungen auf spezielle Zeilenstufenform*

$$\tilde{A} = \begin{pmatrix} 0 & \dots & 0 & 1 & * & \dots & * & 0 & * & \dots & * & 0 & * & \dots & * & 0 & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & * & \dots & * & 0 & * & \dots & * & 0 & \dots \\ 0 & & & & & \dots & & & & & 0 & 1 & * & \dots & * & 0 & \\ 0 & & & & & & & & & \dots & & & & & 0 & 1 & \dots \\ \vdots & & & & & & & & & \dots & & & & & & & \ddots \end{pmatrix}$$

bringen mit $a_{ij_i} = 1$ und $a_{kj_i} = 0$ für $k \neq i$, wenn $j_i := \min\{j \ : \ a_{ij} \neq 0\}$. Ein $*$ steht für ein beliebiges Element aus $\mathbb{K}$. Jede Zeile wird nach einer Reihe von Nullen mit 1 begonnen, und zwar später als in jeder der vorangehenden Zeilen. Oberhalb und unterhalb der führenden 1 sind alle Einträge Null.

Beweis. Suche die erste Spalte j, die nicht identisch Null ist. Durch Zeilenvertauschung (Typ I) läßt sich erreichen, daß der erste Eintrag a_{1j} dieser Spalte $\neq 0$ ist. Teile die erste Zeile durch diesen Eintrag a_{1j} (Typ III). Diese neue erste Zeile wird nun festgehalten. Addiere zu jeder Zeile $i = 2, \dots m$ das $(-a_{ij})$-fache der neuen ersten Zeile (Typ IV). Dadurch wird $a_{ij} = 0$ für alle $i > 1$. Wiederhole das Verfahren für die Zeilen 2 bis m, wobei j nun der kleinste Spaltenindex mit einem $a_{ij} \neq 0$ für $i = 2, \dots, m$ ist, u.s.w. Das Ergebnis ist eine Matrix in Zeilenstufenform, wobei jede Zeile nach einer Reihe von Nullen mit 1 begonnen wird. Durch erneute Umformungen vom Typ IV werden dann (wie in Beispiel 1.7) oberhalb der führenden 1 jeder Zeile die Einträge auf 0 gebracht. □

Definition 1.11 Die Anzahl der Zeilen einer Matrix $A \in M(m \times n, \mathbb{K})$, die nach Überführung in Zeilenstufenform nicht identisch Null sind, heißt der *Rang* von A, geschrieben $\mathrm{rang}(A)$.

Wir werden später sehen, daß der Rang unabhängig von der Wahl der elementaren Zeilenumformungen ist.

Liegt die erweiterte Koeffizientenmatrix in spezieller Zeilenstufenform vor (was nach Satz 1.10 durch elementare Zeilenumformungen immer erreicht werden kann, wobei sich nach Satz 1.3 die Menge der Lösungen des LGS nicht ändert), so läßt sich die Lösung des LGS $A \cdot x = b$ mit $A \in M(m \times n, \mathbb{K})$ und $b \in \mathbb{K}^m$ wie folgt ermitteln (Gaußsches Eliminationsverfahren):

- Ist $\mathrm{rang}(A|b) \neq \mathrm{rang}(A) = r$, d.h. $b_{r+1} \neq 0$ so gibt es keine Lösung, denn die $(r+1)$-te Gleichung $0 = \sum_{j=1}^n a_{r+1,j} x_j = b_{r+1}$ ist nicht lösbar.
- Ist $\mathrm{rang}(A|b) = \mathrm{rang}(A)$, so sind die x_k mit $k \notin \{j_1, \dots, j_r\}$ im Sinne von Definition 1.8.ii), welche also Stufen der Länge ≥ 2 entsprechen, *freie Variablen.* Sie können beliebig aus $\mathbb{K}$ gewählt werden.

- Die verbleibenden x_{j_i} mit $i = 1, \ldots, r$ sind *gebundene Variablen*, deren Wert sich nach Wahl der freien x_k bestimmt zu $x_{j_i} = b_i - \sum_{k \notin \{j_1, \ldots, j_r\}} a_{ik} x_k$.
- Offenbar gilt: Für $n = r$ ist die Lösung eindeutig bestimmt, für $m = r$ ist das LGS für alle $b \in \mathbb{K}^m$ lösbar.

Mit $y = \begin{pmatrix} x_{j_1} \\ \vdots \\ x_{j_r} \end{pmatrix} \in \mathbb{K}^r$ für die gebundenen Variablen und $w = \begin{pmatrix} x_{k_1} \\ \vdots \\ x_{k_{n-r}} \end{pmatrix} \in \mathbb{K}^{n-r}$ mit $k_1 < \cdots < k_{n-r} \notin \{j_1, \ldots, j_r\}$ für die freien Variablen und Zusammenfassung der Matrixelemente $c_{il} := a_{ik_l}$ zu einer Matrix $C = (c_{il}) \in M(r \times (n-r), \mathbb{K})$ ergibt sich die Lösung von $A \cdot x = b$ (in spezieller Zeilenstufenform) zu $y = \tilde{b} - C \cdot w$ mit $w \in \mathbb{K}^{n-r}$ beliebig und $\tilde{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_r \end{pmatrix} \in \mathbb{K}^r$.

Beispiel 1.12 In Beispiel 1.9 mit $b_5 = 0$ ergibt sich nach elementaren Zeilenumformungen

$$\widetilde{(A|b)} = \left(\begin{array}{cccccc|c} 1 & 0 & \frac{11}{9} & 0 & 0 & 0 & b_1 - \frac{2}{9}b_2 - \frac{22}{9}b_3 + \frac{11}{9}b_4 \\ 0 & 1 & \frac{8}{9} & 0 & 0 & 0 & \frac{1}{9}b_2 - \frac{7}{9}b_3 + \frac{8}{9}b_4 \\ 0 & 0 & 0 & 1 & 0 & -1 & b_3 - 2b_4 \\ 0 & 0 & 0 & 0 & 1 & 2 & b_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Als Lösung ergibt sich

$$\begin{pmatrix} x_1 \\ x_2 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} b_1 - \frac{2}{9}b_2 - \frac{22}{9}b_3 + \frac{11}{9}b_4 \\ \frac{1}{9}b_2 - \frac{7}{9}b_3 + \frac{8}{9}b_4 \\ b_3 - 2b_4 \\ b_4 \end{pmatrix} - \begin{pmatrix} \frac{11}{9} & 0 \\ \frac{8}{9} & 0 \\ 0 & -1 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_3 \\ x_6 \end{pmatrix}$$

mit $x_3, x_6 \in \mathbb{R}$ beliebig. Die gebundenen Variablen sind also $x_5 = b_4 - 2x_6$, $x_4 = b_3 - 2b_4 + x_6$, $x_2 = \frac{1}{9}b_2 - \frac{7}{9}b_3 + \frac{8}{9}b_4 - \frac{8}{9}x_3$ und $x_1 = b_1 - \frac{2}{9}b_2 - \frac{22}{9}b_3 + \frac{11}{9}b_4 - \frac{11}{9}x_3$. $\triangleleft$

2 Grundbegriffe

2.1 Gruppen

Unter einer *Verknüpfung* oder *Komposition* auf einer Menge G versteht man eine Vorschrift, die zwei gegebenen Elementen $a, b \in G$ ein neues Element $a * b \in G$ zuordnet, d.h. eine Abbildung

$$* : G \times G \to G \;, \qquad * : (a, b) \mapsto a * b := *(a, b) \;.$$

Beispiele sind die Addition $* = +$ und Multiplikation $* = \cdot$ in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$. Ein weiteres wichtiges Beispiel ist die Menge $G = \mathrm{Abb}(X, X) = \{f : X \to X\}$ der Selbstabbildungen von X mit der Komposition $* = \circ$ als Verknüpfung, $f, g \in \mathrm{Abb}(X, X) \;\Rightarrow\; f \circ g \in \mathrm{Abb}(X, X)$.

Definition 2.1 Eine Menge G zusammen mit einer Verknüpfung $*$ heißt *Gruppe*, falls die folgenden Axiome erfüllt sind

(G1) $(a * b) * c = a * (b * c)$ für alle $a, b, c \in G$ (Assoziativgesetz)

(G2) Es gibt ein neutrales Element $e \in G$ mit $e * a = a$ für alle $a \in G$

(G3) Zu jedem $a \in G$ gibt es ein $a^{-1} \in G$ (das zu a inverse Element) mit $a^{-1} * a = e$

Die Gruppe heißt *kommutativ* (oder *abelsch*), falls außerdem $a * b = b * a$ für alle $a, b \in G$.

Satz 2.2 *Ist G eine Gruppe, so gilt:*

i) *Das neutrale Element ist eindeutig bestimmt, und außerdem gilt $a * e = a$ für alle $a \in G$.*

ii) *Das inverse Element a^{-1} ist für jedes $a \in G$ eindeutig bestimmt, und außerdem gilt $a * a^{-1} = e$ sowie $(a^{-1})^{-1} = a$ und $(ab)^{-1} = b^{-1} * a^{-1}$.*

iii) *Es gelten die Kürzungsregeln $a * b' = a * b \;\Rightarrow\; b = b'$ und $b' * a = b * a \;\Rightarrow\; b = b'$.*

Beweis. Sei $e \in G$ eines der neutralen Elemente und $a \in G$. Zum Inversen $a^{-1} \in G$ gibt es wieder ein Inverses $(a^{-1})^{-1} \in G$, und es gilt

$$\begin{aligned} a * a^{-1} &= e * (a * a^{-1}) = ((a^{-1})^{-1} * a^{-1}) * (a * a^{-1}) = (((a^{-1})^{-1} * a^{-1}) * a) * a^{-1} \\ &= ((a^{-1})^{-1} * (a^{-1} * a)) * a^{-1} = ((a^{-1})^{-1} * e) * a^{-1} = (a^{-1})^{-1} * (e * a^{-1}) \\ &= (a^{-1})^{-1} * a^{-1} = e \end{aligned}$$

und daraus

$$a * e = a * (a^{-1} * a) = (a * a^{-1}) * a = e * a = a\ .$$

Sei $\tilde{e} \in G$ ein weiteres neutrales Element, so gilt aus Sicht von e die Gleichung $e * \tilde{e} = \tilde{e}$ und aus Sicht von $\tilde{e}$ die Gleichung $e * \tilde{e} = e$, also $e = \tilde{e}$. Damit ist i) bewiesen und $a * a^{-1} = e$ aus ii).

Sei $\widetilde{a^{-1}}$ ein weiteres Inverses zu a, so gilt

$$\widetilde{a^{-1}} = \widetilde{a^{-1}} * e = \widetilde{a^{-1}} * (a * a^{-1}) = (\widetilde{a^{-1}} * a) * a^{-1} = e * a^{-1} = a^{-1}\ ,$$

also ist das Inverse eindeutig. Damit ist wegen $a * a^{-1} = e$ das Inverse von a^{-1} durch a selbst gegeben. Schließlich gilt

$$(b^{-1}*a^{-1})*(a*b) = ((b^{-1}*a^{-1})*a)*b = (b^{-1}*(a^{-1}*a))*b = (b^{-1}*e)*b = b^{-1}*b = e$$

so daß $b^{-1} * a^{-1}$ das Inverse zu $a * b$ ist.

Die Kürzungsregeln folgen nach Verknüpfung von links/rechts mit a^{-1} unter Verwendung der Assoziativität. $\square$

Beispiel 2.3 (für Gruppen)

i) $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ mit der Addition als Verknüpfung, mit $e = 0$ als neutralem Element und dem Negativen $a \mapsto -a$ als Inverses.

ii) $\mathbb{Q}^* := \mathbb{Q} \setminus \{0\}$ und $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$ mit der Multiplikation als Verknüpfung und $e = 1$ als neutrales Element.

iii) Die Menge $M(m \times n, \mathbb{K})$ der $(m \times n)$-Matrizen bezüglich der Addition, die wie folgt erklärt wird: Ist $A = (a_{ij})$ und $B = (b_{ij})$, so ist $A+B = (c_{ij})$ mit $c_{ij} := a_{ij} + b_{ij}$.

iv) $S(X) := \{f \in \mathrm{Abb}(X, X) \ : \ f \text{ ist bijektiv}\}$ mit der Komposition $\circ$ als Verknüpfung und der identischen Abbildung $\mathrm{id}_X : x \mapsto x$ als neutrales Element. Diese Gruppe heißt die *symmetrische Gruppe* der Menge X. Falls $X = \{1, 2, \ldots, n\}$, so schreibt man $S_n := S(\{1, 2, \ldots, n\})$. Die Elemente $\sigma \in S_n$ sind *Permutationen* der Zahlen $1, \ldots, n$.

v) Drehungen eines starren Körpers um seinen festgehaltenen Schwerpunkt im $\mathbb{R}^3$.

vi) Drehungen eines unendlichen Kristalls um den Winkel $\frac{2\pi}{n}$ um eine geeignete Achse im Kristall, so daß der Kristall in sich selbst überführt wird.

$\triangleleft$

Einige Bemerkungen:

i) Die Komposition von Abbildungen ist stets assoziativ: Sei $f : X \to Y$, $g : Y \to Z$ und $h : Z \to W$, dann gilt für $x \in X$

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

ii) Besteht X aus mehr als zwei Elementen, so ist $S(X)$ nicht abelsch. Betrachte z.B. $f, g \in S_3$ mit $f : (1,2,3) \mapsto (2,3,1)$ und $g : (1,2,3) \mapsto (2,1,3)$, dann ist $(f \circ g)(1) = 3$ und $(g \circ f)(1) = 1$.

Definition 2.4 Sei G eine Gruppe mit Verknüpfung $*$. Eine nichtleere Teilmenge $H \subset G$ heißt *Untergruppe*, wenn für alle $a, b \in H$ auch $a * b \in H$ und $a^{-1} \in H$ gilt.

Es folgt automatisch $e \in H$ (durch Wahl von $b = a^{-1}$). Ist G eine Gruppe, so sind die Teilmengen $H = G$ und $H = \{e\}$ Untergruppen. Sie heißen *triviale Untergruppen.*

Beispiele sind $\mathbb{Q}^* \subset \mathbb{R}^*$ bezüglich der Multiplikation oder G als Gruppe der Drehungen eines Körpers und H als Untergruppe der Drehungen um eine feste Achse.

2.2 Ringe

Bei Gruppen hat man eine einzige Verknüpfung, z.B. Addition oder Multiplikation. Im weiteren brauchen wir Mengen, deren Elemente mittels Addition *und* Multiplikation verknüpft werden können (z.B. die Zahlen). Solche Mengen (mit brauchbaren Eigenschaften) heißen Ringe.

Definition 2.5 Eine Menge R zusammen mit zwei Verknüpfungen

$$\begin{aligned} + :& R \times R \to R\,, \qquad && + : (a,b) \mapsto a+b \qquad && (Addition) \\ \cdot :& R \times R \to R\,, \qquad && \cdot : (a,b) \mapsto a \cdot b \qquad && (Multiplikation) \end{aligned}$$

heißt *Ring*, wenn folgendes gilt:

(R1) R zusammen mit der Addition ist eine kommutative Gruppe.

(R2) Die Multiplikation ist assoziativ.

(R3) Es gelten die Distributivgesetze

$$a \cdot (b+c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a+b) \cdot c = a \cdot c + b \cdot c$$

für alle $a, b, c \in R$.

Ein Ring heißt kommutativ, falls $a \cdot b = b \cdot a$ für alle $a, b \in R$. Ein Element $1 \in R$ heißt Einselement, wenn $1 \cdot a = a \cdot 1 = a$ für alle $a \in R$.

In (R3) ist die Konvention benutzt, daß die Multiplikation stärker bindet als die Addition.

Ist R ein Ring und $0 \in R$ das Nullelement, d.h. das neutrale Element der kommutativen Gruppe $(R, +)$, dann gilt $0 \cdot a = (0+0) \cdot a = 0 \cdot a + 0 \cdot a$, also $0 \cdot a = 0$ und analog $a \cdot 0 = 0$.

Beispiel 2.6 (für Ringe)

i) die Zahlbereiche $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$.

ii) der Ring der reell- oder kompexwertigen stetigen Funktionen über einem Intervall:

$$R = \{f : I \to \mathbb{K} \text{ stetig}\}\,, \qquad I \subset \mathbb{R}$$

mit

$$(f_1 + f_2)(x) := f_1(x) + f_2(x) \qquad \text{und} \qquad (f_1 \cdot f_2)(x) := f_1(x) \cdot f_2(x)\,.$$

iii) der Ring $\mathbb{K}[t]$ der Polynome in der Variablen t mit Koeffizienten aus $\mathbb{K}$.◁

2.3 Körper

Wenn in einem Ring auch die Multiplikation zu einer Gruppe wird (abgesehen vom Nullelement), sprechen wir von einem Körper.

Definition 2.7 Eine Menge K zusammen mit zwei Verknüpfungen

$$\begin{aligned} + :&K \times K \to K\ , \qquad + : (a,b) \mapsto a+b \qquad (Addition)\\ \cdot :&K \times K \to K\ , \qquad \cdot : (a,b) \mapsto a\cdot b \qquad (Multiplikation)\end{aligned}$$

heißt *Körper*, wenn folgendes gilt:

(K1) K zusammen mit der Addition ist eine kommutative Gruppe. Ihr neutrales Element wird mit 0 bezeichnet und das zu $a \in K$ inverse Element mit $-a$.

(K2) Sei $K^* := K\setminus\{0\}$, dann ist für $a, b \in K^*$ auch $a\cdot b \in K^*$, und K^* mit dieser Multiplikation ist eine kommutative Gruppe. Ihr neutrales Element wird mit 1 bezeichnet und das zu $a \in K^*$ inverse Element mit $a^{-1} = 1/a$. Man schreibt auch $a/b = a^{-1} \cdot b = b \cdot a^{-1}$.

(K3) Es gelten die Distributivgesetze

$$a \cdot (b + c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a + b) \cdot c = a \cdot c + b \cdot c$$

für alle $a, b, c \in K$.

Körper spielen eine zentrale Rolle in linearen Abbildungen und linearen Gleichungssystemen, weil die Division (außer durch 0) ausführbar ist.

Beispiel 2.8 (für Körper)

i) die Zahlbereiche $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.

ii) $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ wenn p eine Primzahl ist (ohne Beweis). Addition und Multiplikation in $\mathbb{Z}$ modulo p ist wie folgt zu verstehen: Ist $a, b \in \{0, 1, \ldots, p-1\}$, so sind $a+b$ und $a\cdot b$ erklärt als die eindeutig bestimmten Zahlen $s, m \in \{0, 1, \ldots, p-1\}$, für die $a+b-s$ und $a\cdot b-m$ durch p teilbar sind. $\triangleleft$

In einem Körper gilt

i) $1 \neq 0$ (da $1 \in K^*$ und $0 \notin K^*$)

ii) $a \cdot 0 = 0 \cdot a = 0$ (verwende $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$)

iii) $a \cdot b = 0 \quad \Rightarrow \quad a = 0$ oder $b = 0$ (aus (K2))

iv) $a \cdot (-b) = -(a \cdot b)$ und $(-a) \cdot (-b) = a \cdot b$

v) $a \cdot b = a \cdot c$ und $a \neq 0 \quad \Rightarrow \quad b = c$

Teil II
Vektorräume

3 Definition und Beispiele

Vektorräume sind der zentrale Gegenstand der linearen Algebra.

Definition 3.1 Sei K ein Körper. Eine Menge V zusammen mit einer inneren Verknüpfung $+ : V \times V \to V$ (der Addition) und einer äußeren Verknüpfung $\cdot : K \times V \to V$ (Multiplikation mit Skalaren) heißt *Vektorraum über K*, wenn

(V1) V zusammen mit der Addition ist eine kommutative Gruppe. (Wie üblich wird das neutrale Element wird mit $0 \in V$ bezeichnet und das zu $v \in V$ inverse Element mit $-v$.)

(V2) Für die Multiplikation mit Skalaren gilt

$$\begin{aligned} (\lambda+\mu)\cdot v &= \lambda\cdot v + \mu \cdot v\ , & \lambda\cdot(v+w) &= \lambda\cdot v+\lambda\cdot w\ , \\ \lambda\cdot(\mu\cdot v) &= (\lambda\mu)\cdot v\ , & 1\cdot v &= v\ , \end{aligned}$$

für alle $\lambda, \mu \in K$ und $v, w \in V$.

Ein Element eines Vektorraumes V heißt *Vektor*.

Die wichtigsten Fälle sind Vektorräume über $K = \mathbb{R}$ und $K = \mathbb{C}$, und schreiben oft auch $\mathbb{K}$ statt K. Wir sprechen dann von reellen bzw. komplexen Vektorräumen.

Satz 3.2 *In einem Vektorraum über K gilt*

i) $0 \cdot v = 0 \in V \quad \forall v \in V$

ii) $\lambda \cdot 0 = 0 \in V \quad \forall \lambda \in K$

iii) $(-1)\cdot v = -v \quad \forall v \in V$

iv) $\lambda \cdot v = 0 \in V \quad \Leftrightarrow \quad \lambda = 0$ *oder* $v = 0$

Beweis. Beweis von i),ii),iii) durch gleiche Methoden wie in Teil I.
iv): Ist $\lambda \cdot v = 0$, aber $\lambda \neq 0$, so gilt $v = 1 \cdot v = (\lambda^{-1}\lambda)\cdot v = \lambda^{-1}\cdot(\lambda \cdot v) = 0$. □

Beispiel 3.3 (für Vektorräume)

i) $K^n = \{x = (x_1, \dots, x_n)\ :\ x_i \in K\}$ ist ein Vektorraum mit Addition

$$(x_1,\dots,x_n) + (y_1,\dots,y_n) := (x_1+y_1,\dots,x_n+y_n)$$

und skalarer Multiplikation

$$\lambda\cdot(x_1,\dots,x_n) := (\lambda x_1,\dots,\lambda x_n)\ .$$

Wir schreiben Vektoren aus K^n oft auch als Spalten statt als Zeilen.

ii) Der Vektorraum $M(m\times n, K)$ der $(m\times n)$-Matrizen mit Einträgen aus K ist ein Vektorraum mit komponentenweiser Addition und skalarer Multiplikation: Schreiben wir $A = (a_{ij}), B = (b_{ij}) \in M(m \times n, K)$, so ist

$$A + B := (a_{ij} + b_{ij}) \ , \qquad \lambda \cdot A := (\lambda a_{ij}) \ .$$

iii) $\mathbb{C}$ kann als reeller Vektorraum aufgefaßt werden durch $\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$, $(\lambda, x + \mathrm{i}y) \mapsto \lambda x + \mathrm{i}\lambda y$.

iv) der Vektorraum der reell- oder kompexwertigen stetigen Funktionen über einem Intervall:

$$\mathcal{C}(I) = \{f : I \to \mathbb{K} \text{ stetig}\} \ , \qquad I \subset \mathbb{R}$$

mit $(\lambda_1, \lambda_2 \in \mathbb{K})$

$$(\lambda_1 f_1 + \lambda_2 f_2)(x) := \lambda_1 f_1(x) + \lambda_2 f_2(x)$$

v) der Vektorraum $\mathbb{K}[t]$ der Polynome in der Variablen t mit Koeffizienten aus $\mathbb{K}$. $\triangleleft$

Definition 3.4 Sei V ein Vektorraum. Eine Teilmenge $W \subset V$ heißt *Untervektorraum*, wenn

(UV1) $W \neq \emptyset$

(UV2) $v, w \in W \quad \Rightarrow \quad v + w \in W$ (d.h. W ist abgeschlossen bezüglich der Addition)

(UV3) $v \in W \ , \ \lambda \in K \quad \Rightarrow \quad \lambda \cdot v \in W$ (d.h. W ist abgeschlossen bezüglich der Multiplikation mit Skalaren)

Ein Untervektorraum ist automatisch ein Vektorraum. Beispiele sind

i) der Nullvektor $W = \{0\}$

ii) Vielfache $W = \{\lambda \cdot v \ : \ \lambda \in K\}$ eines ausgewählten Vektors $v \in V$, speziell jede Gerade in der Ebene durch den Nullpunkt, $W = \{(x, y) \in \mathbb{R}^2 \ : \ ax + by = 0\}$ für $a, b \in \mathbb{R}$.

Satz 3.5 *Seien W_i mit $i \in I$ (Indexmenge) jeweils Untervektorräume von V, so ist der Durchschnitt $W := \bigcap_{i\in I} W_i \subset V$ wieder ein Untervektorraum von V.*

Beweis. $0 \in W_i \ \forall i \in I$, also $0 \in W$ (damit ist W nicht leer). Seien $v, w \in W$, so sind $v, w \in W_i \ \forall i \in I$. Dann ist auch $v + w \in W_i \ \forall i \in I$ und somit $v + w \in W$. Analog ist $\lambda \cdot v \in W$. $\square$

Dagegen ist die Vereinigung von Untervektorräumen im allgemeinen nicht wieder ein Untervektorraum. Man kann eine solche Vereinigung aber zu einen Vektorraum abschließen.

4 Linearkombinationen

Seien $v_1, \ldots, v_r \in V$ (nicht notwendig verschiedene) Vektoren aus V. Ein Vektor $v \in V$ heißt *Linearkombination* von $v_1, \ldots, v_r$, wenn es $\lambda_1, \ldots, \lambda_r \in K$ gibt, so daß

$$v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = \sum_{i=1}^{r} \lambda_i \cdot v_i \ .$$

Wir bezeichnen mit

$$\operatorname{span}_K(v_1, \ldots, v_r) = \left\{ \sum_{i=1}^{r} \lambda_i \cdot v_i \ : \ \lambda_i \in K \right\}$$

den Raum aller Linearkombinationen von $v_1, \ldots, v_r$. Offenbar ist $\operatorname{span}_K(v_1, \ldots, v_r)$ ein Untervektorraum von V, er heißt der durch $v_1, \ldots, v_r$ *aufgespannte* (oder *erzeugte*) *Untervektorraum*.

Die Definition läßt sich verallgemeinern auf Untervektorräume, die von einer Familie $(v_i)_{i \in I}$ aus möglicherweise unendlich vielen Vektoren aufgespannt wird. Dann ist $\operatorname{span}_K(v_i)_{i \in I}$ definiert als die Menge aller *endlichen* Linearkombinationen, d.h. zu jedem $v \in \operatorname{span}_K(v_i)_{i \in I}$ gibt es Indizes $i_1, \ldots, i_r \in I$ und Skalare $\lambda_1, \ldots, \lambda_r \in K$, so daß $v = \sum_{j=1}^{r} \lambda_j v_{i_j}$.

Beispiel 4.1 (für Linearkombinationen)

i) Für $V = \mathbb{R}^3$ und $v_1, v_2 \in V$ ist $\operatorname{span}_{\mathbb{R}}(v_1)$ die Gerade durch 0 und v_1, falls $v_1 \neq 0$, und $\operatorname{span}_{\mathbb{R}}(v_1, v_2)$ die Ebene durch $0, v_1, v_2$, falls $v_1 \neq 0$ und $v_2 \notin \operatorname{span}_{\mathbb{R}}(v_1)$.

ii) Im Vektorraum K^n setzen wir $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, wobei die 1 an der i-ten Stelle steht. Dann ist $\operatorname{span}_K(e_i)_{i=1,\ldots,n} = K^n$. $\triangleleft$

Im weiteren wird es wichtig sein, ob sich ein Vektor $v \in V$ in eindeutiger Weise als Linearkombination von vorgegebenen Vektoren $v_1, \ldots, v_r$ darstellen läßt oder nicht.

Definition 4.2 Sei V ein Vektorraum über K. Eine Familie von endlich vielen Vektoren $v_1, \ldots, v_r$ heißt *linear unabhängig* (bzw. die Vektoren $v_1, \ldots, v_r$ heißen linear unabhängig), wenn aus $\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = 0$ für $\lambda_1, \ldots, \lambda_r \in K$ folgt $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$.

Eine beliebige Familie $(v_i)_{i \in I}$ heißt linear unabhängig, wenn jede endliche Teilfamilie linear unabhängig ist. Entsprechend heißt eine Familie $(v_i)_{i \in I}$, die nicht linear unabhängig ist, *linear abhängig*. In diesem Fall gibt es also $v_{i_1}, \ldots, v_{i_r} \neq 0$ mit $i_1, \ldots, i_r \in I$ und $\lambda_1, \ldots, \lambda_r \neq 0$ mit $\sum_{j=1}^{r} \lambda_j \cdot v_{i_j} = 0$.

Satz 4.3 *Eine Familie* $(v_1, \ldots, v_n)$ *von Vektoren* $v_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{mj} \end{pmatrix} \in K^m$ *ist genau dann linear unabhängig, wenn für die Matrix* $A = (v_{ij}) \in M(m \times n, K)$ *gilt* $\operatorname{rang}(A) = n$.

Beweis. Nach Definition des Matrixprodukts gilt

$$\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n = 0 \quad \Leftrightarrow \quad A \cdot x = 0$$

mit $x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \in K^n$. Die Familie $(v_1, \ldots, v_n)$ ist also genau dann linear unabhängig, wenn $A \cdot x = 0$ nur die triviale Lösung $x = 0$ besitzt. Das erfordert $\operatorname{rang}(A) = n$. $\square$

Beispiel 4.4 Sei z.B. $v_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$, $v_3 = \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix}$ so betrachten wir

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ 4 & 5 & 7 \end{pmatrix} \xrightarrow{IV_{12}(-2)\ ,\ IV_{13}(-4)} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & -3 & 3 \end{pmatrix} \xrightarrow{IV_{23}(-3)} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Damit ist (v_1, v_2, v_3) nicht linear unabhängig. Dagegen ist (v_1, v_2) linear unabhängig, denn Weglassen der 3. Spalte ergibt $\operatorname{rang}(v_1, v_2) = 2$. $\triangleleft$

Wir werden später sehen, daß in beliebigen endlich erzeugten Vektorräumen die lineare Unabhängigkeit stets durch Lösen linearer Gleichungssysteme ermittelt werden kann.

Satz 4.5 *Für eine Familie* $(v_i)_{i \in I}$ *sind folgende Bedingungen äquivalent:*

i) $(v_i)_{i \in I}$ *ist linear unabhängig*

ii) *Jeder Vektor* $v \in \operatorname{span}_K(v_i)_{i \in I}$ *läßt sich in eindeutiger Weise als Linearkombination von Vektoren aus* $(v_i)_{i \in I}$ *darstellen.*

Beweis. i)$\Rightarrow$ii). Sei $v \in \operatorname{span}_K(v_i)_{i \in I}$ auf zwei verschiedene Arten als Linearkombination darstellbar,

$$v = \sum_{i \in I} \lambda_i \cdot v_i = \sum_{i' \in I} \mu_{i'} \cdot v_{i'} ,$$

wobei nur endlich viele Skalare $\lambda_i, \mu_{i'}$ ungleich Null sind. Es gibt also eine endliche Teilmenge $J \subset I$ (Vereinigung der Indizes i, i', für die $\lambda_i \neq 0$ oder $\mu_{i'} \neq 0$), so daß

$$\sum_{j \in J} (\lambda_j - \mu_j) \cdot v_j = 0 .$$

Da nach Voraussetzung i) jede endliche Teilfamilie von $(v_i)_{i\in I}$ linear unabhängig ist, ist $\lambda_j = \mu_j$ für alle $j \in J$ und weiter für alle $i \in I$ (auf dem Komplement $I \setminus J$ ist $\lambda_i = \mu_i = 0$).

ii)$\Rightarrow$i). Der Nullvektor läßt sich als Linearkombination von $(v_i)_{i\in I}$ darstellen, in der sämtliche Skalare Null sind. Ist die Linearkombination eindeutig, so ist $(v_i)_{i\in I}$ linear unabhängig. $\square$

5 Erzeugendensystem, Basis und Dimension

Definition 5.1 Eine Familie $\mathcal{B} = (v_i)_{i\in I}$ von Vektoren $v_i \in V$ heißt *Erzeugendensystem* von V, wenn $V = \mathrm{span}_K(v_i)_{i\in I}$, wenn also jeder Vektor $v \in V$ eine endliche Linearkombination von $(v_i)_{i\in I}$ ist.

Ein Erzeugendensystem $\mathcal{B} = (v_i)_{i\in I}$ heißt *Basis* von V, wenn $\mathcal{B}$ linear unabhängig ist.

Der Vektorraum V heißt *endlich erzeugt*, falls es ein endliches Erzeugendensystem $\mathcal{B} = (v_1, \ldots, v_r)$ von V gibt.

Beispiel 5.2 (für Erzeugendensysteme und Basen)

i) Im Vektorraum K^n ist $\mathcal{B} = (e_1, \ldots, e_n)$, mit $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, ein Erzeugendensystem, denn jeder Vektor $x = (x_1, \ldots, x_n)$ läßt sich schreiben als $x = \sum_{i=1}^n x_i e_i$ mit $x_i \in K$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardbasis* des K^n.

ii) Im Vektorraum $M(m \times n, K)$ ist $\mathcal{B} = (E_{ij})_{\substack{i=1,\ldots m \\ j=1,\ldots,n}}$ mit

$$E_{ij} = \begin{pmatrix} & & & 0 & & & \\ & \mathrm{O} & & \vdots & & \mathrm{O} & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & \mathrm{O} & & \vdots & & \mathrm{O} & \\ & & & 0 & & & \end{pmatrix},$$

wobei die 1 im Schnittpunkt der i-ten Zeile mit der j-ten Spalte steht, ein Erzeugendensystem. Jede Matrix $A = (a_{ij})$ läßt sich darstellen als $A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} E_{ij}$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardmatrixbasis*.

iii) In $\mathbb{C}$, aufgefaßt als reeller Vektorraum, ist $\mathcal{B} = (1, \mathrm{i})$ eine Basis, denn jede komplexe Zahl kann als $z = x \cdot 1 + y \cdot \mathrm{i}$ mit $x, y \in \mathbb{R}$ geschrieben werden.

Fassen wir $\mathbb{C} = \mathbb{C}^1$ dagegen als komplexen Vektorraum auf, dann ist $\mathcal{B} = (1, \mathrm{i})$ zwar ein Erzeugendensystem, aber keine Basis mehr, denn 1 und i sind nicht mehr linear unabhängig über $\mathbb{C}$: $1 \cdot 1 + \mathrm{i} \cdot \mathrm{i} = 0$.

iv) Sei $P_n(x)$ der Vektorraum der reellen oder komplexen Polynome vom Grad $\leq n$ in x. Dann ist $\mathcal{B} = (1, x, x^2, \dots, x^n)$ eine Basis von $P_n(x)$: Jedes Polynom $p(x)$ vom Grad $\leq n$ läßt sich schreiben als $p(x) = \sum_{k=0}^n a_k x^k$ mit $a_k \in \mathbb{K}$ (also ist $\mathcal{B}$ Erzeugendensystem). Aus $p(x) = 0$ folgt nach dem Identitätssatz für Polynome (Satz 9.5 aus dem 1. Semster) $a_k = 0$. also ist $\mathcal{B}$ linear unabhängig. $\triangleleft$

Ist $\mathcal{B} = (v_1, \dots, v_n)$ ein endliches Erzeugendensystem bzw. eine endliche Basis, dann ist jede Permutation (Umordnung) der Vektoren v_i wieder ein endliches Erzeugendensystem bzw. eine endliche Basis. Z.B. ist $\mathcal{B} = (v_1, v_2, v_3)$ mit $v_1 = (1, 0, 0)$, $v_2 = (0, 0, 1)$, $v_3 = (0, 1, 0)$ eine Basis von K^3, aber es ist übersichtlicher, statt $\mathcal{B}$ mit der durch Umordnung erhaltenen Standardbasis (v_1, v_3, v_2) zu arbeiten.

Für die Definition einer Basis gibt es mehrere äquivalente Möglichkeiten. Wir beschränken uns zunächst auf endlich erzeugte Vektorräume:

Satz 5.3 *Für eine Familie $\mathcal{B} = (v_1, \dots, v_n)$ von Vektoren $v_i \in V$, $V \neq \{0\}$, sind folgende Bedingungen äquivalent:*

i) *$\mathcal{B}$ ist ein linear unabhängiges Erzeugendensystem von V (also eine Basis).*

ii) *$\mathcal{B}$ ist ein Erzeugendensystem von V, aber $(v_1, \dots, v_{r-1}, v_{r+1}, \dots, v_n)$ ist für jedes $1 \leq r \leq n$ kein Erzeugendensystem mehr.*

iii) *Zu jedem $v \in V$ gibt es eindeutig bestimmte $\lambda_1, \dots, \lambda_n \in K$ mit $v = \sum_{i=1}^n \lambda_i \cdot v_i$ (Eindeutigkeit der Zerlegung von v nach der Basis).*

iv) *$\mathcal{B}$ ist linear unabhängig, während $(v_1, \dots, v_n, v)$ linear abhängig ist für alle $v \in V$.*

Beweis. i)$\Rightarrow$ii). Wäre $(v_1, \dots, v_{r-1}, v_{r+1}, \dots, v_n)$ ein Erzeugendensystem, dann gibt es $\lambda_1, \dots, \lambda_{r-1}, \lambda_{r+1}, \dots, \lambda_n \in K$ mit $v_r = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + \sum_{j=r+1}^n \lambda_j \cdot v_j$. Damit ist $0 = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + (-1) \cdot v_r + \sum_{j=r+1}^n \lambda_j \cdot v_j$, so daß $\mathcal{B}$ nicht linear unabhängig wäre.

ii)$\Rightarrow$iii). Angenommen, es gibt ein $v \in V$ und für dieses zwei verschiedene Darstellungen

$$v = \sum_{i=1}^n \lambda_i \cdot v_i \qquad \text{und} \qquad v = \sum_{i=1}^n \mu_i \cdot v_i$$

mit $\lambda_r \neq \mu_r$ für mindestens ein r. (Es gibt mindestens eine Darstellung, da $\mathcal{B}$ Erzeugendensystem.) Subtraktion beider Gleichungen und Division durch $(\lambda_r - \mu_r)$ ergibt

$$v_r = \sum_{i=1}^{r-1} \frac{\lambda_i - \mu_i}{\mu_r - \lambda_r} \cdot v_i + \sum_{j=r+1}^n \frac{\lambda_j - \mu_j}{\mu_r - \lambda_r} \cdot v_j$$

Das würde bedeuten, daß $(v_1, \dots, v_{r-1}, v_{r+1}, \dots, v_n)$ ein Erzeugendensystem wäre, denn in einer Linearkombination ließe sich v_r durch die v_i, v_j mit $i, j \neq r$ ausdrücken, im Widerspruch zu ii).

iii)⇒iv). Nach Satz 4.5 ist $\mathcal{B}$ linear unabhängig. Andererseits gibt es für jedes $v \in V$ Skalare $\lambda_1, \dots, \lambda_n \in K$ mit $0 = \sum_{i=1}^n \lambda_i \cdot v_i + (-1) \cdot v$, so daß $(v_1, \dots, v_n, v)$ nicht linear unabhängig ist.

iv)⇒i). Da $(v_1, \dots, v_n, v)$ für alle $v \in V$ linear abhängig ist, gibt es $\lambda_1, \dots, \lambda_n, \lambda \in K$, welche nicht alle gleich 0 sind, mit $0 = \sum_{i=1}^n \lambda_i \cdot v_i + \lambda \cdot v$. Wäre $\lambda = 0$, so auch $\lambda_i = 0$ für alle $i = 1, \dots, n$, da $\mathcal{B}$ linear unabhängig ist. Also ist $\lambda \neq 0$ und somit

$$v = \sum_{i=1}^n \frac{-\lambda_i}{\lambda} \cdot v_i \ .$$

Damit ist $\mathcal{B}$ ein Erzeugendensystem. □

Die wichtigste Eigenschaft einer Basis ist wohl die Eindeutigkeit der Zerlegung eines gegebenen Vektors nach einer Basis des Vektorraums. Ist eine Basis fixiert, dann kann man an Stelle des Vektors $v \in V$ mit einer Folge von Zahlen $(\lambda_1, \dots, \lambda_n)$ mit $\lambda_i \in K$ arbeiten. Diese Zahlen λ_i heißen die *Koordinaten* von $v = \sum_{i=1}^n \lambda_i \cdot v_i$ bezüglich der Basis $(v_1, \dots, v_n)$. Die Beziehung zwischen einem Vektor und seinen Koordinaten wird im Abschnitt über lineare Abildungen weiter untersucht.

Satz 5.4 *Ist V nicht endlich erzeugt, dann gibt es eine unendliche linear unabhängige Familie von Vektoren.*

Beweis. Durch Induktion nach der Anzahl n linear unabhängiger Vektoren von V: Seien $(v_1, \dots, v_n)$ linear unabhängig. Wäre $(v_1, \dots, v_n, v_{n+1})$ linear abhängig für jedes $v_{n+1} \in V$, so wäre $(v_1, \dots, v_n)$ Erzeugendensystem, Widerspruch. □

Satz 5.5 (Austauschlemma von Steinitz) *Es sei $(v_1, \dots, v_n)$ Erzeugendensystem (bzw. eine Basis) eines Vektorraums V über K und $w = \sum_{i=1}^n \lambda_i v_i$ mit $\lambda_r \neq 0$ für ein $r \in \{1, \dots, n\}$. Dann ist auch $(v_1, \dots, v_{r-1}, w, v_{r+1}, \dots, v_n)$ ein Erzeugendensystem (bzw. eine Basis) von V.*

Beweis. Unter den Voraussetzungen gilt

$$v_r = \frac{1}{\lambda_r} \cdot w + \sum_{i=1}^{r-1} \frac{-\lambda_i}{\lambda_r} \cdot v_i + \sum_{j=r+1}^{n} \frac{-\lambda_j}{\lambda_r} \cdot v_j \ .$$

Sei $v \in V$ ein beliebiger Vektor, dann gibt es $\mu_1, \dots, \mu_n \in K$ mit $v = \sum_{i=1}^n \mu_i \cdot v_i$. Einsetzen von v_r liefert

$$v = \sum_{i=1}^{r-1} \left(\mu_i - \frac{\mu_r \lambda_i}{\lambda_r}\right) \cdot v_i + \frac{\mu_r}{\lambda_r} \cdot w + \sum_{j=r+1}^{n} \left(\mu_j - \frac{\mu_r \lambda_j}{\lambda_r}\right) \cdot v_j \ .$$

Damit ist $(v_1, \dots, v_{r-1}, w, v_{r+1}, \dots, v_n)$ ein Erzeugendensystem.

Ist $\mathcal{B} = (v_1, \dots, v_n)$ Basis, dann betrachten wir zur Untersuchung der linearen Unabhängigkeit

$$0 = \sum_{i=1}^{r-1} \mu_i \cdot v_i + \mu \cdot w + \sum_{j=r+1}^{n} \mu_j \cdot v_j \ .$$

Einsetzen von w liefert

$$0 = \sum_{i=1}^{r-1} (\mu_i + \mu\lambda_i) \cdot v_i + \mu\lambda_r \cdot v_r + \sum_{j=r+1}^{n} (\mu_j + \mu\lambda_j) \cdot v_j \ .$$

Da $\mathcal{B}$ linear unabhängig, folgt $\mu_i + \mu\lambda_i = 0$ für $i \neq r$ und $\mu\lambda_r = 0$, also $\mu = 0$ und dann $\mu_i = 0$. $\square$

Beispiel 5.6 Seien $v_1 = \begin{pmatrix} 2 \\ 5 \\ 3 \\ 7 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ 1 \\ -3 \\ 6 \end{pmatrix}$, $v_3 = \begin{pmatrix} 1 \\ 4 \\ 0 \\ 1 \end{pmatrix}$, $v_4 = \begin{pmatrix} 2 \\ -1 \\ 5 \\ 1 \end{pmatrix} \in \mathbb{R}^4$.
Ist die Familie $\mathcal{B} = (v_1, v_2, v_3, v_4)$ linear unabhängig?

Nach Satz 4.3 kann die Frage durch Bestimmen des Rangs der Matrix

$$A = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 5 & 1 & 4 & -1 \\ 3 & -3 & 0 & 5 \\ 7 & 6 & 1 & 1 \end{pmatrix}$$

beantwortet werden. Wir gehen hier einen anderen Weg über mehrfache Anwendung des Austauschlemmas. Wir dürfen die r-te Spalte v_r durch eine Linearkombination $w = \sum_{j=1}^{4} \lambda_j v_j$ ersetzen, wenn $\lambda_r \neq 0$. Das ist aber gerade eine *elementare Spaltenumformung vom Typ IV*, definiert in Analogie zu den Zeilenumformungen aus Abschnitt 1. Vertauschen von Vektoren (Typ I) erhält ebenfalls das Erzeugendensystem. Spaltenumformungen liefern also

$$A \xrightarrow{\substack{IV_{12}(-\frac{1}{2}) \\ IV_{13}(-\frac{1}{2}) \\ IV_{14}(-1)}} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 5 & -\frac{3}{2} & \frac{3}{2} & -6 \\ 3 & -\frac{9}{2} & -\frac{3}{2} & 2 \\ 7 & \frac{5}{2} & -\frac{5}{2} & -6 \end{pmatrix} \xrightarrow{\substack{IV_{23}(1) \\ IV_{24}(-4)}} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 5 & -\frac{3}{2} & 0 & 0 \\ 3 & -\frac{9}{2} & -6 & 20 \\ 7 & \frac{5}{2} & 0 & -16 \end{pmatrix}$$

$$\xrightarrow{\substack{III_2(2) \\ IV_{34}(\frac{10}{3})}} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 5 & -3 & 0 & 0 \\ 3 & -9 & -6 & 0 \\ 7 & 5 & 0 & -16 \end{pmatrix}$$

Damit ist $\mathcal{B}$ linear unabhängig. Durch weitere Spaltenumformungen läßt sich $\mathcal{B}$ auch in die Standardbasis überführen. $\triangleleft$

Würde sich bei diesem Verfahren der Nullvektor ergeben, so kann dieser aus dem Erzeugendensystem weggelassen werden. Allgemein gilt:

Satz 5.7 (Basisauswahlsatz) *Aus jedem endlichen Erzeugendensystem läßt sich eine Basis auswählen. Insbesondere hat jeder endlich erzeugte Vektorraum eine Basis.*

Beweis. Man nehme aus einem endlichen Erzeugendensystem einzelne Vektoren weg, so daß die reduzierte Familie immer noch ein Erzeugendensystem des Vektorraums bleibt. Ist das nicht mehr möglich, so liegt eine Basis vor. □

Wir zeigen nun, daß alle Basen eines endlich erzeugten Vektorraums aus der gleichen Anzahl an Vektoren bestehen.

Satz 5.8 (Austauschsatz) *In einem Vektorraum V über K sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis und $(w_1, \dots, w_r)$ eine linear unabhängige Familie von Vektoren aus V. Dann ist $r \leq n$, und es gibt paarweise verschiedene Indizes $i_1, \dots, i_r \in \{1, 2, \dots, n\}$, so daß man nach Austausch von v_{i_j} durch w_j für alle $1 \leq j \leq r$ wieder eine Basis von V erhält. Nach Permutation der Indizes (Umnumerierung) erreicht man, daß $\mathcal{B}^* = (w_1, \dots, w_r, v_{r+1}, \dots, v_n)$ eine Basis von V ist.*

Beweis. Wir tauschen schrittweise die Basisvektoren aus. Im ersten Schritt finden wir ein $1 \leq s \leq n$, so daß $(v_1, \dots, v_{s-1}, w_1, v_{s+1}, \dots, v_n)$ eine Basis ist. Nun nennen wir $v_1 \mapsto v_s$ und ordnen durch Vertauschen von $v_s \leftrightarrow w_1$ die Basis um, so daß $(w_1, v_2, \dots, v_n)$ eine Basis ist.

Dann gibt es $\mu_1, \lambda_2, \dots, \lambda_n \in K$ mit $w_2 = \mu_1 \cdot w_1 + \sum_{i=2}^n \lambda_i \cdot v_i$. Mindestens eines der λ_i ist von 0 verschieden, denn sonst wäre $w_2 = \mu_1 \cdot w_1$, und die Familie $(w_1, \dots, w_r)$ wäre nicht linear unabhängig. Es läßt sich deshalb ein v_s mit $2 \leq s \leq n$ durch w_2 austauschen, so daß nach Umnumerierung $(w_1, w_2, v_3 \dots, v_n)$ eine Basis von V ist.

Ist $r \leq n$, so führt die wiederholte Anwendung dieses Verfahrens auf eine Basis $(w_1, \dots, w_r, v_{r+1}, \dots, v_n)$. Wäre $r > n$, dann erhalten wir im n-ten Schritt eine Basis $(w_1, \dots, w_n)$ von V. Dann ist w_{n+1} nach dieser Basis zerlegbar, so daß $(w_1, \dots, w_r)$ für $r > n$ nicht mehr linear unabhängig wäre. □

Satz 5.9 *Sei V ein endlich erzeugter Vektorraum. Dann besteht jede Basis von V aus der gleichen Anzahl von Vektoren.*

Beweis. Seien zwei Basen $\mathcal{B}_1 = (v_1, \dots, v_n)$ und $\mathcal{B}_2 = (w_1, \dots, w_r)$ von V gegeben. Ist $n < r$, dann wäre nach Satz 5.8 $\mathcal{B}_2$ nicht linear unabhängig, also keine Basis. Ist $n > r$, dann wäre nach Satz 5.8 $\mathcal{B}_1$ nicht linear unabhängig, also keine Basis. Folglich gilt $n = r$. □

Definition 5.10 In einem Vektorraum V über K heißt die durch

$$\dim_K V := \begin{cases} 0 & \text{falls } V = \{0\}, \\ n & \text{falls } V \text{ endlich erzeugt ist und eine aus } n \text{ Vektoren} \\ & \text{bestehende Basis besitzt,} \\ \infty & \text{falls } V \text{ nicht endlich erzeugt ist} \end{cases}$$

definierte Zahl die *Dimension* von V.

Offenbar ist $\dim_K(K^n) = n$ und $\dim_{\mathbb{R}}(\mathbb{C}) = 2$, aber $\dim_{\mathbb{C}}(\mathbb{C}) = 1$. Wir schreiben auch $\dim(V)$ statt $\dim_K(V)$, wenn der Körper klar ist. Die Dimension ist eine entscheidende Charakterisierung eines Vektorraums.

Satz 5.11 *Ist $W \subset V$ Untervektorraum eines endlich erzeugten Vektorraums V, so ist auch W endlich erzeugt, und es gilt*

i) $\dim(W) \leq \dim(V)$,

ii) *aus* $\dim(W) = \dim(V)$ *folgt* $W = V$.

Beweis. i) Wäre W nicht endlich erzeugt, dann gäbe es nach Satz 5.4 eine unendliche linear unabhängige Familie, die auch in V linear unabhängig wäre, Widerspruch. Ebenso kann es nach dem Austauschsatz höchstens $n := \dim_K(V)$ linear unabhängige Vektoren in W geben.

ii) Sei $\dim(W) = \dim(V) = n$ und $(w_1, \ldots, w_n)$ eine Basis von W. Wäre $V \neq W$, so gibt es ein $v \in V \setminus W$, das keine Linearkombination von $(w_1, \ldots, w_n)$ ist. Damit wäre $(w_1, \ldots, w_n, v)$ linear unabhängig, Widerspruch. □

Dieser Satz ist sehr hilfreich. Wenn wir schon wissen, daß ein Vektorraum V die Dimension n hat und n linear unabhängige Vektoren $v_1, \ldots, v_n \in V$ gegeben sind, dann ist $W = \operatorname{span}(v_1, \ldots, v_n)$ ein Untervektorraum, in dem $(v_1, \ldots, v_n)$ eine Basis ist. Aus der Gleichheit der Dimensionen folgt $W = V$, also ist $(v_1, \ldots, v_n)$ ein Erzeugendensystem von V. Es ist wesentlich einfacher, die lineare Unabhängigkeit mit dem Gaußschen Algorithmus für lineare Gleichungssysteme zu überprüfen als die Eigenschaft des Erzeugendensystems. Daraus wird die Bedeutung der Dimension ersichtlich.

Ein weiteres nützliches Hilfsmittel ist:

Satz 5.12 (Basisergänzungssatz) *In einem endlich erzeugten Vektorraum V sei eine Familie $(w_1, \ldots, w_r)$ von linear unabhängigen Vektoren gegeben. Dann läßt sich diese Familie zu einer Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ von V ergänzen.*

Beweis. Man nehme irgendeine Basis von V und wende den Austauschsatz an. □

In nicht endlich erzeugten (also unendlich-dimensionalen) Vektorräumen gibt es einige Besonderheiten. Zwar gilt unter Benutzung des *Auswahlaxioms*:

Satz 5.13 *Jeder Vektorraum besitzt eine Basis.*

Jedoch läßt sich in nicht endlich erzeugten Vektorräumen eine Basis im allgemeinen nicht angeben. Im Vektorraum der reellen Zahlenfolgen $V = \{(a_1, a_2, \dots) \ : \ a_i \in \mathbb{R}\}$ sind die Vektoren

$$\{(1,0,0,0,\dots),(0,1,0,0,\dots),(0,0,1,0,\dots),\dots\}$$

zwar linear unabhängig, aber sie bilden keine Basis (auch kein Erzeugendensystem), denn z.B. wird die Folge $(1,1,1,\dots)$ nicht davon *als endliche Linearkombination* erzeugt.

Wir werden später unitäre und euklidische Vektorräume einführen, die auch unendlich-dimensional sein können. Dort gibt es sogenannte Orthonormalbasen, so daß sich jeder Vektor als eindeutige unendliche (aber konvergente) Linearkombination darstellen läßt. Die Konvergenz erfordert dann Methoden der Analysis. Unter Verwendung des Auswahlaxioms existiert zwar auch dort eine Basis, so daß jeder Vektor eine endliche Linearkombination ist, diese Basis ist aber viel größer (überabzählbar) als die sehr natürliche Orthonormalbasis.

Beispiel 5.14 Es sei

$$V = \mathcal{C}^1_{per}([0,2\pi]) := \{f : [0,2\pi] \to \mathbb{R} \text{ stetig differenzierbar}, \ f(0) = f(2\pi)\}$$

der Vektorraum der 2π-periodischen stetig differenzierbaren Funktionen. Man kann zeigen, daß die Familie $(v_n, w_n)_{n\in\mathbb{N}}$ mit $v_n(x) = \cos(nx)$ und $w_n = \sin((n+1)x)$ eine Orthonormalbasis ist, d.h. jede Funktion $f \in V$ eindeutig geschrieben werden kann als unendliche *konvergente* Linearkombination

$$f(x) = \lambda_0 + \sum_{n=1}^{\infty} \big(\lambda_n \cdot \cos(nx) + \mu_n \cdot \sin(nx)\big)\ , \qquad \lambda_n, \mu_n \in \mathbb{R}\ ,\ x \in [0,2\pi]\ . \ \triangleleft$$

6 Summen von Vektorräumen

Sei $(v_1, \dots, v_n)$ eine Basis von V, dann erzeugt jeder Basisvektor v_i einen eindimensionalen Untervektorraum $W_i = Kv_i := \operatorname{span}_K(v_i)$. Da jeder Vektor $v \in V$ als Summe $v = \sum_{i=1}^n w_i$ mit $w_i \in Kv_i$ dargestellt werden kann, ist es sinnvoll, V selbst als Summe der Untervektorräume aufzufassen. Diese Summe läßt sich dann verallgemeinern auf höherdimensionale Untervektorräume:

Definition 6.1 Sei V ein Vektorraum und $W_1, \dots, W_r \subset V$ Untervektorräume. Dann heißt

$$W = W_1 + \dots + W_r := \{v \in V \ : \ \text{es gibt } v_j \in W_j \text{ mit } v = v_1 + \dots + v_r\ \}$$

die *Summe* von $W_1, \dots, W_r$.

Offenbar ist W wieder ein Untervektorraum von V, und es gilt $\dim(\sum_{i=1}^r W_i) \leq \sum_{i=1}^r \dim(W_i)$. Für $r = 2$ können wir mehr zeigen:

Satz 6.2 *Für endlich-dimensionale Untervektorräume $W_1, W_2 \subset V$ gilt $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$.*

Beweis. $W_1 \cap W_2$ ist ein endlich-dimensionaler Untervektorraum von V. Man nehme eine Basis $(v_1, \dots, v_m)$ von $W_1 \cap W_2$ und ergänze sie zu Basen $(v_1, \dots, v_m, w_1, \dots, w_r)$ von W_1 und $(v_1, \dots, v_m, w'_1, \dots, w'_s)$ von W_2. Damit wird $W_1 + W_2$ von $\mathcal{B} = (v_1, \dots, v_m, w_1, \dots, w_r, w'_1, \dots, w'_s)$ erzeugt. Wir zeigen: $\mathcal{B}$ ist linear unabhängig, also Basis. Dazu sei

$$0 = \underbrace{\sum_{i=1}^m \lambda_i v_i + \sum_{j=1}^r \mu_j w_j}_{=v \in W_1} + \underbrace{\sum_{k=1}^s \mu'_k w_k}_{=-v \in W_2} .$$

Das bedeutet $v \in W_1 \cap W_2$, also $\mu_j = 0$ und $\mu'_k = 0$ und somit $v = 0$. Aus der linearen Unabhängigkeit von $(v_1, \dots, v_m)$ folgt schließlich auch $\lambda_i = 0$. Damit ist ist $\dim(W_1 + W_2) = m + r + s$ sowie $\dim(W_1) = m + r$ und $\dim(W_2) = m + s$. □

Beispiel 6.3 Sei $W_1 = \operatorname{span}(v_1, v_2, v_3)$ und $W_2 = \operatorname{span}(v_4, v_5)$ mit $v_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$, $v_3 = \begin{pmatrix} 1 \\ 3 \\ 7 \end{pmatrix}$ sowie $v_4 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, $v_5 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Nach Beispiel Beispiel 4.4 gilt $\dim(W_1) = 2$, insbesondere $W_1 = \operatorname{span}(v_1, v_2)$. Die Dimension $\dim(W_2) = 2$ ist klar. Setzen wir $A = (v_1, v_2, v_4, v_5)$, so suchen wir zur Bestimmung von $W_1 \cap W_2$ die Lösungsmenge $x \in \mathbb{R}^4$ des LGS $Ax = 0$ durch elementare Zeilenumformungen. Die rechte Seite $b = 0$ kann weggelassen werden:

$$\begin{pmatrix} 1 & 2 & 2 & 0 \\ 2 & 3 & 1 & 1 \\ 4 & 5 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & 2 & 0 \\ 0 & -1 & -3 & 1 \\ 0 & -3 & -8 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & 2 & 0 \\ 0 & -1 & -3 & 1 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$
$$\mapsto \begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & -1 & 0 & -5 \\ 0 & 0 & 1 & -2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 & -6 \\ 0 & -1 & 0 & -5 \\ 0 & 0 & 1 & -2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$

Damit ergibt sich $\begin{pmatrix} x_1 \\ x_2 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ -5 \\ 2 \end{pmatrix} \cdot x_5$ und $x_5 \in \mathbb{R}$ beliebig. Das bedeutet $\sum_{i=1,2,4,5} x_i v_i = x_5 \cdot (6v_1 - 5v_2 + 2v_4 + v_5) = 0$, d.h. Linearkombinationen von

$$\underbrace{5v_2 - 6v_1}_{\in W_1} = \underbrace{2v_4 + v_5}_{\in W_2} = \begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}$$

spannen $W_1 \cap W_2$ auf. Somit ist $\dim(W_1 \cap W_2) = 1$ und nach der Dimensionsformel (Satz 6.2) $\dim(W_1 + W_2) = 3$. $\triangleleft$

Aus Satz 6.2 erhält man schrittweise eine allgemeine Dimensionsformel für Summen von Vektorräumen, z.B.

$$\begin{aligned}\dim\Big(\sum_{i=1}^{r} W_i\Big) &= \dim\Big(\sum_{i=1}^{r-1} W_i\Big) + \dim(W_r) - \dim\Big(W_r \cap \sum_{i=1}^{r-1} W_i\Big) = \dots \\ &= \sum_{i=1}^{r} \dim(W_i) - \sum_{l=2}^{r} \dim\Big(W_l \cap \sum_{i=1}^{l-1} W_i\Big) .\end{aligned}$$

Die Dimensionen der entstehenden Durchschnitte $(W_l \cap (W_1 + \dots + W_{l-1})$ für $l = 2, \dots, r$ lassen sich im allgemeinen nicht besser ausdrücken.

Es gilt $\dim(W_1 \cap W_2) = 0$ genau dann, wenn $W_1 \cap W_2 = \{0\}$, d.h. wenn beliebige von 0 verschiedene Vektoren $w_1 \in W_1$ und $w_2 \in W_2$ linear unabhängig sind.

Definition 6.4 Sei V ein Vektorraum und $W_1, \dots, W_k$ Untervektorräume von V. Ein Vektorraum W heißt *direkte Summe* der Untervektorräume $W_1, \dots, W_k$, bezeichnet mit

$$W = W_1 \oplus W_2 \oplus \dots \oplus W_k ,$$

wenn gilt:

(DS1) $W = W_1 + \dots + W_k$

(DS2) Von Null verschiedene Vektoren $w_1 \in W_1, \dots, w_k \in W_k$ sind linear unabhängig in V.

Beispiel 6.5 Sei $V = \mathbb{R}^3$ und $v_1, v_2, v_3 \in V$ linear unabhängig, dann ist z.B. $V = \mathbb{R}v_1 \oplus \mathbb{R}v_2 \oplus \mathbb{R}v_3$ oder $V = \mathrm{span}_{\mathbb{R}}(v_1, v_3) \oplus \mathbb{R}v_2$. $\triangleleft$

Satz 6.6 *Für Untervektorräume $W_1, \dots, W_k$ eines endlich-dimensionalen Vektorraums V sind folgende Bedingungen äquivalent:*

i) $V = W_1 \oplus W_2 \oplus \dots \oplus W_k$

ii) *Ist für jedes $i = 1, \dots, k$ eine Basis $\mathcal{B}^{(i)} := (v_1^{(i)}, \dots v_{r_i}^{(i)})$ von W_i gegeben, dann ist $\mathcal{B} = (v_1^{(1)}, \dots v_{r_1}^{(1)}, \dots v_1^{(k)}, \dots v_{r_k}^{(k)})$ eine Basis von V.*

iii) *Jeder Vektor $v \in V$ ist eindeutig darstellbar als $v = w_1 + \dots + w_k$ mit $w_j \in W_j$.*

iv) $V = W_1 + \dots + W_k$ *und* $\dim(W_1) + \dots + \dim(W_k) = \dim(V)$.

Beweis. *i)* $\Rightarrow$ *ii)* Nach (DS1) ist V durch $\mathcal{B}$ erzeugt. Nach (DS2) sind in der Linarkombination

$$0 = \underbrace{\sum_{i=1}^{r_1} \lambda_i^{(1)} v_i^{(1)}}_{w_1} + \cdots + \underbrace{\sum_{i=1}^{r_k} \lambda_i^{(k)} v_i^{(k)}}_{w_k}$$

alle $w_i = 0$, und damit $\lambda_i^{(j)} = 0$. Also ist $\mathcal{B}$ Basis von V.

ii) $\Rightarrow$ *iii)* Ist $\mathcal{B}$ Basis, dann gibt es eine eindeutige Zerlegung $v = \underbrace{\sum_{i=1}^{r_1} \lambda_i^{(1)} v_i^{(1)}}_{w_1} + \cdots + \underbrace{\sum_{i=1}^{r_k} \lambda_i^{(k)} v_i^{(k)}}_{w_k}$.

iii) $\Rightarrow$ *i)* Offenbar ist $V = \sum_{j=1}^{k} W_j$, und $0 = w_1 + \cdots + w_k$ hat die eindeutige Lösung $w_i = 0$, d.h. (DS2).

ii) $\Leftrightarrow$ *iv)* folgt aus der Definition der Dimension. □

Satz 6.7 *Sei V eine endlich-dimensionaler Vektorraum und $W \subset V$ ein Untervektorraum. Dann gibt es einen (nicht eindeutig bestimmten) Untervektorraum $W' \subset V$ mit $V = W \oplus W'$. (W' heißt* direkter Summand *von V zu W.)*

Beweis. Ist $\dim(W) = \dim(V)$, so ist $V = W \oplus \{0\}$. Sei nun $\dim(W) < \dim(V)$, $(v_1, \ldots, v_n)$ eine Basis von V und $(w_1, \ldots, w_r)$ eine Basis von W. Nach dem Basisaustauschsatz gibt es nach möglicher Umnumerierung der Indizes eine Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ von V. Also ist $W' := \mathrm{span}(v_{r+1}, \ldots, v_n)$ ein direkter Summand. □

Teil III
Lineare Abbildungen

7 Definition und Beispiele

Definition 7.1 Seien V, W Vektorräume über K. Eine Abbildung $F : V \to W$ heißt *linear* (genauer: K-linear) oder *Homomorphismus* von Vektorräumen (über K), wenn

(L) $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2)$ für alle $v_1, v_2 \in V$ und $\lambda_1, \lambda_2 \in K$.

Eine lineare Abbildung $F : V \to W$ heißt

- *Isomorphismus*, wenn F bijektiv ist,
- *Endomorphismus*, wenn $W = V$ ist,
- *Automorphismus*, wenn F bijektiv und $W = V$ ist.

Beispiel 7.2

i) *Skalentransformationen (Dilatationen).* Sei V ein reeller Vektorraum und $c > 0$, dann definiert $F_c : v \mapsto c \cdot v$ einen Automorphismus von V (eine lineare Abbildung $F_c : V \to V$, die bijektiv ist mit $F_c^{-1} = F_{\frac{1}{c}} : v \mapsto \frac{1}{c} \cdot v$).

ii) *Drehungen in der Ebene.* Sei $V = \mathbb{R}^2$ und F_α die Drehung um den Nullpunkt mit dem Winkel α,

$$F_\alpha : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto F_\alpha \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} x\cos\alpha - y\sin\alpha \\ x\sin\alpha + y\cos\alpha \end{pmatrix} .$$

F ist ebenfalls ein Automorphismus mit $F_\alpha^{-1} = F_{-\alpha}$.

iii) *Matrizen.* Sei $V = W = \mathbb{R}^2$ und

$$F_a : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} .$$

Dann ist $F_a : \mathbb{R}^2 \to \mathbb{R}^2$ eine lineare Abbildung, welche durch die Matrix $a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M(2 \times 2, \mathbb{R})$ parametrisiert wird. Die beiden Beispiele i) und ii) sind Spezialfälle mit $a = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$ (für $V = \mathbb{R}^2$) und $a = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. Für allgemeines a ist F ein Endomorphismus. Es wird wichtig sein zu untersuchen, wann F bijektiv ist, also ein Automorphismus ist.

iv) *Projektionen der Ebene auf eine Gerade durch den Nullpunkt.* Sei $V = \mathbb{R}^2$ und (v_1, v_2) eine Basis, z.B. $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Dann definiert die eindeutige Zerlegung von v nach der Basis $v = \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2$ zwei Projektionen

$$\begin{aligned} F_1 &: \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \mapsto \lambda_1 \cdot v_1 \ , \\ F_2 &: \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \mapsto \lambda_2 \cdot v_2 \ . \end{aligned}$$

Diese Projektionen F_1, F_2 sind lineare Abbildungen.

v) In Analogie zu Beispiel iii) definieren wir $F_a : \mathbb{R}^2 \to \mathbb{R}^1$ durch $F_a : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto a_{11}x_1 + a_{12}x_2$. Dann ist F_a eine lineare Abbildung, die durch die Matrix $a = \begin{pmatrix} a_{11} & a_{12} \end{pmatrix} \in M(1 \times 2, \mathbb{R})$ parametrisiert wird. Beispiel iv) für $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ist ein Spezialfall mit $F_1 \Leftrightarrow a = (1\ 0)$ und $F_2 \Leftrightarrow a = (0\ 1)$.

vi) *Projektionen auf einen Untervektorraum.* In Verallgemeinerung von Beispiel iv) sei V ein n-dimensionaler Vektorraum, $W \subset V$ ein Untervektorraum und $V = W \oplus W'$. Dann ist die Projektion $F_W : V \to W$ eine lineare Abbildung, die wie folgt erhalten wird. Man nehme eine Basis $(w_1, \dots, w_r)$ von W und eine Basis $(w'_{r+1}, \dots, w'_n)$ von W' (also $n - r$ Vektoren von V, so daß $(w_1, \dots, w_r, w'_{r+1}, \dots, w'_n)$ eine Basis von V ist) und zerlege $v \in V$ nach dieser Basis, $v = \sum_{i=1}^r \lambda_i \cdot v_i + \sum_{j=r+1}^n \mu_j \cdot w'_j$. Dann ist $F_W : v \mapsto \sum_{i=1}^r \lambda_i \cdot v_i$. Diese Abbildung F_W ist unabhängig vom direkten Summanden W' (welcher nicht eindeutig war) und unabhängig von der Basis von W.

vii) *Differentiation.* Für ein offenes Intervall $I =]a, b[\subset \mathbb{R}$ seien $V = \mathcal{C}^1(I)$ und $W = \mathcal{C}(I)$ die (unendlich-dimensionalen) Vektorräume der einmal stetig differenzierbaren bzw. der stetigen reellwertigen Funktionen auf I, mit $(\lambda_1 \cdot f_1 + \lambda_2 \cdot f_2)(x) := \lambda_1 \cdot f_1(x) + \lambda_2 \cdot f_2(x)$ für $x \in]a, b[$. Dann ist das Differential $D : V \to W$ mit $D : f \mapsto f'$ eine lineare Abbildung. Ist $V = \mathcal{C}^\infty(I)$ der Vektorraum der beliebig oft differenzierbaren Funktionen über I, dann ist das Differential $D : V \to V$ ein Endomorphismus.

viii) *Integration.* Für ein abgeschlossenes Intervall $I = [a, b] \subset \mathbb{R}$ definiert das Integral $\int : f \mapsto \int_a^b f(x)\, dx$ eine lineare Abbildung $\int : \mathcal{C}(I) \to \mathbb{R}$.

Diese sehr unterschiedlichen Beispiele zeigen, daß lineare Abbildungen häufig auftreten. Kenntnisse der allgemeinen Eigenschaften von linearen Abbildungen und von Methoden zu ihrer Untersuchung sind deshalb sehr wichtig.

8 Bild und Kern einer linearen Abbildung

Satz 8.1 *Sei $F : V \to W$ eine lineare Abbildung. Dann gilt*

i) $F(0) = 0$, $F(v - w) = F(v) - F(w)$, $F(\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n) = \lambda_1 \cdot F(v_1) + \cdots + \lambda_n \cdot F(v_n)$.

ii) *Ist die Familie $(v_i)_{i\in I}$ linear abhängig in V, so ist die Familie $(F(v_i))_{i\in I}$ linear abhängig in W.*

iii) *Sind $V' \subset V$ und $W' \subset W$ Untervektorräume, so ist das Bild $F(V') \subset W$ Untervektorraum von W und das Urbild $F^{-1}(W') \subset V$ Untervektorraum von V.*

iv) $\dim(F(V)) \leq \dim(V)$.

v) *Ist F ein Isomorphismus, so ist auch das Urbild $F^{-1} : W \to V$ eine lineare Abbildung (und damit ein Isomorphismus).*

Beweis. i) ist klar.

ii) Sei $0 = \sum_{i\in I} \lambda_i \cdot v_i$, wobei mindestens einer und insgesamt endlich viele Skalare λ_i ungleich 0 sind. Dann ist auch $0 = F(0) = \sum_{i\in I} \lambda_i \cdot F(v_i)$.

iii) $0 \in V'$ und $0 = F(0) \in F(V')$, also $F(V') \neq \varnothing$. Sind $w_1, w_2 \in F(V')$, dann gibt es v_1, v_2 mit $w_1 = F(v_1)$ und $w_2 = F(v_2)$. Also $\lambda_1 \cdot w_1 + \lambda_2 \cdot w_2 = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) = F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) \in F(V')$, und $F(V')$ ist Untervektorraum.

Sind $v_1, v_2 \in F^{-1}(W')$, so ist $F(v_1), F(v_2) \in W'$ und auch $\lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) \in W'$, da W' Untervektorraum. Also ist $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) \in W'$ und damit $\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \in F^{-1}(W')$, und $F^{-1}(W')$ ist Untervektorraum.

iv) Sind $F(v_1), \ldots, F(v_n)$ linear unabhängig, so sind auch $v_1, \ldots, v_n$ linear unabhängig, denn das Gegenteil führt wegen ii) zu einem Widerspruch.

v) Seien $w_1, w_2 \in W$ und $v_1, v_2 \in V$ mit $F(v_1) = w_1$ und $F(v_2) = w_2$. Dann ist $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot w_1 + \lambda_2 \cdot w_2$. Anwenden von F^{-1} liefert $F^{-1}(\lambda_1 \cdot w_1 + \lambda_2 \cdot w_2) = \lambda_1 \cdot F^{-1}(w_1) + \lambda_2 \cdot F^{-1}(w_2)$. □

Seien V, W Vektorräume über K, dann bezeichnen wir mit

$$\mathrm{Hom}_K(V, W) := \{F : V \to W \ : \ F \text{ ist } K\text{-linear}\}$$

die Menge aller K-linearen Abbildungen (aller Homomorphismen) von V nach W. Die Menge $\mathrm{Hom}_K(V, W)$ ist ein Vektorraum mit Verknüpfungen

$$(\lambda_1 \cdot F_1 + \lambda_2 \cdot F_2)(v) := \lambda_1 \cdot F_1(v) + \lambda_2 \cdot F_2(v) \ .$$

Insbesondere ist der Nullvektor die lineare Abbildung $0 : v \mapsto 0$ für alle $v \in V$, und die negative Abbildung ist $-F : v \mapsto -F(v)$. Ist der Körper klar, dann schreiben wir $\mathrm{Hom}(V, W)$ statt $\mathrm{Hom}_K(V, W)$.

Ist $V = W$, so schreiben wir $\mathrm{End}(V) := \mathrm{Hom}(V, V)$, und $\mathrm{End}(V)$ ist wieder ein Vektorraum. Entsprechend bezeichnet

$$\mathrm{Aut}_K(V) := \{F : V \to V \ : \ F \text{ ist } K\text{-linear und bijektiv}\}$$

die Menge der bijektiven linearen Abbildungen von V nach V. Jedoch ist $\mathrm{Aut}(V)$ kein Vektorraum, denn für $F \in \mathrm{Aut}(V)$ ist $0 \cdot F = 0 \notin \mathrm{Aut}(V)$.

Lineare Abbildungen lassen sich ganz analog wie allgemeine Abbildungen hintereinander ausführen.

Satz 8.2 *Sind U, V, W Vektorräume und $G : U \to V$ sowie $F : V \to W$ lineare Abbildung, dann ist auch die Komposition $F \circ G : U \to W$ eine lineare Abbildung.*

Beweis.

$$\begin{aligned}(F \circ G)(\lambda_1 \cdot u_1 + \lambda_2 \cdot u_2) &:= F(G(\lambda_1 \cdot u_1 + \lambda_2 \cdot u_2)) \\ &= F(\lambda_1 \cdot G(u_1) + \lambda_2 \cdot G(u_2)) \\ &= \lambda_1 \cdot F(G(u_1)) + \lambda_2 \cdot F(G(u_2)) \\ &= \lambda_1 \cdot (F \circ G)(u_1) + \lambda_2 \cdot (F \circ G)(u_2) \ .\end{aligned}$$ □

Ohne Beweis erwähnen wir:

Satz 8.3 *Ist V ein Vektorraum, so ist $\mathrm{End}(V)$ ein Ring mit der Komposition als Multiplikation, und $\mathrm{Aut}(V)$ ist eine Gruppe mit der Komposition als Verknüpfung.*

Definition 8.4 Ist $F : V \to W$ eine lineare Abbildung, so heißt

- $\mathrm{im}(F) := F(V) \subset W$ das *Bild* von F,
- $\ker(F) := F^{-1}(0) \subset V$ der *Kern* von F.

Satz 8.5 *Sei $F : V \to W$ linear. Dann gilt:*

i) $\mathrm{im}(F) \subset W$ *und* $\ker(F) \subset V$ *sind Untervektorräume.*

ii) *F surjektiv* $\Leftrightarrow$ $\mathrm{im}(F) = W$

iii) *F injektiv* $\Leftrightarrow$ $\ker(F) = \{0\}$

iv) *Ist F injektiv und sind $v_1, \ldots, v_n$ linear unabhängig, so sind auch $F(v_1), \ldots, F(v_n)$ linear unabhängig.*

Beweis. i) ist in Satz 8.1.iii) bewiesen und ii) ist die Definition der Surjektivität.

iii) $0 \in \ker(F)$. Ist F injektiv, so folgt aus $F(v) = 0 = F(0)$, daß $v = 0$. Gibt es umgekehrt $v_1 \neq v_2 \in V$ mit $F(v_1) = F(v_2)$, so wäre $F(v_1 - v_2) = 0$, also $0 \neq v_1 - v_2 \in \ker(F)$. Widerspruch

iv) Sei $\lambda_1 \cdot F(v_1) + \cdots + \lambda_n \cdot F(v_n) = 0$, dann gilt $0 = F(\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n)$ nach Linearität und $\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n = 0$ nach iii). Also ist $\lambda_i = 0$. □

Definition 8.6 Ist $F : V \to W$ linear, so heißt die Zahl $\operatorname{rang}(F) := \dim(\operatorname{im}(F))$ der *Rang* der Abbildung F.

Der Rang ist eine wichtige Charakterisierung einer linearen Abbildung.

Satz 8.7 *Ist V ein endlich-dimensionaler Vektorraum und $F : V \to W$ eine lineare Abbildung, so gilt*

$$\dim(V) = \dim(\operatorname{im}(F)) + \dim(\ker(F)) .$$

Beweis. Wir wissen $\dim(\operatorname{im}(F)) \leq \dim(V)$ nach Satz 8.1.iv) und $\dim(\ker(F)) \leq \dim(V)$ nach Satz 8.5.i). Seien also $(v_1, \ldots, v_k)$ eine Basis von $\ker(F)$ und $(w_1, \ldots, w_r)$ eine Basis von $\operatorname{im}(F)$. Wir wählen $u_1, \ldots, u_r \in V$ mit $F(u_i) = w_i$ für alle $1 \leq i \leq r$. Für $v \in V$ gibt es eine eindeutige Darstellung $F(v) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r$. Für die so bestimmten λ_i konstruieren wir $v' := \lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r$. Dann gilt $F(v - v') = 0$, also $v - v' \in \ker(F)$. Es existieren also eindeutig bestimmte $\mu_1, \ldots, \mu_k$ mit $v - v' = \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k$. Also gilt

$$v = \lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r + \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k .$$

Damit ist V durch die Familie $\mathcal{B} = (u_1, \ldots, u_r, v_1, \ldots, v_k,)$ erzeugt. Zum Beweis der linearen Unabhängigkeit sei $\lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r + \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k = 0$. Anwenden von F ergibt $F(\lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r = 0$, und damit $\lambda_i = 0$. Das liefert $\mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k = 0$ und damit $\mu_j = 0$. Damit ist $\mathcal{B}$ eine Basis von V aus $\dim(\operatorname{im}(F)) + \dim(\ker(F))$ Vektoren. □

Satz 8.8 *Zwischen endlich-dimensionalen Vektorräumen V, W gibt es genau dann einen Isomorphismus, wenn $\dim(V) = \dim(W)$.*

Beweis. F ist Isomorphismus, wenn linear, injektiv und surjektiv. F injektiv heißt $\dim(\ker(F)) = 0$ und F surjektiv heißt $\operatorname{im}(F) = W$. Aus dem vorigen Satz folgt dann die Behauptung. □

Satz 8.9 *Ist $\dim(V) = \dim(W) < \infty$ und ist F linear, dann sind für F Injektivität, Surjektivität und Bijektivität äquivalent.*

Beweis. injektiv$\Rightarrow$ surjektiv: F injektiv, dann $\ker(F) = \{0\}$ und $\dim(V) = \dim(\operatorname{im}(F))$. Also $\dim(\operatorname{im}(F)) = \dim(W)$ und dann $\operatorname{im}(F) = W$.

surjektiv$\Rightarrow$injektiv: F surjektiv, dann $\dim(\operatorname{im}(F)) = \dim(W) = \dim(V)$. Also $\dim(\ker(F)) = 0$ und $\ker(F) = \{0\}$, somit F injektiv. □

Satz 8.10 (Faktorisierungssatz) *Sei $F : V \to W$ linear und $\mathcal{B} = (u_1, \ldots, u_r, v_1, \ldots, v_k)$ eine Basis von V mit $\ker(F) = \operatorname{span}(v_1, \ldots, v_k)$. Mit $U := \operatorname{span}(u_1, \ldots, u_r)$ gilt:*

i) $V = U \oplus \ker(F)$

ii) *Die Einschränkung* $F\big|_U : U \to \mathrm{im}(F)$ *ist ein Isomorphismus.*

iii) *Sei* $P_U : V = U \oplus \ker(F) \to U$ *die Projektion auf den ersten Summanden, definiert für* $v = u + v' \in V$ *mit* $u \in U$ *und* $v' \in \ker(F)$ *durch* $P_U(v) := u$*, so gilt* $F = F\big|_U \circ P_U$*.*

Beweis. i) ist klar, denn U und $\ker(F)$ sind linear unabhängig.

ii) $\ker(F\big|_U) = \ker(F) \cap U = \{0\}$, also ist $F\big|_U : U \to \mathrm{im}(F)$ injektiv, außerdem surjektiv und damit bijektiv.

iii) Für $v = u + v'$ mit $u \in U$ und $v' \in \ker F$ ist $F(v) = F(u) + F(v') = F(u) = F\big|_U(u) = F\big|_U(P(v))$. □

Zu beachten ist, daß U in der direkten Summe $V = U \oplus \ker(F)$ nicht eindeutig definiert ist. Die $F\big|_U$ sind also, je nach Wahl von U, verschiedene Isomorphismen.

Ist $A \in M(m \times n, K)$, so wird durch $F_A(x) := A \cdot x$ für $x \in K^n$ eine lineare Abbildung $F_A : K^n \to K^m$ definiert. Dazu gibt es folgende Umkehrung:

Satz 8.11 *Sei* $F : K^n \to K^m$ *eine lineare Abbildung. Dann existiert genau eine Matrix* $A_F \in M(m \times n, K)$ *mit* $F(x) = A_F \cdot x$ *für alle* $x \in K^n$*. Für diese Matrix* A_F *gilt* $A_F := (F(e_1), \ldots, F(e_n))$*, d.h. die* j*-te Spalte von* A_F *ist der Vektor* $F(e_j) \in K^m$*, wobei* $e_j \in K^n$ *der* j*-te Standardbasisvektor ist.*

Damit definiert die Zuordnung $F \mapsto A_F$ *eine bijektive Abbildung* $M_m^n : \mathrm{Hom}(K^n, K^m) \to M(m \times n, K)$*.*

Beweis. Sei $A = (a_{ij}) \in M(m \times n, K)$ und $a_j := \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in K^m$ die j-te Spalte von A. Dann gilt für alle $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$

$$A \cdot x = \begin{pmatrix} x_1 a_{11} + x_2 a_{12} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + x_2 a_{m2} + \cdots + x_n a_{mn} \end{pmatrix} = \sum_{j=1}^{n} x_j a_j \ .$$

Sei nun $F : K^n \to K^m$ linear und $A_F := (F(e_1), \ldots, F(e_n))$, dann gilt $A_F \cdot x = \sum_{j=1}^n x_j F(e_j) = F(\sum_{j=1}^n x_j e_j) = F(x)$.

Sei $A' = (a'_1, \ldots, a'_n) \in M(m \times n, K)$ mit $A' \neq A$. Dann gibt es ein $r \in \{1, \ldots, n\}$, so daß für die r-te Spalte gilt $a'_r \neq a_r$. Für $x = e_r$ gilt dann $A' \cdot x = a'_r \neq a_r = F(x)$. Damit ist A_F durch F eindeutig bestimmt. □

Die wichtigen Begriffe von Bild, Kern und Rang einer linearen Abbildung übertragen sich somit auf Matrizen.

Definition 8.12 Für $A \in M(m \times n, K)$ heißt

- $\ker(A) = \{x \in K^n \ : \ A \cdot x = 0 \in K^m\} \subset K^n$ der *Kern* von A,
- $\mathrm{im}(A) = \{b \in K^m \ : \ \text{es gibt ein } x \in K^n \text{ mit } A \cdot x = b\} \subset K^m$ das *Bild* von A,
- $\mathrm{rang}(A) = \dim(\mathrm{im}(A))$ der Rang von A.

Damit ist $\ker(A)$ die Lösungsmenge des homogenen LGS $A \cdot x = 0$. Der Faktorisierungssatz liefert dann eine Methode zur Bestimmung von $\mathrm{im}(A)$: Ist $(v_1, \ldots, v_{n-r})$ eine Basis von $\ker A \subset K^n$, so ergänze man sie zu einer Basis $(v_1, \ldots, v_{n-r}, u_1, \ldots, u_r)$ von K^n. Dann ist $(A \cdot u_1, \ldots, A \cdot u_r)$ eine Basis von $\mathrm{im}(A)$.

Beispiel 8.13 Durch $F((w, x, y, z)) = (w + x - y, x + y - z, w + 2x - z)$ für $w, x, y, z \in \mathbb{R}$ werde eine lineare Abbildung $F : \mathbb{R}^4 \to \mathbb{R}^3$ definiert. Daraus lesen wir ab:

$$F(e_1) = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad F(e_2) = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad F(e_3) = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad F(e_4) = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}.$$

Folglich wird F durch die Matrix $A_F = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 1 & 2 & 0 & -1 \end{pmatrix} \in M(3 \times 4, \mathbb{R})$ dargestellt. Zur Berechnung des Kerns als Lösung von $A \cdot x = 0$ führen wir folgende elementare Zeilenumformungen durch:

$$A_F \overset{IV_{13}(-1), IV_{23}(-1)}{\longrightarrow} \begin{pmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{IV_{21}(-1)}{\longrightarrow} \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Daraus lesen wir als Lösung $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$ ab mit $x_3, x_4 \in \mathbb{R}$ beliebig. Folglich ist $\dim(\ker A) = 2$, und eine Basis von $\ker A$ ist (v_1, v_2) mit

$$v_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Wir können sie z.B. zur Basis (v_1, v_2, e_1, e_2) von $\mathbb{R}^4$ ergänzen. Somit ist $(F(e_1), F(e_2)$ eine Basis von $\mathrm{im}(A)$. ◁

Wir zeigen im nächsten Abschnitt, daß die hier gegebene Definition des Rangs einer Matrix mit Definition 1.11 übereinstimmt.

9 Zur Theorie linearer Gleichungssysteme

Definition 9.1 Eine Teilmenge $X \subset V$ eines K-Vektorraums V heißt *affiner Unterraum*, falls es ein $v \in V$ und einen Untervektorraum $W \subset V$ gibt, so daß

$$X = v + W := \{x \in V \ : \ \text{es gibt ein } w \in W \text{ mit } x = v + w\} \ .$$

Beispiel 9.2 Ist $V = \mathbb{R}^2$ die reelle Zahlenebene, dann sind die eindimensionalen Untervektorräume $W \subset V$ die Geraden durch den Nullpunkt. Ein aus der Geraden W durch 0 hervorgehender affiner Unterraum $X = v + W$ ist dann die Parallele zu W durch v. ◁

Satz 9.3 *Ist $X = v + W$ ein affiner Unterraum, dann gilt*

i) *Für beliebiges $v' \in X$ ist $X = v' + W$.*

ii) *Ist $v' \in V$ und $W' \subset V$ ein Untervektorraum mit $v' + W' = v + W$, so folgt $W = W'$ und $v' - v \in W$.*

Beweis. i) Wegen $v' \in X$ gibt es ein $w' \in W$ mit $v' = v + w'$. Sei $x = v + w \in X$ ein beliebiges Element, so schreibt sich $x = v' + w - w'$. Also ist $X \subset v' + W$. Durch Vertauschen von v, v' folgt $v' + W = X$.

ii) Seien $x, x' \in X = v + W$, dann ist $x - x' \in W$. Gäbe es eine zweite Darstellung $X = v' + W'$, dann ist $x - x' \in W'$ für alle x, x', also $W = W'$. Dann folgt $v' - v \in W$. □

Da also in einem affinen Unterraum $X = v + W$ der Untervektorraum eindeutig ist, definieren wir die Dimension $\dim(X) := \dim(W)$ wenn $X = v + W$.

Wir untersuchen den *Lösungsraum*

$$\text{Lös}(A, b) := \{x \in K^n \ : \ Ax = b\}$$

eines linearen Gleichungssystems $Ax = b$ zu gegebener Matrix $A \in M(m \times n, K)$ und gegebenem Vektor $b \in K^m$. Ist $F : K^n \to K^m$ die durch $A \in M(m \times n, K)$ definierte lineare Abbildung, $F(x) = A \cdot x$, so ist $\text{Lös}(A, b) = F^{-1}(b)$ und insbesondere $\text{Lös}(A, 0) = F^{-1}(0) = \ker(F)$. Damit gilt stets $0 \in \text{Lös}(A, 0)$, und diese Lösung heißt die *triviale Lösung* des homogenen Systems. Geht die erweiterte Koeffizientenmatrix $(A|b)$ durch elementare Zeilenumformungen in eine erweiterte Koeffizientenmatrix $(\tilde{A}|\tilde{b})$ über, dann gilt $\text{Lös}(A, b) = \text{Lös}(\tilde{A}, \tilde{b})$ nach Satz 1.3.

Satz 9.4 *Gegeben sei das lineare Gleichungssystem $Ax = b$ aus m Gleichungen mit n Unbekannten. Ist* $\text{rang}(A) = r$, *dann gilt für die Lösungsräume:*

i) $\text{Lös}(A, 0) \subset K^n$ *ist ein Untervektorraum der Dimension $n - r$.*

ii) $\text{Lös}(A, b) \subset K^n$ *ist entweder leer oder ein affiner Raum der Dimension $n - r$. Ist $v \in$* $\text{Lös}(A, b)$ *eine beliebige Lösung, dann gilt* $\text{Lös}(A, b) = v + \text{Lös}(A, 0)$.

Beweis. i) Sei $F : K^n \to K^m$ die durch $F(x) = A \cdot x$ definierte lineare Abbildung. Nach Satz 8.5 ist $\text{Lös}(A,0) = \ker(F)$ ein Untervektorraum. Nach Satz 8.7 ist seine Dimension $\dim(\ker(F)) = \dim(K^n) - \dim(\text{im}(F)) = n - r$ wegen $\text{rang}(A) := \dim(\text{im}(F))$.

ii) Seien $v, v' \in \text{Lös}(A,b)$ zwei Lösungen des linearen Gleichungssystems, also $Av = b$ und $Av' = b$. Dann ist $A(v - v') = 0$, also $v - v' \in \text{Lös}(A,0)$. □

Der Satz besagt, daß man eine allgemeine Lösung des inhomogenen linearen Gleichungssystems erhält, indem man zu einer *speziellen* Lösung des inhomogenen Systems die *allgemeine* Lösung des homogenen Systems addiert. Sei also $(w_{r+1}, \dots, w_n)$ eine Basis von $\text{Lös}(A,0)$ (wenn $r < n$) und $v \in \text{Lös}(A,b)$ eine beliebige spezielle Lösung des inhomogenen Systems, so ist

$$\text{Lös}(A,b) = v + Kw_{r+1} + \dots + Kw_n \qquad \text{mit } Kw := \text{span}_K(w) \;.$$

Satz 9.5 *Der Lösungsraum eines linearen Gleichungssystems $Ax = b$ ist genau dann nicht leer, wenn* $\text{rang}(A,b) = \text{rang}(A)$.

Beweis. Die Matrizen $A \in M(m \times n, K)$ und $(A,b) \in M(m \times (n+1), K)$ beschreiben lineare Abbildungen $A : K^n \to K^m$ bzw. $A' : K^{n+1} \to K^m$ mit $A(e_j) = A'(e_j) = \sum_{i=1}^m a_{ij} e_i$ für $1 \leq j \leq n$ und $A'(e_{n+1}) = b$. Damit ist $\text{im}(A) \subset \text{im}(A')$, also $\text{rang}(A) \leq \text{rang}(A')$, sowie $b \in \text{im}(A')$. Ist nun $\text{rang}(A) = \text{rang}(A')$, dann ist $\text{im}(A) = \text{im}(A')$ und damit $b \in \text{im}(A)$, das Gleichungssystem ist also lösbar.

Ist $\text{rang}(A) < \text{rang}(A')$, dann muß es Vektoren $v \in \text{im}(A')$ geben, die nicht in $\text{im}(A)$ liegen. Da $\text{span}(A'(e_j))_{j=1,\dots,n} \subset \text{im}(A)$, bleiben nur die b enthaltenden Linearkombinationen, welche nicht in $\text{im}(A)$ liegen. Das bedeutet $b \notin \text{im}(A)$. □

Insbesondere ist der Lösungsraum $\text{Lös}(A,b)$ für $A \in M(m \times n, K)$ nichtleer, wenn $\text{rang}(A) = m$, da dann die durch A beschriebene lineare Abbildung surjektiv ist. Das bedeutet, daß $Ax = b$ für jedes $b \in K^m$ lösbar ist. In diesem Fall heißt das lineare Gleichungssystem *universell lösbar*. Ist $\text{rang}(A) < m$, so gibt es nicht für alle $b \in K^m$ eine Lösung. Es ist nur für jene $b \in K^m$ lösbar, für die die $\text{rang}(A) = \text{rang}(A,b)$ gilt.

Wir sagen, das lineare Gleichungssystem $Ax = b$ ist *eindeutig lösbar*, wenn $\text{Lös}(A,b)$ nur aus einem Element besteht.

Satz 9.6 *Das lineare Gleichungssystem $Ax = b$ für $A \in M(m \times n, K)$ und $b \in K^m$ ist genau dann eindeutig lösbar, wenn* $\text{rang}(A) = \text{rang}(A,b) = n$.

Beweis. Nach Satz 9.5 gibt es eine Lösung, und nach Satz 9.4 ist diese eindeutig. □

Satz 9.7 *Die Definitionen 8.12 und 1.11 für den Rang einer Matrix $A \in M(m \times n, K)$ stimmen überein.*

Beweis. Es sei $\tilde{A} \in M(m \times n, K)$ eine aus A durch elementare Zeilenumformungen hervorgehende Matrix. Der Lösungsraum eines LGS bleibt unverändert bei elementaren Zeilenumformungen, d.h. es gilt $\text{Lös}(A, 0) = \text{Lös}(\tilde{A}, 0)$ und damit $r := \dim(\text{im}(A)) = \dim(\text{im}(\tilde{A}))$ nach der Dimensionsformel aus Satz 8.7. Sei $\tilde{r}$ die Anzahl der Nicht-Null-Zeilen von $\tilde{A}$ und $V_{\tilde{r}} = \text{span}(e_1, \dots, e_{\tilde{r}}) \subset K^m$ der durch die ersten $\tilde{r}$ Basisvektoren $e_1, \dots, e_{\tilde{r}} \in K^m$ aufgespannte Untervektorraum. Dann gilt $\tilde{A} \cdot x \in V_{\tilde{r}}$ für beliebige $x \in K^n$. Andererseits ist nach dem Gaußschen Eliminationsverfahren die Gleichung $\tilde{A} \cdot x = \tilde{b}$ für beliebige $\tilde{b} \in V_{\tilde{r}}$ lösbar, d.h. $\text{im}(\tilde{A}) = V_{\tilde{r}}$ und somit $\tilde{r} = r$. $\square$

10 Die darstellende Matrix einer linearen Abbildung

Im folgenden geht es darum, die Äquivalenz zwischen Matrizen und linearen Abbildungen auf beliebige endlich-dimensionale Vektorräume zu verallgemeinern. Der erste Schritt ist

Satz 10.1 *Gegeben seien endlich-dimensionale Vektorräume V und W und Vektoren $v_1, \dots, v_r \in V$ und $w_1, \dots, w_r \in W$. Dann gilt:*

i) *Sind die $v_1, \dots, v_r$ linear unabhängig, so gibt es mindestens eine lineare Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $1 \leq i \leq r$.*

ii) *Ist $(v_1, \dots, v_r)$ eine Basis, so gibt es genau eine lineare Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $1 \leq i \leq r$. Für diese Abbildung F gilt*

   - $\text{im}(F) = \text{span}(w_1, \dots, w_r)$
   - *F injektiv $\Leftrightarrow$ $(w_1, \dots, w_r)$ ist linear unabhängig.*

Beweis. ii) Ist $v = \lambda_1 \cdot v_1 + \dots + \lambda_r \cdot v_r \in V$, so definieren wir $F(v) := \lambda_1 \cdot w_1 + \dots + \lambda_r \cdot w_r$. Da die λ_i eindeutig bestimmt sind, ist auch $F(v)$ eindeutig. Zu zeigen ist aber, daß F wirklich linear ist. Sei $v' = \mu_1 \cdot v_1 + \dots + \mu_r \cdot v_r$ ein weiterer Vektor. Dann gilt

$$\begin{aligned} \lambda \cdot v + \mu \cdot v' &= \sum_{i=1}^{r} (\lambda\lambda_i + \mu\mu_i) \cdot v_i \qquad \text{und} \\ F(\lambda \cdot v + \mu \cdot v') &= \sum_{i=1}^{r} (\lambda\lambda_i + \mu\mu_i) \cdot w_i = \lambda\Big(\sum_{i=1}^{r} \lambda_i \cdot w_i\Big) + \mu\Big(\sum_{i=1}^{r} \mu_i \cdot w_i\Big) \\ &= \lambda \cdot F(v) + \mu \cdot F(v') . \end{aligned}$$

Klar ist, daß $\text{im}(F) \subset \text{span}(w_1, \dots, w_r)$. Andererseits ist $w = \lambda_1 \cdot w_1 + \dots + \lambda_r w_r = F(\lambda_1 \cdot v_1 + \dots + \lambda_r \cdot v_r)$ und damit $\text{span}(w_1, \dots, w_r) \subset \text{im}(F)$.

Sei $(w_1, \dots, w_r)$ linear abhängig, so gibt es mindestens ein $\lambda_i \neq 0$, so daß $0 = \lambda_1 w_1 + \dots \lambda_r w_r = F(\lambda_1 \cdot v_1 + \dots + \lambda_r \cdot v_r)$. Da $(v_1, \dots, v_r)$ eine Basis ist,

ist $\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r \neq 0$, und damit ist F nicht injektiv. Das bedeutet F injektiv $\Rightarrow$ $(w_1, \ldots, w_r)$ linear unabhängig. Ist umgekehrt $(w_1, \ldots, w_r)$ linear unabhängig, dann betrachten wir $F(v) = 0$ mit $v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r$, also $F(\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r$, was zu $\lambda_i = 0$ und damit $v = 0$ führt. Also ist $\ker(F) = \{0\}$ und F ist injektiv.

i) Ist $(v_1, \ldots, v_r)$ linear unabhängig, aber keine Basis, dann können wir eine Basis $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_n)$ von V finden. Für diese geben wir beliebige Vektoren $w_{r+1}, \ldots, w_n \in W$ vor und finden für jede Wahl von $w_{r+1}, \ldots, w_n$ eine lineare Abbildung F mit $F(v_i) = w_i$. □

Satz 10.2 *Sei $\mathcal{B} = (w_1, \ldots, w_n)$ Basis eines Vektorraums W.*

i) *Es gibt genau einen Isomorphismus $\Phi_{\mathcal{B}} : K^n \to W$ mit $\Phi_{\mathcal{B}}(e_i) = w_i$ für alle $1 \leq i \leq n$, wobei $(e_i)_{i=1,\ldots,n}$ die Standardbasis ist.*

ii) *Ist $W = K^n$ und $\mathcal{B} = (w_1, \ldots, w_n)$ eine Basis von W, dann gilt $\Phi_{\mathcal{B}}(x) = A_{\mathcal{B}} \cdot x$ mit $A_{\mathcal{B}} = (w_1, \ldots, w_n) \in M(n \times n, K)$, für alle $x \in K^n$. (Die Spalten von $A_{\mathcal{B}}$ sind also die nebeneinandergeschriebenen Basisvektoren w_i.)*

Beweis. i) folgt aus Satz 10.1 und ii) ist Satz 8.11. □

Wir können nun die Verallgemeinerung von Satz 8.11 angeben:

Satz 10.3 *Gegeben seien K-Vektorräume V mit Basis $\mathcal{A} = (v_1, \ldots, v_n)$ und W mit Basis $\mathcal{B} = (w_1, \ldots, w_m)$.*

i) *Dann gibt es zu jeder linearen Abbildung $F : V \to W$ genau eine Matrix $A = (a_{ij}) \in M(m \times n, K)$, so daß $F(v_j) = \sum_{i=1}^m a_{ij} \cdot w_i$.*

ii) *Die so erhaltene Abbildung*

$$M^{\mathcal{A}}_{\mathcal{B}} : \mathrm{Hom}(V, W) \to M(m \times n, K) , \qquad M^{\mathcal{A}}_{\mathcal{B}} : F \mapsto A = M^{\mathcal{A}}_{\mathcal{B}}(F)$$

ist ein Isomorphismus von Vektorräumen über K, und insbesondere gilt

$$M^{\mathcal{A}}_{\mathcal{B}}(\lambda \cdot F + \mu \cdot G) = \lambda \cdot M^{\mathcal{A}}_{\mathcal{B}}(F) + \mu \cdot M^{\mathcal{A}}_{\mathcal{B}}(G) .$$

Ist $V = K^n$ und $W = K^m$ und sind $\mathcal{A}$ und $\mathcal{B}$ die Standardbasen, dann ist $M^{\mathcal{A}}_{\mathcal{B}} = M^n_m : \mathrm{Hom}(K^n, K^m) \to M(m \times n, K)$ die Konstruktion aus Satz 8.11.

Definition 10.4 Mit den Bezeichnungen aus Satz 10.3 heißt $A = M^{\mathcal{A}}_{\mathcal{B}}(F)$ die *darstellende Matrix* von $F : V \to W$ bezüglich der Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W.

Beweis von Satz 10.3. i) Durch Komposition von F mit den Isomorphismen $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ aus Satz 10.2:

$$A := (\Phi_{\mathcal{B}})^{-1} \circ F \circ \Phi_{\mathcal{A}} : K^n \to K^m , \qquad A \cdot x := (\Phi_{\mathcal{B}})^{-1}(F(\Phi_{\mathcal{A}}(x))) .$$

Umgekehrt ergibt sich F aus A durch $F = \Phi_{\mathcal{B}} \circ A \circ (\Phi_{\mathcal{A}})^{-1} : V \to W$ mit $A \circ x := A \cdot x$. Zunächst ist $(\Phi_{\mathcal{A}})^{-1}(v_j) = e_j \in K^n$. Matrixmultiplikation liefert

$$A \cdot ((\Phi_{\mathcal{A}})^{-1}(v_j)) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \sum_{i=1}^{m} a_{ij} \cdot e_i \in K^m .$$

Anwenden von $\Phi_{\mathcal{B}}$ liefert $\Phi_{\mathcal{B}}(A \cdot ((\Phi_{\mathcal{A}})^{-1}(v_j))) = \sum_{i=1}^{m} a_{ij} \cdot w_i \in W$.

ii) Als Komposition linearer Abbildungen ist $M^{\mathcal{A}}_{\mathcal{B}} : \mathrm{Hom}(V,W) \to M(m \times n, K)$ mit

$$A = M^{\mathcal{A}}_{\mathcal{B}}(F) = (\Phi_{\mathcal{B}})^{-1} \circ F \circ \Phi_{\mathcal{A}} : K^n \to K^m .$$

linear. Da $\mathcal{A}$ eine Basis ist, gibt es nach Satz 10.1.ii) genau eine lineare Abbildung F mit $F(v_j) = \widetilde{w_j} := \sum_{i=1}^{m} a_{ij} \cdot w_i$. Also ist $M^{\mathcal{A}}_{\mathcal{B}}$ bijektiv. $\square$

Beispiel 10.5 Wir betrachten die durch $F((x,y,z)) = (x-y, 2x+y-z)$ definierte lineare Abbildung $F : \mathbb{R}^3 \to \mathbb{R}^2$ bezüglich der Basen (v_1, v_2, v_3) von $\mathbb{R}^3$ und (w_1, w_2) von $\mathbb{R}^2$ mit

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} , \; v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} , \; v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} , \; w_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \; w_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} .$$

Damit gilt

$$F(v_1) = F((0,1,1)) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} , \; F(v_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \; F(v_3) = \begin{pmatrix} 0 \\ 3 \end{pmatrix} .$$

Diese Vektoren sind nun bezüglich der Basis (w_1, w_2) darzustellen, $w_1 a_{1j} + w_2 a_{2j} = F(v_j)$. Es handelt sich also um drei lineare Gleichungssysteme aus 2 Gleichungen mit 2 Unbekannten:

$$\begin{aligned}
\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} & \Rightarrow \quad \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} &= \begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} \\
\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \Rightarrow \quad \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} &= \begin{pmatrix} 0 \\ 3 \end{pmatrix} & \Rightarrow \quad \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} &= \begin{pmatrix} 2 \\ -1 \end{pmatrix} .
\end{aligned}$$

Die darstellende Matrix ist folglich $A = \begin{pmatrix} -\frac{1}{3} & 1 & 2 \\ -\frac{1}{3} & 0 & -1 \end{pmatrix}$. $\triangleleft$

Sind also die Basen der Vektorräume V und W festgelegt, dann kann man die Vektorräume mit den Standardräumen K^n, K^m der Koordinaten identifizieren und lineare Abbildungen $F : V \to W$ mit Matizen $A \in M(m \times n, K)$. Eine oftmals sinnvolle Wahl besteht aus Basen, in denen für die Koordinatenmatrix die Einheitsmatrix entsteht:

Definition 10.6 Die Matrix

$$E_r = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in M(r \times r, K) ,$$

in der sämtliche Einträge auf der *Diagonalen* gleich 1 und alle anderen Einträge gleich 0 sind, heißt $(r \times r)$-*Einheitsmatrix*. Ihre Komponenten sind also $E_r = (\delta_{ij})$, wobei

$$\delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

das *Kronecker-Symbol* ist.

Satz 10.7 *Sei $F : V \to W$ linear und $\dim(V) = n$, $\dim(W) = m$, $\dim(\mathrm{im}(F)) = \mathrm{rang}(F) = r$. Dann gibt es Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W mit*

$$M_{\mathcal{B}}^{\mathcal{A}}(F) = \begin{pmatrix} E_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} .$$

Dabei ist $0_{p\times q} \in M(p \times q, K)$ die Matrix, in der sämtliche Einträge gleich 0 sind.

Beweis. Wir wählen eine Basis $(w_1, \dots, w_r)$ von $\mathrm{im}(F)$ und ergänzen sie zu einer Basis $\mathcal{B} = (w_1, \dots, w_r, w_{r+1}, \dots, w_m)$ von W. Dann wählen wir wie im Beweis von Satz 8.7 Vektoren $u_1, \dots u_r \in V$ mit $F(u_i) = w_i$ für alle $1 \leq i \leq r$ und ergänzen sie mit einer Basis $(v_1, \dots, v_k)$ von $\ker(F)$ zu einer Basis $\mathcal{A} = (u_1, \dots, u_r, v_1, \dots, v_k)$ von V. Dann hat $M_{\mathcal{B}}^{\mathcal{A}}(F)$ die angegebene Form. □

11 Multiplikation von Matrizen

Seien $G : K^r \to K^n$ und $F : K^n \to K^m$ lineare Abbildungen, dargestellt (bezüglich der Standardbasen) nach Satz 8.11 durch Matrizen $B = M_n^r(G) \in M(n \times r, K)$ und $A = M_m^n(F) \in M(m \times n, K)$.

Nach Satz 8.2 ist die Komposition $F \circ G : K^r \to K^m$ wieder eine lineare Abbildung, welche durch eine Matrix $C = M_m^r(F \circ G) \in M(m \times r, K)$ dargestellt wird. Das bedeutet $C = M_m^r\big((M_m^n)^{-1}(A) \circ (M_n^r)^{-1}(B)\big)$, und auf diese Weise wird ein Produkt

$$\begin{aligned} \cdot : M(m \times n, K) \times M(n \times r, K) &\to M(m \times r, K) , \\ A \cdot B &:= M_m^r\big((M_m^n)^{-1}(A) \circ (M_n^r)^{-1}(B)\big) \end{aligned}$$

von Matrizen *passender Größe* definiert. Im Produkt muß die Zahl der Spalten der linken Matrix gleich der Zahl der Zeilen der rechten Matrix sein.

Satz 11.1 *Ist* $A = (a_{ij}) \in M(m \times n, K)$, $B = (B_{jk}) \in M(n \times r, K)$ *und* $A \cdot B = C = (c_{ik}) \in M(r \times s, K)$, *so gilt* $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$.

Beweis. Sei $F = (M_n^m)^{-1}(A) : K^n \to K^m$ und $G = (M_r^n)^{-1}(B) : K^r \to K^n$ sowie $F \circ G = (M_r^m)^{-1}(C) : K^r \to K^m$. Seien $(\mathcal{E}_1, \ldots, \mathcal{E}_r)$ die Standardbasis im K^r, $(\epsilon_1, \ldots, \epsilon_n)$ die Standardbasis im K^n und $(e_1, \ldots, e_m)$ die Standardbasis im K^m. Dann gilt

$$
\begin{aligned}
(F \circ G)(\mathcal{E}_k) &= \sum_{i=1}^{m} c_{ik} \cdot e_i && \text{(Satz 10.3 für } F \circ G) \\
&= F(G(\mathcal{E}_k)) && \text{(Definition von } F \circ G) \\
&= F\Big(\sum_{j=1}^{n} b_{jk} \cdot \epsilon_j\Big) && \text{(Satz 10.3 für } G) \\
&= \sum_{j=1}^{n} b_{jk} \cdot F(\epsilon_j) && \text{(Linearität von } F) \\
&= \sum_{j=1}^{n} b_{jk} \cdot \Big(\sum_{i=1}^{m} a_{ij} \cdot e_i\Big) && \text{(Satz 10.3 für } F) \\
&= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} b_{jk}\Big) \cdot e_i && \text{(Umordnen der Summen)}
\end{aligned}
$$

Die Behauptung folgt nun aus der Eindeutigkeit der Koeffizienten zur Basis $(e_i)_{i=1,\ldots,m}$ von K^m. □

Ein Spezialfall dieser Rechenregel ist das schon zuvor erklärte Matrixprodukt $A \cdot v \in K^m$ einer Matrix $A \in M(m \times n, K)$ mit einem Spaltenvektor $v \in K^n$. Mit den Bezeichnungen aus Satz 11.1 gilt: Die i-te Zeile von A ist $(a_{i1}\ a_{i2}\ \ldots\ a_{in})$, und die j-te Spalte von B ist $\begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}$. Damit folgt: Das Matrixelement c_{ij} von $C = A \cdot B$ ist das "Produkt" der i-ten Zeile von A mit der j-ten Spalte von B.

Beispiel 11.2 Sei $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 2 \end{pmatrix}$ und $B = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \end{pmatrix}$, dann gilt

$$A \cdot B = \begin{pmatrix} 3 & 2 & 5 \\ 1 & 0 & 1 \\ 4 & 4 & 8 \end{pmatrix} \quad \text{und} \quad B \cdot A = \begin{pmatrix} 7 & 10 \\ 3 & 4 \end{pmatrix}.$$

Daraus ist ersichtlich, daß selbst wenn sowohl $A \cdot B$ als auch $B \cdot A$ erklärt sind, im allgemeinen $A \cdot B \neq B \cdot A$ gilt. ◁

Beispiel 11.3 Eine Drehung in der Ebene um den Ursprung mit Winkel α wird durch die Matrix $A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ beschrieben. Entsprechend beschreibt $B = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}$ eine Drehung mit Winkel β. Die Hintereinanderausführung beider Drehungen ist durch das Matrixprodukt gegeben:

$$\begin{aligned} A \cdot B &= \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \\ &= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\sin\alpha\cos\beta - \cos\alpha\sin\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & \cos\alpha\cos\beta - \sin\alpha\sin\beta \end{pmatrix} . \end{aligned}$$

Andererseits ist die Hintereinanderausführung beider Drehungen wieder eine Drehung mit Gesamtwinkel $\alpha + \beta$. Daraus folgen die Additionstheoreme für Sinus und Cosinus,

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta\,, \qquad \cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

◁

Satz 11.4 *Sind Matrizen $A, A' \in M(m \times n, K)$, $B, B' \in M(n \times s, K)$ und $C \in M(r \times s, K)$ gegeben sowie $\lambda \in K$, so gilt:*

i) $A \cdot (B + B') = A \cdot B + A \cdot B'$ *und* $(A + A') \cdot B = A \cdot B + A' \cdot B$ *(Distributivgesetze)*

ii) $A \cdot (\lambda \cdot B) = (\lambda \cdot A) \cdot B = \lambda \cdot (A \cdot B)$

iii) $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ *(Assoziativgesetz)*

iv) $E_m \cdot A = A \cdot E_n = A$ *(Neutralität der Einheitsmatrix)*

Beweis. i) und ii) folgen aus der Summendarstellung des Matrixprodukts.

iii) Unter Verwendung der Assoziativität der Komposition beliebiger Abbildungen (Bemerkung i) nach Beispiel 2.3) gilt

$$\begin{aligned} (A \cdot B) \cdot C &= M_m^s((M_m^r)^{-1}(A \cdot B) \circ (M_r^s)^{-1}(C)) \\ &= M_m^s(((M_m^n)^{-1}(A) \circ (M_n^r)^{-1}(B)) \circ (M_r^s)^{-1}(C)) \\ &= M_m^s((M_m^n)^{-1}(A) \circ ((M_n^r)^{-1}(B) \circ (M_r^s)^{-1}(C))) \\ &= M_m^s((M_m^n)^{-1}(A) \circ (M_n^s)^{-1}(B \cdot C)) = A \cdot (B \cdot C) \end{aligned}$$

iv) Sei $E_m \cdot A = (\alpha_{ij}) \in M(m \times n, K)$, dann gilt $\alpha_{ij} = \sum_{k=1}^m \delta_{ik} a_{kj} = a_{ij}$, denn zu gegebenem Index i überlebt in der Summe nur der Term mit $i = k$. Analog folgt $A \cdot E_n = A$. □

Im weiteren schreiben wir oft AB statt $A \cdot B$ für das Produkt der Matrizen.

Satz 11.5 *Die Menge $M(n, K) := M(n \times n, K)$ der quadratischen Matrizen ist ein Ring mit Einselement E_n.* □

Definition 11.6 Die Abbildung $t : M(m \times n, K) \to M(n \times m, K)$ definiert durch $t : A = (a_{ij}) \mapsto A^t := (a_{ji})$ heißt *Transposition von Matrizen.*

Satz 11.7 *Wenn das Matrixprodukt AB erklärt ist, dann gilt $(A \cdot B)^t = B^t \cdot A^t$.*

Beweis. Sei $A = (a_{ij}) \in M(m \times n, K)$, $B = (b_{jk}) \in M(n \times r, K)$ und $C = AB = (c_{ik}) \in M(m \times r, K)$. Dann ist $c_{ik} = \sum_{j=1}^n a_{ij}b_{jk}$ und $C^t = (c'_{ki})$ mit $c'_{ki} = c_{ik}$. Weiter ist $B^t = (b'_{kj})$ mit $b'_{kj} = b_{jk}$ und $A^t = (a'_{ji})$ mit $a'_{ji} = a_{ij}$. Dann ist $B^tA^t = D = (d_{ki})$ mit $d_{ki} = \sum_{j=1}^n b'_{kj}a'_{ji} = \sum_{j=1}^n b_{jk}a_{ij} = \sum_{j=1}^n a_{ij}b_{jk} = c_{ik} = c'_{ki}$ und somit $D = C^t$. □

Definition 11.8 Eine quadratische Matrix $A \in M(n, K)$ heißt *invertierbar*, wenn es ein $A^{-1} \in M(n, K)$ gibt mit $A \cdot A^{-1} = A^{-1} \cdot A = E_n$.

Satz 11.9 *Die Menge*

$$GL(n, K) = \{A \in M(n, K) \ : \ A \text{ ist invertierbar}\}$$

der invertierbaren Matrizen ist eine Gruppe mit der Matrixmultiplikation als Verknüpfung und der Einheitsmatrix E_n als neutralem Element.

Diese Gruppe $GL(n, K)$ heißt die *allgemeine lineare Gruppe.*

Beweis. Da E_n ein neutrales Element ist, ist nur zu zeigen, daß aus $A, B \in GL(n, K)$ auch $AB \in GL(n, K)$ folgt. Die Matrix $B^{-1}A^{-1}$ erfüllt $(B^{-1}A^{-1})AB = E_n$ unter Verwendung des Assoziativgesetzes. Aus den allgemeinen Gruppeneigenschaften folgt, daß das neutrale Element und das zu A inverse Element eindeutig sind. □

Satz 11.10

i) *$A \in M(n, K)$ ist genau dann invertierbar, wenn* $\operatorname{rang}(A) = n$.

ii) *Für $A \in GL(n, K)$ gilt $(A^t)^{-1} = (A^{-1})^t$.*

Beweis. i) Ist $\operatorname{rang}(A) = \operatorname{rang}(F) = n$, dann ist F surjektiv nach Satz 8.5.ii) und dann bijektiv nach Satz 8.9.

ii) Wir haben $(A^{-1})^tA^t = (AA^{-1})^t = E_n^t = E_n$, so daß $(A^{-1})^t$ das Inverse zu A^t ist. □

Ist $A \in M(n, K)$ invertierbar, dann ist das lineare Gleichungssystem $Ax = b$ wegen $\operatorname{rang}(A) = n$ eindeutig lösbar mit Lösung $x = A^{-1}b$ für beliebiges $b \in K^n$. Im nächsten Abschnitt werden wir eine Methode zur Bestimmung der inversen Matrix gewinnen.

12 Elementarmatrizen

Die elementaren Zeilenumformungen, die eine Matrix $B \in M(m \times n, K)$ (z.B. die erweiterte Koeffizientenmatrix (A, b), welche ein ein lineares Gleichungssystem beschreibt) in Zeilenstufenform bringen, lassen sich durch Linksmultiplikation von B mit geeigneten $(m \times m)$-Matrizen beschreiben. Diese *Elementarmatrizen* sind:

Typ I: $B \xrightarrow{\mathrm{I}_{ik}} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = P_{ik} \cdot B$ mit

$$P_{ik} = \begin{pmatrix} 1 & & & \mid & & & & \mid & & & \\ & \ddots & & & & & & & & & \\ & & 1 & \mid & & & & \mid & & & \\ - & - & - & 0 & - & - & - & 1 & - & - & - \\ & & & \mid & 1 & & & \mid & & & \\ & & & & & \ddots & & & & & \\ & & & \mid & & & 1 & \mid & & & \\ - & - & - & 1 & - & - & - & 0 & - & - & - \\ & & & \mid & & & & \mid & 1 & & \\ & & & & & & & & & \ddots & \\ & & & \mid & & & & \mid & & & 1 \end{pmatrix}$$

$$= E_m - E_{ii} - E_{kk} + E_{ik} + E_{ki}$$

(Angedeutet sind die i-te und k-te Zeile und Spalte. Alle nicht bezeichneten Einträge sind 0.)

Typ II: ist Spezialfall $\lambda = 1$ von Typ IV

Typ III: $B \xrightarrow{\mathrm{III}_i} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = S_i(\lambda) \cdot B$ mit

$$S_i(\lambda) = \begin{pmatrix} 1 & & & \mid & & & \\ & \ddots & & & & & \\ & & 1 & \mid & & & \\ - & - & - & \lambda & - & - & - \\ & & & \mid & 1 & & \\ & & & & & \ddots & \\ & & & \mid & & & 1 \end{pmatrix} = E_m + (\lambda - 1)E_{ii}$$

Typ IV: $B \overset{\mathrm{IV}_{ik}(\lambda)}{\longrightarrow} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = Q_{ik}(\lambda) \cdot B$ mit

$$Q_{ik}(\lambda) = \begin{pmatrix} 1 & & & | & & & & | & & & \\ & \ddots & & & & & & & & & \\ & & 1 & | & & & & | & & & \\ - & - & - & 1 & - & - & - & 0 & - & - & - \\ & & & | & 1 & & & | & & & \\ & & & & & \ddots & & & & & \\ & & & | & & & 1 & | & & & \\ - & - & - & \lambda & - & - & - & 1 & - & - & - \\ & & & | & & & & | & 1 & & \\ & & & & & & & & & \ddots & \\ & & & | & & & & | & & & 1 \end{pmatrix} = E_m + \lambda E_{ki}$$

(Angedeutet sind die i-te und k-te Zeile und Spalte.)

Analog sind *elementare Spaltenumformungen* einer Matrix $A \in M(m \times n, K)$ gegeben durch die entsprechenden *Rechtsmultiplikationen* von A mit Elementarmatrizen $P, Q, S \in M(n \times n, K)$.

Tatsächlich genügen die Matrizen $S_i(\lambda)$ und $Q(1)$ für Typ II und III, denn Typ I eine Kombination aus mehreren Umformungen vom Typ II und III (ersteres als Spezialfall von IV):

$$P_{ik} = Q_{ik}(1)\ S_k(-1)Q_{ki}(-1)S_i(-1)Q_{ik}(1)S_i(-1)\ ,$$

denn

$$(i,k) \overset{\mathrm{III}_i(-1)}{\longrightarrow} (-i,k) \overset{\mathrm{II}_{ik}}{\longrightarrow} (-i,k-i) \overset{\mathrm{III}_i(-1)}{\longrightarrow} (i,k-i) \overset{\mathrm{II}_{ki}}{\longrightarrow} (k,k-i) \overset{\mathrm{III}_k(-1)}{\longrightarrow} (k,i-k) \overset{\mathrm{II}_{ik}}{\longrightarrow} (k,i)\ .$$

Für Typ IV gilt

$$Q_{ij}(\lambda) = S_i(\tfrac{1}{\lambda})Q_{ik}(1)S_i(\lambda)\ ,$$

denn

$$(i,k) \overset{\mathrm{III}_i(\lambda)}{\longrightarrow} (\lambda i,k) \overset{\mathrm{II}_{ik}}{\longrightarrow} (\lambda i, k+\lambda i) \overset{\mathrm{III}_i(\frac{1}{\lambda})}{\longrightarrow} (i, k+\lambda i)$$

Die Elementarmatrizen sind invertierbar mit

$$P_{ik}^{-1} = P_{ik}\ , \qquad (Q_{ik}(\lambda))^{-1} = Q_{ik}(-\lambda)\ , \qquad (S_i(\lambda))^{-1} = S_i(\tfrac{1}{\lambda})\ .$$

Eine wichtige Anwendung der Elementarmatrizen ist der folgende Satz:

Satz 12.1 *Jede invertierbare Matrix $A \in GL(n,K)$ ist ein endliches Produkt von Elementarmatrizen.*

Beweis. Da $\text{rang}(A) = n$, führen elementare Zeilenumformungen mit Typ I und IV erst zu einer Dreiecksmatrix mit Diagonalelementen $d_{ii} \neq 0$ für alle $1 \leq i \leq n$, nach weiterer Umformung mit Typ IV dann zu einer Diagonalmatrix mit gleichen Diagonalelementen $d_{ii} \neq 0$, und schließlich mit Typ III zur Einheitsmatrix. Damit gilt $B_r \cdots B_2 \cdot B_1 \cdot A = E_n$, wobei jedes B_s eine Elementarmatrix $Q_{ik}(\lambda)$ oder $S_i(\lambda)$ ist. Folglich gilt $A = B_1^{-1} \cdot B_2^{-1} \cdots B_r^{-1}$ und $A^{-1} = B_r \cdots B_2 \cdot B_1$. □

Eine nützliche Anwendung dieses Beweises besteht in einer effizienten Berechnungsmethode für die *inverse Matrix.* Durch Vergleich von $B_r \cdots B_2 \cdot B_1 \cdot A = E_n$ mit $A^{-1} = B_r \cdots B_2 \cdot B_1 \cdot E_n$ lesen wir ab:

Satz 12.2 *Dieselben Zeilenumformungen, welche eine invertierbare Matrix $A \in GL(n, K)$ in E_n überführen, überführen E_n in A^{-1}.* □

Beispiel 12.3

$$\left(\begin{array}{cc|cc} 9 & 4 & 1 & 0 \\ 11 & 5 & 0 & 1 \end{array}\right) \xrightarrow{Q_{12}(-\frac{11}{9})} \left(\begin{array}{cc|cc} 9 & 4 & 1 & 0 \\ 0 & \frac{1}{9} & -\frac{11}{9} & 1 \end{array}\right) \xrightarrow{Q_{21}(-36)} \left(\begin{array}{cc|cc} 9 & 0 & 45 & -36 \\ 0 & \frac{1}{9} & -\frac{11}{9} & 1 \end{array}\right)$$
$$\xrightarrow{S_1(\frac{1}{9})S_2(9)} \left(\begin{array}{cc|cc} 1 & 0 & 5 & -4 \\ 0 & 1 & -11 & 9 \end{array}\right)$$

Folglich ist $\begin{pmatrix} 9 & 4 \\ 11 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5 & -4 \\ -11 & 9 \end{pmatrix}$, was durch Ausmultiplizieren leicht zu überprüfen ist. ◁

Ein anderer Weg, diese Rechnung zu verstehen, interpretiert $A \cdot A^{-1} = E_n$ als n lineare Gleichungssysteme $A \cdot x^{(p)} = b^{(p)}$. Dabei sind die $x^{(p)}$ bzw. $b^{(p)}$ mit $1 \leq p \leq n$ gerade die n Spalten von A^{-1} bzw. E_n. Anstatt nun die Berechnung von $x^{(p)}$ durch Zeilenumformung jeder dieser erweiterten Koeffizientenmatrizen $(A, b^{(p)})$ separat durchzuführen, schreiben wir alle Spalten $b^{(p)}$ nebeneinander und formen die entstehende Matrix (A, E_n) gleichzeitig um. (Es ist stets dieselbe Rechnung, die A in E_n überführt, unabhängig von b).

Man muß zunächst nicht wissen, ob A invertierbar ist, um dieses Verfahren durchzuführen. Wir können sogar nichtquadratische Matrizen $A \in M(m \times n, K)$ zulassen. Sei also $\text{rang}(A) = r$, dann führen die Zeilenumformungen zu $(m - r)$ Zeilen, in denen sämtliche Einträge gleich 0 sind. Wir erreichen dann bestenfalls die Blockdarstellung

$$(A, E_m) \mapsto \left(\begin{array}{c|c} D_{r\times n} & L_{r\times m} \\ 0_{(m-r)\times n} & M_{(m-r)\times m} \end{array}\right),$$

wobei die Größe der Matrizen entsprechend $D_{r\times n} \in M(r \times n, K)$ usw. gekennszeichnet ist. Nach Konstruktion als Matrix in spezieller Zeilenstufenform sind r der $n \geq r$ *Spalten* von $D_{r\times n}$ die paarweise verschiedenen Standardbasisvektoren $e_1, \ldots, e_r \in K^r$.

Es gibt also Elementarmatrizen $B_1, \ldots, B_r \in GL(m,K)$ mit $B_r \cdots B_1 \cdot A = \begin{pmatrix} D_{r\times n} \\ 0_{(m-r)\times n} \end{pmatrix}$ und $B_r \cdots B_1 \cdot E_m = \begin{pmatrix} L_{r\times m} \\ M_{(m-r)\times m} \end{pmatrix} =: L^{-1}$. Durch Transposition entsteht die Matrix $\begin{pmatrix} D_{r\times n} \\ 0_{(m-r)\times n} \end{pmatrix}^t = \begin{pmatrix} D^t_{r\times n} & 0_{n\times(m-r)} \end{pmatrix} \in M(n\times m, K)$, wobei nun r der $n \geq r$ *Zeilen* von $D^t_{r\times n} =: \tilde{D}_{n\times r}$ die transponierten Standardbasisvektoren $e^t_1, \ldots, e^t_r \in K^r$ sind. Elementare Zeilenumformungen überführen diese Matrix in $\begin{pmatrix} E_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{pmatrix} \in M(n\times m, K)$. Es gibt also Elementarmatrizen $C_1, \ldots, C_s \in GL(n,K)$ mit

$$C_s \cdots C_2 \cdot C_1 \cdot \begin{pmatrix} D_{r\times n} \\ 0_{(m-r)\times n} \end{pmatrix}^t = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

und somit nach Rücktransposition

$$\begin{pmatrix} D_{r\times n} \\ 0_{(m-r)\times n} \end{pmatrix} \cdot (C_s \cdots C_2 \cdot C_1)^t = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Das Produkt dieser Elementarmatrizen $\tilde{R} = C_s \cdots C_2 \cdot C_1$ erhalten wir wieder durch die korrespondierende Zeilenumformung der Einheitsmatrix:

$$\left(\begin{array}{cc|c} \tilde{D}_{n\times r} & 0_{n\times(m-r)} & E_n \end{array}\right) \mapsto \left(\begin{array}{cc|c} E_r & 0 & \tilde{R} \\ 0 & 0 & \end{array}\right).$$

Damit haben wir folgenden Satz bewiesen:

Satz 12.4 *Zu jeder Matrix $A \in M(m\times n, K)$ mit* $\operatorname{rang}(A) = r$ *gibt es invertierbare Matrizen $L \in GL(m,K)$ und $R \in GL(n,K)$, so daß $A = L\cdot \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \cdot R^{-1}$. (Dabei ist $R = \tilde{R}^t$.)*

Beispiel 12.5

$$\left(\begin{array}{ccc|cc} 1 & 2 & 3 & 1 & 0 \\ 4 & 5 & 6 & 0 & 1 \end{array}\right) \xrightarrow{Q_{12}(-4)} \left(\begin{array}{ccc|cc} 1 & 2 & 3 & 1 & 0 \\ 0 & -3 & -6 & -4 & 1 \end{array}\right)$$
$$\xrightarrow{Q_{21}(\frac{2}{3})} \left(\begin{array}{ccc|cc} 1 & 0 & -1 & -\frac{5}{3} & \frac{2}{3} \\ 0 & -3 & -6 & -4 & 1 \end{array}\right) \xrightarrow{S_2(-\frac{1}{3})} \left(\begin{array}{ccc|cc} 1 & 0 & -1 & -\frac{5}{3} & \frac{2}{3} \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} \end{array}\right).$$

Damit ist $L^{-1} = \begin{pmatrix} -\frac{5}{3} & \frac{2}{3} \\ \frac{4}{3} & -\frac{1}{3} \end{pmatrix}$. Die übliche Rechnung für das Inverse liefert $L = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$.

Nach Transposition berechnen wir

$$\left(\begin{array}{rr|rrr} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ -1 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{Q_{13}(1)\,Q_{23}(-2)} \left(\begin{array}{rr|rrr} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{array}\right),$$

also $\tilde{R}^t = R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$. Diese Matrix ist nun leicht zu invertieren: $R^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

Folglich gilt

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Man kann Satz 12.4 auch abstrakt verstehen: Die Matrix $A = (a_{ij}) \in M(m \times n, K)$ entspricht einer linearen Abbildung $F : K^n \to K^m$ mit $F(e_j) = \sum_{i=1}^m a_{ij} \cdot e_i$. Ist $\operatorname{rang}(F) = \operatorname{rang}(A) = r$, dann existieren nach Satz 10.7 Basen $\mathcal{A} = (v_1, \ldots, v_n)$ von K^n und $\mathcal{B} = (w_1, \ldots, w_m)$ von K^m, so daß $M^{\mathcal{A}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \tilde{A} = (\tilde{a}_{kl})$. Die mit L und R bewirkte Umrechnung von A in $\tilde{A}$ ist ein Beispiel für eine *Koordinatentransformation.*

Satz 12.6 *Für $A \in M(m \times n, K)$ ist* $\operatorname{rang}(A) = \operatorname{rang}(A^t)$.

Beweis. Sei $\operatorname{rang}(A) = r$. Nach Satz 10.7 gibt es angepaßte Basen $\mathcal{A}$ in K^n und $\mathcal{B}$ in K^m, so daß $A = L \cdot \begin{pmatrix} E_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} \cdot R^{-1}$ für invertierbare Matrizen $L \in GL(m, K)$ und $R \in GL(n, K)$. Dann ist $A^t = (R^{-1})^t \cdot B_r \cdot L^t$ mit $B_r := \begin{pmatrix} E_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{pmatrix}$. Mit R ist auch R^t invertierbar, also ist $(R^{-1})^t : K^n \to K^m$ ein Isomorphismus. Somit gilt $\dim(\operatorname{im}(A^t)) = \dim(\operatorname{im}(B_r \cdot L^t) \leq r$. Durch Umkehrung von $A \mapsto A^t$ (oder Verwendung der Surjektivität von L^t) folgt die Behauptung. □

13 Kommutative Diagramme und Basiswechsel

Koordinatentransformationen werden hervorgerufen durch einen Basiswechsel in einem Vektorraum. Sind in einem Vektorraum V zwei Basen $\mathcal{A} = (v_1, \ldots, v_n)$ und $\tilde{\mathcal{A}} = (\tilde{v}_1, \ldots, \tilde{v}_n)$ gegeben, so existieren nach Satz 10.2 Isomorphismen $\Phi_{\mathcal{A}} : K^n \to V$ und $\Phi_{\tilde{\mathcal{A}}} : K^n \to V$ mit $\Phi_{\mathcal{A}}(e_i) = v_i$ und $\Phi_{\tilde{\mathcal{A}}}(e_i) = v_i$. Die identitische Abbildung auf V induziert dann einen Isomorphismus $\mathcal{T}_{\mathcal{A}}^{\tilde{\mathcal{A}}} := \Phi_{\tilde{\mathcal{A}}}^{-1} \circ \Phi_{\mathcal{A}} : K^n \to K^n$, den man am besten als *kommutatives Diagramm* schreibt:

$$\begin{tikzcd} K^n \arrow[r, "\Phi_{\mathcal{A}}"] \arrow[d, "T_{\tilde{\mathcal{A}}}^{\mathcal{A}}=\Phi_{\tilde{\mathcal{A}}}^{-1}\circ\Phi_{\mathcal{A}}"'] & V \arrow[d, "id_V"] \\ K^n \arrow[r, "\Phi_{\tilde{\mathcal{A}}}"'] & V \end{tikzcd} \qquad \textit{bzw.} \qquad \begin{tikzcd} K^n \arrow[dr, "\Phi_{\mathcal{A}}"] \arrow[dd, "T_{\tilde{\mathcal{A}}}^{\mathcal{A}}=\Phi_{\tilde{\mathcal{A}}}^{-1}\circ\Phi_{\mathcal{A}}"'] & \\ & V \\ K^n \arrow[ur, "\Phi_{\tilde{\mathcal{A}}}"'] & \end{tikzcd}$$

Dabei bedeutet Kommutativität, daß die Komposition der Abbildungen unabhängig vom Weg ist.

Die darstellende Matrix der identischen Abbildung ist somit $T_{\tilde{\mathcal{A}}}^{\mathcal{A}} = (T_{ij}) = \mathcal{M}_{\tilde{\mathcal{A}}}^{\mathcal{A}}(id_V)$ mit $id_V(v_j) = v_j = \sum_{i=1}^n T_{ij}\tilde{v}_i$. Sei $v = \sum_{j=1}^n x_j v_j = \sum_{i=1}^n y_i \tilde{v}_i \in V$, dann gilt

$$v = \sum_{j=1}^n x_j v_j = \sum_{i,j=1}^n T_{ij} x_j \tilde{v}_i = \sum_{i=1}^n y_i \tilde{v}_i \qquad \Rightarrow \quad y = T_{\tilde{\mathcal{A}}}^{\mathcal{A}} \cdot x\ .$$

Kennt man also die *Transformationsmatrix* $T_{\tilde{\mathcal{A}}}^{\mathcal{A}}$, so kann man die neuen Koordinaten y von $v \in V$ bezüglich der Basis $\tilde{\mathcal{A}}$ aus den alten Koordinaten x bezüglich der Basis $\mathcal{A}$ berechnen. Besonders wichtig ist der Fall $V = K^n$. Dann sind $\Phi_{\mathcal{A}} =: A \in M(n \times n, K)$ und $\Phi_{\tilde{\mathcal{A}}} = \tilde{A} \in M(n \times n, K)$ selbst Matrizen, gegeben durch die jeweiligen Basisvektoren als Spalten. In diesem Fall ist nach Definition des Matrixprodukts $T_{\tilde{\mathcal{A}}}^{\mathcal{A}} = \tilde{A}^{-1} \cdot A$. An Stelle der Lösung eines linearen Gleichungssystems tritt nun die einfachere Aufgabe der Berechnung einer inversen Matrix.

Allgemein läßt sich die darstellende Matrix einer linearen Abbildung $F : V \to W$ bezüglich Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W als kommutatives Diagramm auffassen:

$$\begin{tikzcd} K^n \arrow[r, "\Phi_{\mathcal{A}}"] \arrow[d, "M_{\mathcal{B}}^{\mathcal{A}}(F)=\Phi_{\mathcal{B}}^{-1}\circ F\circ\Phi_{\mathcal{A}}"'] & V \arrow[d, "F"] \\ K^m \arrow[r, "\Phi_{\mathcal{B}}"'] & W \end{tikzcd}$$

Durch Kombination mit dem Basiswechsel erhalten wir:

Satz 13.1 (Transformationsformel) *Es sei $F : V \to W$ eine lineare Abbildung sowie $\mathcal{A}, \tilde{\mathcal{A}}$ Basen von V und $\mathcal{B}, \tilde{\mathcal{B}}$ Basen von W, mit* $\dim(V) = n$ *und*

$\dim(W) = m$. *Dann ist folgendes Diagramm kommutativ:*

$$\begin{tikzcd}
K^n \arrow[rrr, "M^{\mathcal{A}}_{\mathcal{B}}(F)"] \arrow[dr, "\Phi_{\mathcal{A}}"] \arrow[dd, "T^{\mathcal{A}}_{\tilde{\mathcal{A}}}"'] & & & K^m \arrow[dl, "\Phi_{\mathcal{B}}"'] \arrow[dd, "T^{\mathcal{B}}_{\tilde{\mathcal{B}}}"] \\
& V \arrow[r, "F"] & W & \\
K^n \arrow[rrr, "M^{\tilde{\mathcal{A}}}_{\tilde{\mathcal{B}}}(F)"'] \arrow[ur, "\Phi_{\tilde{\mathcal{A}}}"'] & & & K^m \arrow[ul, "\Phi_{\tilde{\mathcal{B}}}"]
\end{tikzcd}$$

Insbesondere gilt $M^{\tilde{\mathcal{A}}}_{\tilde{\mathcal{B}}}(F) = T^{\mathcal{B}}_{\tilde{\mathcal{B}}} \cdot M^{\mathcal{A}}_{\mathcal{B}}(F) \cdot (T^{\mathcal{A}}_{\tilde{\mathcal{A}}})^{-1}$.

Beweis. Alle Dreiecke und Teilvierecke sind kommuativ, also auch das gesamte Diagramm. □

Insbesondere läßt sich die darstellende Matrix einer linearen Abbildung $F : K^n \to K^m$ bezüglich anderer als der Standardbasen sehr leicht berechnen. Wir betrachten erneut Beispiel 10.5:

Beispiel 13.2 Wir betrachten die durch $F((x, y, z)) = (x - y, 2x + y - z)$ definierte lineare Abbildung $F : \mathbb{R}^3 \to \mathbb{R}^2$ bezüglich der Basen $\tilde{\mathcal{A}} = (v_1, v_2, v_3)$ von $\mathbb{R}^3$ und $\tilde{\mathcal{B}} = (w_1, w_2)$ von $\mathbb{R}^2$ mit

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \; v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \; w_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; w_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Bezüglich der Standardbasen ist $M^m_n(F) = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -1 \end{pmatrix}$ (Bilder der Basisvektoren als Spalten). Die Transformationsmatrizen sind nach Satz 10.2 und wegen $\Phi_{\mathcal{A}} = id$ und $\Phi_{\mathcal{B}} = id$ gegeben als

$$\mathcal{T}^{\mathcal{A}}_{\tilde{\mathcal{A}}} = \Phi^{-1}_{\tilde{\mathcal{A}}} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$$

und

$$\mathcal{T}^{\mathcal{B}}_{\tilde{\mathcal{B}}} = \Phi^{-1}_{\tilde{\mathcal{B}}} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}.$$

Somit gilt

$$M^{\tilde{\mathcal{A}}}_{\tilde{\mathcal{B}}}(F) = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & 1 & 2 \\ -\frac{1}{3} & 0 & -1 \end{pmatrix},$$

in Übereinstimmung mit der ersten Rechnung. ◁

Schließlich können wir den Zusammenhang zwischen Matrixprodukt und Komposition linearer Abbildungen (Abschnitt 11) für beliebige endlich-dimensionale Vektorräume formulieren:

Satz 13.3 *Gegeben seien Vektorräume U, V, W mit Basen $\mathcal{A}, \mathcal{B}, \mathcal{C}$ sowie lineare Abbildungen $G : U \to V$ und $F : V \to W$. Dann gilt $\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(F \circ G) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(F) \cdot \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(G)$.*

Beweis. Wir betrachten das folgende Diagramm:

$$\begin{tikzcd}
K^r \arrow[rrr, "\Phi_{\mathcal{A}}"] \arrow[dd, "A\cdot B"'] \arrow[dr, "B"] & & & U \arrow[dl, "G"'] \arrow[dd, "F\circ G"] \\
& K^n \arrow[r, "\Phi_{\mathcal{B}}"] \arrow[dl, "A"'] & V \arrow[dr, "F"] & \\
K^m \arrow[rrr, "\Phi_{\mathcal{C}}"'] & & & W
\end{tikzcd}$$

mit $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(F) = \Phi_{\mathcal{C}}^{-1} \circ F \circ \Phi_{\mathcal{B}} : K^n \to K^m$ und $B = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(G) = \Phi_{\mathcal{B}}^{-1} \circ G \circ \Phi_{\mathcal{A}} : K^r \to K^n$. Alle Teildiagramme sind kommutativ, damit auch das äußere Viereck. □

Teil IV
Determinanten

14 Definition und Berechnungsmethoden

Bei der Determinante hantelt es sich um eine wichtige Charakterisierung quadratischer Matrizen. Die Determinante ist ein Kriterium für die Invertierbarkeit einer Matrix. Sie tritt außerdem auf beim Eigenwertproblem für Matrizen.

Definition 14.1 Eine Abbildung

$$\det : M(n,K) \to K\ , \qquad \det : A \mapsto \det A$$

heißt *Determinante*, wenn folgendes gilt:

(D1) $\det$ ist linear in jeder Zeile, d.h. ist die i-te Zeile $a_i = \lambda' a_i' + \lambda'' a_i'' \in K^n$, so gilt

$$\det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} = \lambda' \det \begin{pmatrix} \vdots \\ a_i' \\ \vdots \end{pmatrix} + \lambda'' \det \begin{pmatrix} \vdots \\ a_i'' \\ \vdots \end{pmatrix} .$$

(Die mit $\vdots$ symbolisierten Zeilen $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$ sind in jeder der drei Matrizen identisch.)

Insbesondere gilt: Entsteht B aus A durch Multiplikation der i-ten Zeile mit $\lambda \neq 0$, ist also eine Zeilenumformung vom Typ III mit $B = S_i(\lambda) \cdot A$, so ist $\det B = \lambda \cdot \det A$.

(D2) $\det$ ist alternierend, d.h. hat A zwei gleiche Zeilen, so gilt $\det A = 0$.

(D3) $\det$ ist normiert auf $\det E_n = 1$.

Man kann zeigen, daß solche Determinanten existieren. Wir beschränken uns auf den Beweis der Eindeutigkeit der Determinante. Dazu leiten wir aus der Definition weitere Eigenschaften her, die insbesondere eine Berechnungsmethode beinhalten:

Satz 14.2 *Die Determinante hat folgende weitere Eigenschaften:*

(D4) $\det(\lambda \cdot A) = \lambda^n \det A \quad \forall \lambda \in K$.

(D5) Ist eine Zeile von A identisch Null, so folgt $\det A = 0$.

(D6) Die Determinante ändert das Vorzeichen bei Zeilenumformungen von Typ I: Entsteht B aus A durch Zeilenvertauschung, also $B = P_{ik} \cdot A$, so gilt $\det B = -\det A$.

(D7) *Zeilenumformungen von Typ IV lassen die Determinante unverändert: Entsteht B aus A durch Addition der λ-fachen i-ten Zeile zur k-ten Zeile, also $B = Q_{ik}(\lambda) \cdot A$, so gilt* $\det B = \det A$.

(D8) *Ist* $A = \begin{pmatrix} a_{11} & . & \dots & . \\ 0 & a_{22} & \dots & . \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$ *eine obere Dreiecksmatrix, so gilt* $\det A = a_{11}a_{22}\cdots a_{nn}$.

(D9) *Sei $n \geq 2$, und $A \in M(n, K)$ habe die folgende Blockdarstellung:*

$$A = \begin{pmatrix} A_1 & C \\ 0 & A_2 \end{pmatrix} , \qquad A_1 \in M(n_1, K) \ , \ A_2 \in M(n_2, K) \ , \ n_1 + n_2 = n \ .$$

Dann gilt $\det A = \det A_1 \cdot \det A_2$.

(D10) $\det A = 0 \quad \Leftrightarrow \quad \operatorname{rang}(A) < n$.

(D11) *Es gilt der Determinantenmultiplikationssatz* $\det(A \cdot B) = \det A \cdot \det B$ *für alle $A, B \in M(n, K)$. Insbesondere ist* $\det A^{-1} = \frac{1}{\det A}$ *für $A \in GL(n, K)$. Anders formuliert:* $\det : GL(n, K) \to K^*$ *ist ein Gruppenhomomorphismus.*

Beweis. D4) Nach (D1) gilt

$$\det \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} = \lambda \det \begin{pmatrix} a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} = \lambda^2 \det \begin{pmatrix} a_1 \\ a_2 \\ \lambda a_3 \\ \vdots \\ \lambda a_n \end{pmatrix} = \cdots = \lambda^n \det \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

D5) folgt aus (D1) mit $\lambda = 0$

D6) Ist $B = P_{ik}A$, so berechnen wir

$$\det A + \det B = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i \\ \vdots \end{pmatrix}$$

$$\overset{(D2)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_k \\ \vdots \end{pmatrix}$$

$$\overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} \overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i + a_k \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} \overset{(D2)}{=} 0$$

(D7) Ist $B = Q_{ik}(\lambda)A$, so gilt

$$\det B = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k + \lambda a_i \\ \vdots \end{pmatrix} \overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \lambda \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix} \overset{(D2)}{=} \det A$$

(D8) Für Diagonalmatrizen folgt durch wiederholte Anwendung von (D1)

$$\det \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \overset{(D1)}{=} a_{11} \det \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \overset{(D1)}{=} \dots$$
$$\overset{(D1)}{=} a_{11}a_{22}\cdots a_{nn} \det E_n \overset{(D3)}{=} a_{11}a_{22}\cdots a_{nn}$$

Sind für eine allgemeine obere Dreiecksmatrix alle $a_{ii} \neq 0$, so bringen wir sie durch Zeilenumformung vom Typ IV in Diagonalform *mit den gleichen Diagonalelementen*. Ist $a_{ii} = 0$ und $a_{kk} \neq 0$ für alle $k > i$, so führt Zeilenumformung vom Typ IV auf die Zeile $a_i = 0$.

(D9) Wir können Zeilenumformungen verwenden, die die beiden Blöcke nicht mischen. Durch Umformungen nur der ersten n_1 Zeilen und dann nur der letzten n_2 Zeilen erreicht man

$$\det A = (-1)^{s_1+s_2} \det \begin{pmatrix} D_1 & C' \\ 0 & D_2 \end{pmatrix} ,$$

wobei D_1, D_2 obere Dreiecksmatrizen sind und s_1, s_2 die Anzahl der Zeilenumformungen vom Typ I im oberen bzw unteren Block sind. Die Behauptung folgt nun aus (D8).

(D10) Durch elementare Zeilenumformung überführen wir A in eine obere Dreiecksmatrix mit Diagonalelementen $a_{11}, \dots, a_{nn}$. Dann ist $\det A = \pm a_{11} \cdots a_{nn}$. Aus $\det A = 0$ und der Zeilenstufenform folgt $a_{nn} = 0$ and damit $\det A = 0$ nach (D5), also $\operatorname{rang}(A) < n$. Ist umgekehrt $\operatorname{rang}(A) < n$, so führt elementare Zeilenumformung auf eine obere Dreiecksmatrix mit $a_{nn} = 0$, so daß $\det A = 0$.

(D11) Ist $\operatorname{rang}(A) < n$, so ist auch $\operatorname{rang}(AB) < n$ als Dimension des Bildes der Hintereinanderausführung linearer Abbildungen. Dann ist $\det(A) = \det(AB) = 0$. Ist $\operatorname{rang}(A) = n$, so ist A invertierbar und nach Satz 12.1 ein endliches Produkt von Elementarmatrizen. Also ist $\det(AB) = \det(C_1 \cdots C_r \cdot B)$, wobei $C_p = Q_{ik}(\lambda)$ oder $C_p = S_i(\lambda)$ für $1 \leq p \leq r$. Es gilt $\det(Q_{ik}(\lambda)B) = \det B$ nach (D7) und $\det(S_i(\lambda)B) = \lambda \det B$ nach (D1). Also ist $\det(AB) = \lambda_1 \cdots \lambda_s \det B$, wobei λ_p die Koeffizienten in den Umformungen vom Typ III sind. Andererseits ist $\det(A) = \det(AE_n) = \lambda_1 \cdots \lambda_s$, was die Behauptung liefert. □

Zusammengefaßt haben wir damit aus den Axiomen ein Berechnungsverfahren für die Determinante einer Matrix abgeleitet: Wir überführen mit elementaren Zeilenumformungen vom Typ I und IV die Matrix $A \in M(n, K)$ in Zeilenstufenform $\tilde{A} = (a_{ij})$: Sind dabei s Vertauschungen von Zeilen (Typ I) notwendig, dann ist $\det A = (-1)^s a_{11} \cdots a_{nn}$. Das Verfahren beweist die *Eindeutigkeit* der Determinante: Gäbe es zwei Abbildungen $\det$ und $\widetilde{\det}$, die (D1), (D2) und (D3) erfüllen, so folgt aus beiden die Berechnungsvorschrift für die durch Zeilenumformung erhaltene Dreiecksmatrix $\tilde{A} = (a_{ij})$ mit $\det A = \widetilde{\det} A = (-1)^s a_{11} \cdots a_{nn}$.

Beispiel 14.3

$$\det \begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 9 \end{pmatrix} \overset{I_{12}}{=} -\det \begin{pmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 6 & 7 & 9 \end{pmatrix} \overset{IV_{13}(-2)}{=} -\det \begin{pmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 0 & -1 & -1 \end{pmatrix}$$

$$\overset{IV_{23}(1)}{=} -\det \begin{pmatrix} 3 & 4 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = -3\,. \qquad \triangleleft$$

Satz 14.4 *Es gilt* $\det(A^t) = \det(A)$ *für alle* $A \in M(n, K)$.

Beweis. Ist $\det A = 0$, so ist $\operatorname{rang}(A^t) = \operatorname{rang}(A) < n$ und damit $\det(A^t) = 0$. Ist $\det A \neq 0$, so ist A invertierbar und damit darstellbar als endliches Produkt von Elementarmatrizen $A = C_1 \cdots C_r$ mit $C_p = Q_{ik}(\lambda)$ oder $C_p = S_i(\lambda)$. Nach dem Determinatenmultiplikationssatz ist $\det A = \det C_1 \cdots \det C_r$ und $\det A^t = \det C_1^t \cdots \det C_r^t$. Es gilt $(Q_{ik}(\lambda))^t = Q_{ki}(\lambda)$ und $(S_i(\lambda))^t = S_i(\lambda)$. Die Behauptung folgt nun aus $\det(Q_{ik}(\lambda)) = 1$ für alle i, k, λ und $\det(S_i(\lambda)) = \lambda$. □

Daraus ergeben sich mehrere Folgerungen:

Satz 14.5 *(D2') Die Determinate ist linear in jeder Spalte.*
(D3') besitzt $A \in M(n, K)$ *zwei gleiche Spalten, so gilt* $\det A = 0$. □

Satz 14.6 *Für* $A \in M(n, K)$ *sind äquivalent:*

i) $\det A \neq 0$

ii) $\operatorname{rang}(A) = n$

iii) *A ist invertierbar*

iv) *Die Zeilen von A sind linear unabhängig.*

v) *Die Spalten von A sind linear unabhängig.* $\square$

Satz 14.7 *Die Gruppe*

$$SL(n,K) := \{A \in GL(n,K) \ : \ \det A = 1\}$$

ist eine Untergruppe von $GL(n,K)$. Sie heißt die die spezielle lineare Gruppe*.*

Beweis. Für $A, B \in SL(n,K)$ folgt aus den Eigenschaften der Determinante $\det(AB) = 1$ und $\det A^{-1} = 1$. $\square$

Die Determinantenbildung für Matrizen läßt sich sofort auf die Determinate für Endomorphismen endlich-dimensionaler Vektorräume verallgemeinern. Sei V ein n-dimensionaler Vektorraum über K und $F : V \to V$ eine lineare Abbildung (ein Endomorphismus). Wir wählen eine beliebige Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V und stellen F in dieser Basis dar:

$$F(v_j) = \sum_{i=1}^{n} a_{ij} \cdot v_i \ , \qquad a_{ij} \in K \ .$$

Dann ist $A = (a_{ij}) = M^{\mathcal{B}}_{\mathcal{B}}(F)$ die darstellende Matrix von F bezüglich der Basis $\mathcal{B}$. Wir definieren $\det F := \det(M^{\mathcal{B}}_{\mathcal{B}}(F))$. Diese Definition ist sinnvoll, denn sie hängt nicht von der Wahl der Basis ab: Sei $\mathcal{A}$ eine andere Basis von V, dann gilt nach Satz 13.1 $M^{\mathcal{B}}_{\mathcal{B}}(F) = T^{\mathcal{A}}_{\mathcal{B}} \cdot M^{\mathcal{A}}_{\mathcal{A}}(F) \cdot (T^{\mathcal{A}}_{\mathcal{B}})^{-1}$ und deshalb $\det(M^{\mathcal{B}}_{\mathcal{B}}(F)) = \det(M^{\mathcal{A}}_{\mathcal{A}}(F))$.

In direkter Verallgemeinerung der Determinanteneigenschaften gilt:

Satz 14.8 *Sind $F, G \in \mathrm{End}(V)$, so gilt*

i) $\det F \neq 0 \quad \Leftrightarrow \quad F \in \mathrm{Aut}(V)$

ii) $F \in \mathrm{Aut}(V) \quad \Rightarrow \quad \det F^{-1} = \frac{1}{\det F}$

iii) $\det(F \circ G) = \det F \cdot \det G$

Eine weitere nützliche Anwendung ist die Definition der *Orientierung* von Automorphismen und von Basen.

Definition 14.9 Sei V ein endlich-dimensionaler Vektorraum über K. Ein Automorphismus $F \in \mathrm{Aut}(V)$ heißt *orientierungstreu*, falls $\det F > 0$, ansonsten *orientierngsuntreu*.

Zwei Basen $\mathcal{A}$ und $\mathcal{B}$ von V heißen *gleich orientiert*, wenn für die darstellende Matrix der Identität gilt $\det(M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)) > 0$, ansonsten *ungleich orientiert*.

Offenbar gilt:

Satz 14.10 *Die Menge*

$$\mathrm{Aut}^+(V) := \{F \in \mathrm{Aut}(V) \ : \quad \det F > 0\}$$

der orientierungstreuen Automorphismen von V ist eine Untergruppe von $\mathrm{Aut}(V)$*. Insbesondere ist*

$$GL^+(m, K) := \{A \in GL(n, K) \ : \quad \det A > 0\}$$

eine Untergruppe von $GL(n, K)$.

15 Laplacescher Entwicklungssatz und komplementäre Matrix

In Beweisen und für speziell gewählte Matrizen sind auch rekursive und abstrakte Berechnungsformeln nützlich:

Satz 15.1 (Entwicklungssatz von Laplace) *Ist $A = (a_{ij}) \in M(n, K)$ und seien die Matrizen $A_{ij} \in M(n-1, K)$ aus A durch Streichen der i-ten Zeile und der j-ten Spalte erhalten. Dann gilt für beliebiges $1 \leq i \leq n$*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \qquad \textit{Entwicklung nach der } i\textit{-ten Zeile}$$

und für beliebiges $1 \leq j \leq n$

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \qquad \textit{Entwicklung nach der } j\textit{-ten Spalte.}$$

Beweis. (für die Zeilenentwicklung). Wir schreiben die i-te Zeile $a_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ von $A = (a_{ij})$ als Linearkombination der kanonischen Basen

$$a_i = a_{i1}(1, 0, 0, \ldots, 0) + a_{i2}(0, 1, 0, \ldots, 0) + \cdots + a_{in}(0, 0, \ldots, 0, 1)$$

Anwenden von (D1) ergibt

$$\det A = \sum_{j=1}^{n} a_{ij} \det A'_{ij} \ ,$$

$$A'_{ij} = \begin{pmatrix} a_{11} & \ldots & a_{1,j-1} & a_{1,j} & a_{1,j+1} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \ldots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \ldots & a_{i-1,n} \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ a_{i+1,1} & \ldots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \ldots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \ldots & a_{nn} \end{pmatrix} .$$

In jeder der Matrizen A'_{ij} bringen wir mit Zeilenumformungen vom Typ IV (welche die Determinante nicht ändern) alle anderen Einträge der j-ten Spalte auf 0:

$$\det A'_{ij} = \det A''_{ij} \;, \quad A''_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & 0 & a_{1,j+1} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \dots & a_{i-1,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{i+1,1} & \dots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & 0 & a_{n,j+1} & \dots & a_{nn} \end{pmatrix} .$$

Durch Vertauschen mit jeweils benachbarten Zeilen bringen wir die i-te Zeile von A''_{ij} an die j-te Stelle, wobei sich die Reihenfolge aller anderen Zeilen nicht ändert. Dazu sind $|i-j|$ Zeilenvertauschungen erforderlich. Die 1 steht nun auf der Diagonale: Für $i < j$ erhalten wir:

$$\det A''_{ij} = (-1)^{|i-j|} \det A'''_{ij} \;, \quad A'''_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & 0 & a_{1,j+1} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{j,1} & \dots & a_{j,j-1} & 0 & a_{j,j+1} & \dots & a_{j,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{j+1,1} & \dots & a_{j+1,j-1} & 0 & a_{j+1,j+1} & \dots & a_{j+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & 0 & a_{n,j+1} & \dots & a_{nn} \end{pmatrix} .$$

Durch elementare Zeilenumformung vom Typ I und IV überführen wir A'''_{ij} in eine obere Dreiecksmatrix und lesen die Determinante ab. Bei diesen Umformungen wird an der j-ten Zeile und Spalte von A'''_{ij} nichts geändert. Deshalb stimmt $\det A'''_{ij}$ mit der Determinante jener Matrix $A_{ij} \in M(n-1, K)$ überein, die aus A'''_{ij} durch Weglassen des Kreuzes aus j-ter Zeile und j-ter Spalte erhalten wird. Dieselbe Matrix A_{ij} entsteht aber auch aus A'_{ij} und damit auch aus A selbst durch Weglassen der i-ten Zeile und der j-ten Spalte. Unter Verwendung von $(-1)^{|i-j|} = (-1)^{i+j}$ folgt die Behauptung. Analog für die Spaltenentwicklung. □

Wir können nun über den Laplaceschen Entwicklungssatz die Existenz der Determinante beweisen:

Satz 15.2 *Für alle* $n \in \mathbb{N} \setminus \{0\}$ *gibt es eine Abbildung* $\det : M(n, K) \to K$ *mit den Eigenschaften (D1), (D2) und (D3) aus Definition 14.1.*

Beweis. Durch Induktion nach n. Für $n = 1$ setzen wir $\det(a) := a$. Dann ist (D2) eine leere Aussage, (D1) und (D3) sind klar. Sei die Existenz bis $M(n-1, K)$ bewiesen. Wir definieren $\det : M(n, K) \to K$ über den Laplaceschen Entwicklungssatz und Entwicklung nach der j-ten *Spalte*. Zu zeigen sind $(D1) - (D3)$.

(D1) Betrachtet werde die k-te Zeile von $A \in M(n, K)$. Es gilt $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = (-1)^{k+j} a_{kj} A_{kj} + \sum_{i \neq k} (-1)^{i+j} a_{ij} \det A_{ij}$. Im ersten Term ist A_{kj} unabhängig von der k-ten Zeile, und a_{kj} ist linear. In jedem anderen Term $i \neq k$ ist $\det A_{ij}$ linear in der k-ten Zeile nach Induktionsvoraussetzung, und a_{ij} ist unabhängig von der k-ten Zeile.

(D2) Seien die k-te und l-te Zeile von $A \in M(n, K)$ gleich, und $k < l$. Ist $i \neq k$ und $i \neq l$, so sind die entsprechenden Zeilen auch in $A_{ij} \in M(n-1, K)$ gleich, damit $\det A_{ij} = 0$. Im Entwicklungssatz verbleibt $\det A = (-1)^{k+j} a_{kj} \det A_{kj} + (-1)^{l+j} a_{lj} \det A_{lj}$. Die Annahme (Gleichheit det k-ten und l-ten Zeile von A) ergibt zunächst $a_{kj} = a_{lj}$. Ist $n = 2$, so ist $k = 1$ und $l = 2$, damit $(-1)^{k+j} = -(-1)^{l+j}$ und $\det A_{1j} = \det A_{2j}$ wegen Gleichheit beider Zeilen von $A \in M(n, K)$. Sei also $n \geq 3$ und $a_i \in K^n$ die i-te Zeile von A mit $a_k = a_l =: b \in K^n$. Seien $a_i', b' \in K^{n-1}$ die aus a_i, b durch Weglassen des j-ten Eintrags entstehenden Zeilenvektoren. Dann ist

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_{k-1} \\ b \\ a_{k+1} \\ \vdots \\ a_{l-1} \\ b \\ a_{l+1} \\ \vdots \\ a_n \end{pmatrix}, \qquad A_{kj} = \begin{pmatrix} a_1' \\ \vdots \\ a_{k-1}' \\ a_{k+1}' \\ \vdots \\ a_{l-1}' \\ b' \\ a_{l+1}' \\ \vdots \\ a_n' \end{pmatrix}, \qquad A_{lj} = \begin{pmatrix} a_1' \\ \vdots \\ a_{k-1}' \\ b' \\ a_{k+1}' \\ \vdots \\ a_{l-1}' \\ a_{l+1}' \\ \vdots \\ a_n' \end{pmatrix}.$$

Durch $l-1-k$ Vertauschungen benachbarter Zeilen bringen wir die $(l-1)$-te Zeile b' von A_{kj} in die k-te Zeile, alle Zeilen $a_{k+1}', \ldots, a_{l-1'}$ verschieben sich um eine Zeile nach unten. Das Ergebnis dieser $l-1-k$ Zeilenvertauschungen ist A_{lj}, d.h. es gilt $\det A_{kj} = (-1)^{l-1-k} \det A_{lj}$ und damit $(-1)^{k+j} \det A_{kj} - (-1)^{l+k} \det A_{lj}$. Das liefert (D2).

(D3) Es gilt $\det E_n = \sum_{i=1}^{n} (-1)^{i+j} \delta_{ij} \det(E_n)_{ij} = (-1)^{2j} \det(E_n)_{jj} = \det(E_{n-1}) = 1$. $\square$

Die Berechnung der Determinante von $A \in M(n, K)$ wird durch Satz 15.1 auf die Berechnung der Determinanten kleinerer Matrizen zurückgeführt. Als Beispiel berechnen wir

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (-1)^{1+1} a_{11} \det(a_{22}) + (-1)^{1+2} a_{12} \det(a_{21}) = a_{11}a_{22} - a_{12}a_{21} \;.$$

Einfache Rechenregel: Für 2×2-Matrizen ist die Determinante gleich dem Produkt der Hauptdiagonalelemente minus dem Produkt der Nebendiagonalelemente. Eine ähnliche graphische Rechenregel gibt es auch für 3×3-Matrizen (Regel von Sarrus). Ein Analogon für $M(n, K)$ mit $n \geq 4$ wäre falsch!

Beispiel 15.3 Wir berechnen erneut Beispiel 14.3 durch Entwicklung nach der 1. Zeile:

$$\begin{aligned} &\det \begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 9 \end{pmatrix} = 0 \cdot \det \begin{pmatrix} 4 & 5 \\ 7 & 9 \end{pmatrix} + (-1) \cdot \det \begin{pmatrix} 3 & 5 \\ 6 & 9 \end{pmatrix} + 2 \cdot \det \begin{pmatrix} 3 & 4 \\ 6 & 7 \end{pmatrix} \\ &= -1 \cdot 3 \cdot 9 + 1 \cdot 6 \cdot 5 + 2 \cdot 3 \cdot 7 - 2 \cdot 6 \cdot 4 = -27 + 30 + 42 - 48 = -3 \;, \end{aligned}$$

in Übereinstimmung mit der ersten Rechnung. $\triangleleft$

Allgemein wird durch den Entwicklungssatz die Determinante von $A = (a_{ij})$ rekursiv durch Summen und Differenzen von Produkten $a_{i_1 j_1} \cdots a_{i_n j_n}$ der Einträge a_{ij} berechnet. Durch Entwicklung nach der jeweils obersten Zeile sieht man, daß in jedem dieser Produkte jeder Zeilenindex genau einmal vorkommt. Andererseits kommt auch jeder Spaltenindex genau einmal vor: Damit a_{ij} in einem Produkt vorkommen kann, darf zuvor kein a_{kj} aufgetreten sein, da sonst die j-te Spalte gestrichen wäre. Nach dem Auftreten von a_{ij} wird die j-te Spalte weggelassen und weitere a_{kj} können im Produkt nicht vorkommen. Eine solche injektive/surjektive/bijektive Zuordnung eines Spaltenindex $j = \sigma(i) \in \{1, \dots, n\}$ zu jedem Zeilenindex $i \in \{1, \dots, n\}$ heißt *Permutation*. Ein Beispiel ist durch folgende Tabelle gegeben:

i	1	2	3	4	5
$\sigma(i)$	2	5	3	1	4

Sei S_n die Menge (und Gruppe) aller Permutationen der Menge $\{1, 2, \dots, n\}$, dann gilt

$$\det A = \sum_{\sigma \in S_n} \lambda(\sigma)\; a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \;.$$

Dabei ist $\lambda(\sigma) \in \pm 1$, denn jede Permutation kommt vor und zwar mit Vorfaktor $+1$ oder -1 entsprechend dem Entwicklungssatz. Man kann zeigen, daß $\lambda(\sigma) = \operatorname{sign}(\sigma)$ das *Vorzeichen der Permutation* ist, das wie folgt berechnet werden kann: Eine Permutation τ heißt *Transposition*, wenn sie zwei Elemente aus $\{1, \dots, n\}$ austauscht und alle anderen an ihrer Stelle beläßt. Jede Permutation kann als (nicht eindeutige) Hintereinanderausführung von Transpositionen erhalten werden. Dann ist

$$\operatorname{sign}(\sigma) = (-1)^{\text{Zahl der Transpositionen in } \sigma} \;.$$

Das Vorzeichen ist eindeutig, auch wenn die Zahl der Transpositionen selbst nicht eindeutig ist.

Alternativ kann man das Vorzeichen aus der Zahl der *Fehlstellen* der Permutation ablesen. Eine Fehlstelle ist ein Paar (i, j) mit $i < j$ und $\sigma(i) > \sigma(j)$. Dann ist

$$\operatorname{sign}(\sigma) = (-1)^{\text{Zahl der Fehlstellen in } \sigma} .$$

Damit gilt (insgesamt hier ohne Beweis)

Satz 15.4 (*Formel von Leibniz*)

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma)\ a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} .$$

Es gibt $n!$ Permutationen $\sigma \in S_n$, so daß die Zahl der Produkte in der Formel von Leibniz mit wachsendem n sehr groß wird. Deshalb ist die Formel von Leibniz vor allem aus theoretischer Sicht bedeutsam. Sind z.B. die Einträge der Matrix stetige/differenzierbare Funktionen, so ist auch die Determinante eine stetige/differenzierbare Funktion. Für praktische Berechnungen sind Zeilenumformungen geeigneter.

Definition 15.5 Sei $A = (a_{ij}) \in M(n, K)$. Dann heißt die Matrix

$$A^\sharp = (a^\sharp_{ij}) \in M(n, K) , \qquad a^\sharp_{ij} := (-1)^{i+j} \det A_{ji} .$$

die zu A *komplementäre Matrix*, wobei $A_{ji} \in M(n-1, K)$ durch Streichen der j-ten Zeile und der i-ten Spalte entsteht.

Man beachte die Vertauschung der Reihenfolge von i, j: Im Matrixelement $a^\sharp_{ij}$ steht A_{ji}, nicht A_{ij}!

Satz 15.6 *Ist $A \in M(n, K)$ und sei $A^\sharp$ die zu A komplementäre Matrix, dann gilt $A \cdot A^\sharp = A^\sharp \cdot A = (\det A)E_n$. Insbesondere gilt $A^{-1} = \frac{1}{\det A} A^\sharp$ für invertierbare Matrizen $A \in GL(n, K)$.*

Beweis. Wir berechnen die Komponenten von $A^\sharp \cdot A$:

$$(A^\sharp \cdot A)_{kj} = \sum_{i=1}^{n} a^\sharp_{ki} a_{ij} = \sum_{i=1}^{n} (-1)^{i+k} a_{ij} \det A_{ik} .$$

Ist $j = k$ beliebig, so entsteht gerade der Spaltenentwicklungssatz von Laplace: $\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$. Für $j \neq k$ betrachten wir die Matrix $B = (b_{il}) \in M(n, K)$, die aus A entsteht, wenn man die k-te Spalte von A durch die j-te Spalte von A ersetzt. Es ist also $b_{il} = a_{il}$ für $l \neq k$ und $b_{ik} = a_{ij}$. Da B zwei gleiche Spalten besitzt, ist $\det B = 0$. Wir entwickeln $\det B$ nach der k-ten Spalte:

$$0 = \det B = \sum_{i=1}^{n} (-1)^{i+k} b_{ik} \det B_{ik} = \sum_{i=1}^{n} (-1)^{i+k} a_{ij} \det A_{ik}$$

wegen $A_{ik} = B_{ik}$. Folglich ist $(A^\sharp \cdot A)_{kj} = (\det A)\delta_{kj}$. $\square$

Mit $A^{-1} = \frac{1}{\det A} A^\sharp$ für $A \in GL(n, K)$ haben wir eine weitere Methode zur Berechnung der inversen Matrix kennengelernt. Für $n > 3$ ist diese Methode jedoch sehr aufwendig. Allerdings ist diese abstrakte Darstellung in der Analysis sehr nützlich, denn die Determinantenbildung hängt als Produkt der a_{ij} stetig und sogar differenzierbar von den Einträgen a_{ij} ab. Daraus folgt, daß für $A \in GL(n, K)$ die Abbildung $A \mapsto A^{-1}$ differenzierbar ist.

Eine weitere Anwendung der Determinanten besteht in einem Lösungsverfahren für lineare Gleichungssysteme mit invertierbaren Matrizen:

Satz 15.7 (Cramersche Regel) *Seien $A = (a_{ij}) \in GL(n, K)$ und $b \in K^n$ gegeben und sei $x = (x_1, \dots, x_n)^t \in K^n$ die eindeutig bestimmte Lösung des linearen Gleichungssystems $A \cdot x = b$. Bezeichnen wir mit $a_1, \dots, a_n$ die Spalten von A, also $a_j = (a_{1j}, \dots, a_{nj})^t$, dann gilt*

$$x_j = \frac{\det(a_1, \dots, a_{j-1}, b_j, a_{j+1}, \dots, a_n)}{\det A} .$$

Beweis. Die endeutig bestimmte Lösung des linearen Gleichungssystems ist durch $x = A^{-1} \cdot b$ gegeben. In Komponenten gilt damit

$$x_j = \sum_{i=1}^n (A^{-1})_{ji} b_i = \frac{1}{\det A} \sum_{i=1}^n (-1)^{i+j} b_i (\det A_{ij}) .$$

Nach dem Spaltenentwicklungssatz von Laplace ist $\sum_{i=1}^n (-1)^{i+j} b_i (\det A_{ij})$ gerade die Determinante einer Matrix $B = (b_{il})$, deren Einträge auf der j-ten Spalte durch $b_{ij} = b_i$ gegeben sind und deren andere Spalten identisch sind mit den Spalten von A. $\square$

Wieder liegt die Bedeutung der Cramerschen Regel in theoretischen Betrachtungen wie der Schlußfolgerung, daß die Lösung des linearen Gleichungssystems $Ax = b$ für $A \in GL(n, K)$ differenzierbar von der rechten Seite $b \in K^n$ sowie den Einträgen der Matrix $A \in GL(n, K)$ abhängt.

Zum Abschluß geben wir ohne Beweis noch die Verallgemeinerung von $\det(A \cdot B) = \det A \cdot \det B$, hier mit $A, B \in M(n, K)$, auf Produkte nichtquadratischer Matrizen an:

Satz 15.8 (Binet-Cauchy) *Es seien $A = (a_1, \dots, a_{n+k}) \in M(n \times (n + k), K)$ und $B = (b_1, \dots, b_{n+k}) \in M(n \times (n + k), K)$ zwei rechteckige Matrizen, gebildet aus den Spaltenvektoren $a_i, b_i \in K^n$, und $k \in \mathbb{N}$. Für $1 \leq m_1 < m_2 < \dots < m_n \leq$*

$n+k$ seien quadratische Matrizen $A^{m_1m_2\ldots m_n} := (a_{m_1}, a_{m_2}, \ldots, a_{m_n}) \in M(n,K)$ und $B^{m_1m_2\ldots m_n} := (b_{m_1}, b_{m_2}, \ldots, b_{m_n}) \in M(n,K)$ definiert. Dann gilt

$$\det(A \cdot B^t) = \sum_{1 \leq m_1 < m_2 < \cdots < m_n \leq n+k} (\det A^{m_1m_2\ldots m_n})(\det B^{m_1m_2\ldots m_n}) \, .$$

Die Summe läuft über die $\binom{n+k}{k} = \frac{(n+k)!}{n!k!}$ verschiedenen Möglichkeiten, n der $n+k$ Spalten der Matrizen auszuwählen.

Ein Beweis findet sich z.B. in G. Fischer: Lineare Algebra, Kapitel 3.3. Man kann sich auf Matrizen $A, B \in M(n \times l, K)$ mit $l \geq n$ beschränken, da sich zeigen läßt, daß für $A, B \in M(n \times l, K)$ mit $l < n$ stets $\det(A \cdot B^t) = 0$ gilt. Aus Satz 15.8 folgt insbesondere $\det(A \cdot A^t) = \sum_{1 \leq m_1 < m_2 < \cdots < m_n \leq n+k} (\det A^{m_1m_2\ldots m_n})^2$ (Gramsche Determinante), d.h. für reellwertige Matrizen ist $\det(A \cdot A^t) \geq 0$.

Teil V
Eigenwerte und Diagonalisierbarkeit

Wir hatten im Abschnitt über Matrizen (Satz 10.7) gesehen, daß zu jeder linearen Abbildung $F : V \to W$ angepaßte Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W existieren, so daß die darstellende Matrix die Form $M^{\mathcal{A}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ hat. Natürlich können wir $W = V$ wählen und dann entsprechende Basen $\mathcal{A}, \mathcal{B}$ von V finden. Die Frage ist: Gibt es zu einem Endomorphismus $F : V \to V$ auch *eine* Basis $\mathcal{B}$ von V, so daß für die darstellende Matrix gilt: $M^{\mathcal{B}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$? Die Antwort ist: Nein!

16 Definitionen und Eigenschaften

Es geht nun darum, zu gegebenem Endomorphismus die Basis so sinnvoll wie möglich zu wählen.

Definition 16.1 Sei $F : V \to V$ Endomorphismus eines Vektorraums V über K. Ein Skalar $\lambda \in K$ heißt *Eigenwert* von F, wenn es einen Vektor $v \neq 0$ von V gibt, so daß $F(v) = \lambda \cdot v$. Jeder Vektor $v \neq 0$ von V mit $F(v) = \lambda \cdot v$ heißt *Eigenvektor* von F zum Eigenwert λ.

Zu beachten ist, daß es zu einem Eigenwert λ mehrere Eigenvektoren geben kann.

Definition 16.2 Ein Endomorphismus $F \in \mathrm{End}(V)$ heißt *diagonalisierbar*, wenn es eine Basis von V aus Eigenvektoren von F gibt.

In diesem Fall gilt:

Satz 16.3 *Ist* $\dim(V) = n$, *so ist* $F \in \mathrm{End}(V)$ *genau dann diagonalisierbar, wenn es eine Basis* $\mathcal{B} = (v_1, \ldots, v_n)$ *von* V *gibt, so daß* $M^{\mathcal{B}}_{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$.

Beweis. Für die darstellende Matrix $M^{\mathcal{B}}_{\mathcal{B}}(F) = (a_{ij})$ gilt $F(v_j) = \sum_{i=1}^{n} a_{ij} \cdot v_i$. Damit sind die v_i die Eigenvektoren zu den Eigenwerten λ_i. □

Das ist die optimalste Situation für eine Basis zu gegebenem Endomorphismus. Jedoch ist nicht klar, daß jeder Endomorphismus auch diagonalisierbar ist. Zunächst untersuchen wir, ob die Eigenvektoren linear unabhängig sind:

Satz 16.4 *Sei* $F \in \mathrm{End}(V)$ *ein Endomorphismus und seien* $v_1, \ldots, v_m$ *Eigenvektoren zu paarweise verschiedenen Eigenwerten* $\lambda_1, \ldots, \lambda_m$ *von* V*. Dann sind die* $(v_1, \ldots, v_m)$ *linear unabhängig.*

Beweis. Wir beweisen den Satz durch Induktion nach der Anzahl m linear unabhängiger Eigenvektoren. Für $m = 1$ ist nichts zu zeigen. Angenommen, $v_2, \ldots, v_m$ seien linear unabhängig. Wir betrachten

$$\mu_1 v_1 + \cdots + \mu_m v_m = 0$$

Anwenden von F einerseits und Subtraktion des λ_1-fachen dieser Gleichung ergibt

$$0 = \mu_1(\lambda_1 - \lambda_1)v_1 + \mu_2(\lambda_2 - \lambda_1)v_2 + \cdots + \mu_m(\lambda_m - \lambda_1)v_2 \ .$$

Da $\lambda_1 \neq \lambda_i$ für $2 \leq i \leq m$ und $(v_2, \ldots, v_m)$ linear unabhängig, folgt $\mu_j = 0$ für alle $1 \leq j \leq m$. Damit ist $(v_1, \ldots, v_m)$ linear unabhängig. □

Als direkte Konsequenz ergibt sich:

Satz 16.5 *Seien* $\lambda_1, \ldots, \lambda_m$ *paarweise verschiedene Eigenwerte eines Endomorphismus* $F \in \mathrm{End}(V)$ *und sei* $\dim(V) = n$*. Dann gilt:*

i) $m \leq n$

ii) *Ist* $m = n$*, dann ist* F *diagonalisierbar.*

Beweis. i) Für $m > n$ wären nach dem vorigen Satz mehr als n Vektoren aus V, nämlich Eigenvektoren zu $\lambda_1, \ldots, \lambda_m$, linear unabhängig. Das ist durch die Dimension ausgeschlossen.

ii) Ist $\dim(V) = n$, dann bilden n linear unabhängige Vektoren eine Basis. Damit ist $(v_1, \ldots, v_n)$ eine Basis aus Eigenvektoren, und F ist diagonalisierbar. □

Es gibt natürlich Endomorphismen mit weniger als $\dim(V)$ paarweise verschiedenen Eigenwerten, die trotzdem diagonalisierbar sind. Ein Beispiel ist $\mathrm{id}_V \in \mathrm{End}(V)$. Es gibt nur einen Eigenwert $\lambda = 1$, aber jede Basis von V diagonalisiert id_V. Die Untersuchung der Eigenräume zu gegebenem Eigenwert ist also entscheidend:

Definition 16.6 Ist $F \in \mathrm{End}(V)$ und ist $\lambda \in K$ Eigenwert von F, dann heißt

$$\mathrm{Eig}(F; \lambda) := \{v \in V \ : \ F(v) = \lambda \cdot v\}$$

der *Eigenraum* von F zum Eigenwert λ.

Zu beachten ist, daß der Nullvektor im Eigenraum liegt, $0 \in \mathrm{Eig}(F; \lambda)$, obwohl er kein Eigenvektor ist. Der Grund ist i) im folgenden Satz:

Satz 16.7 *Sei* $\mathrm{Eig}(F;\lambda)$ *der Eigenraum von* $F \in \mathrm{End}(V)$ *zum Eigenwert* λ. *Dann gilt:*

i) $\mathrm{Eig}(F;\lambda) \subset V$ *ist Untervektorraum*

ii) λ *ist Eigenwert von* $F \quad \Leftrightarrow \quad \mathrm{Eig}(F;\lambda) \neq \{0\} \quad \Leftrightarrow \quad \dim(\mathrm{Eig}(F;\lambda)) > 0$

iii) $\mathrm{Eig}(F;\lambda) \setminus \{0\}$ *ist die Menge der Eigenvektoren von* F *zum Eigenwert* λ

iv) $\mathrm{Eig}(F;\lambda) = \ker(F - \lambda \cdot \mathrm{id}_V)$

v) *Ist* $\lambda_1 \neq \lambda_2$, *so folgt* $\mathrm{Eig}(F;\lambda_1) \cap \mathrm{Eig}(F;\lambda_2) = \{0\}$

Beweis. i) Sind $v_1, v_2 \in \mathrm{Eig}(F;\lambda)$ und $\mu_1, \mu_2 \in K$, so folgt

$$F(\mu_1 v_1 + \mu_2 v_2) = \mu_1 F(v_1) + \mu_2 F(v_2) = \mu_1(\lambda v_1) + \mu_2(\lambda v_2) = \lambda(\mu_1 v_1 + \mu_2 v_2)$$

und damit $\mu_1 v_1 + \mu_2 v_2 \in \mathrm{Eig}(F;\lambda)$.

ii) Es gibt ein $v \neq 0$ mit $F(v) = \lambda v$. Dann ist $v \in \mathrm{Eig}(F;\lambda) \neq \{0\}$.

iii) ist klar

iv) Sei $v \in \mathrm{Eig}(F;\lambda)$. Dann gilt:

$$(F - \lambda \cdot \mathrm{id}_V)(v) = \lambda v - \lambda v = 0 ,$$

also $v \in \ker(F - \lambda \cdot \mathrm{id}_V)$. Ebenso folgt die Umkehrung.

v) Ist $\lambda_1 \neq \lambda_2$ und $v \in \mathrm{Eig}(F;\lambda_1)$, so ist

$$(F - \lambda_2 \cdot \mathrm{id}_V)(v) = (\lambda_1 - \lambda_2)v .$$

Damit ist $v \in \ker(F - \lambda_2 \cdot \mathrm{id}_V)$ genau dann, wenn $v = 0$. □

17 Das charakteristische Polynom

Es geht nun um ein Verfahren zur Bestimmung der Eigenwerte und Eigenvektoren eines Endomorphismus, wobei die Determinante eine entscheidende Rolle spielt.

Satz 17.1 *Sei* $F \in \mathrm{End}(F)$ *Endomorphismus eines endlich-dimensionalen Vektorraums* V. *Ein Skalar* $\lambda \in K$ *ist genau dann Eigenwert von* F, *wenn*

$$\det(F - \lambda \cdot \mathrm{id}_V) = 0 .$$

Beweis. Nach Satz 16.7.ii) und Satz 16.7.iv) ist λ genau dann Eigenwert von F, wenn $\dim(\ker(F - \lambda \cdot \mathrm{id}_V)) > 0$. Das ist gleichbedeutend mit

$$\dim(\mathrm{im}(F - \lambda \cdot \mathrm{id}_V)) = \mathrm{rang}(F - \lambda \cdot \mathrm{id}_V) < \dim(V) .$$

Nach der Determinanteneigenschaft (D10) ist diese Eigenschaft äquivalent zu $\det(F - \lambda \cdot \mathrm{id}_V) = 0$. □

Die Determinante eines Endomorphismus $F \in \mathrm{End}(V)$ berechnet sich als die Determinante der darstellenden Matrix $M_{\mathcal{A}}^{\mathcal{A}}(F) = A = (a_{ij}) \in M(n, K)$ bezüglich einer beliebigen Basis $\mathcal{A} = (w_1, \ldots, w_n)$ von V. Es gilt

$$M_{\mathcal{A}}^{\mathcal{A}}(F - t \cdot \mathrm{id}_V) = A - t \cdot E_n \ ,$$

denn $M_{\mathcal{A}}^{\mathcal{A}}$ ist eine lineare Abbildung (Satz 10.3) und $M_{\mathcal{A}}^{\mathcal{A}}(\mathrm{id}_V) = E_n$ unabhängig von der Wahl der Basis. Damit ist die Suche nach Eigenwerten von $F \in \mathrm{End}(V)$ zurückgeführt auf die Bestimmung der *Nullstellen* der Abbildung

$$P_A : K \to K \ : \qquad P_A : t \mapsto \det(A - t \cdot E_n) \ .$$

Nach der Formel von Leibniz gilt

$$\det(A - t \cdot E_n) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t) + Q \ ,$$

wobei in Q aus Summen und Differenzen von Produkten der Matrixelemente besteht, in denen mindestens zwei Nichtdiagonalelemente a_{ij} mit $i \neq j$ auftreten (welche kein t beinhalten): Wenn in einem solchen Produkt a_{ij} auftritt, dann berechnen sich nach dem Entwicklungssatz von Laplace die weiteren Faktoren zu $(-1)^{i+j} \det A_{ij}$, aber A_{ij} enthält *nicht* die Diagonalelemente $a_{ii} - t$ und $a_{jj} - t$. Ausmultiplikation der Produkte und Ordnen nach Potenzen von t ergibt:

$$\begin{aligned} \det(A - t \cdot E_n) &= \sum_{k=0}^{n} \alpha_k t^k \ , \\ &\alpha_n = (-1)^n \ , \ \alpha_{n-1} = (-1)^{n-1}(a_{11} + \cdots + a_{nn}) \ , \ \ldots , \ \alpha_0 = \det A \ . \end{aligned}$$

Dabei heißt im zweithöchsten Term $\mathrm{tr}(A) := a_{11} + \cdots + a_{nn}$ die *Spur* der Matrix $A \in M(n, K)$. Der niedrigste Term ist unabhängig von t und stimmt damit mit der Rechnung für $t = 0$ überein, was gerade die Determinante von A ergibt. Die anderen Koeffizienten $\alpha_1, \ldots, \alpha_{n-2}$ sind schwieriger zu charakterisieren. Insgesamt entsteht mit

$$\sum_{k=0}^{n} \alpha_k t^k \ , \qquad \alpha_k \in K$$

ein *Polynom in t mit Koeffizienten im Körper K vom Grad n*. In Verallgemeinerung von Abschnitt 9 aus dem letzten Semester bezeichnen wir mit $K[t]$ den Vektorraum der Polynome in t mit Koeffizienten aus K.

Definition 17.2 Das spezielle Polynom $P_A(t) = \det(A - t \cdot E_n) \in K[t]$ heißt das *charakteristische Polynom* der Matrix $A \in M(n, K)$.

Sinnvollerweise heißt ein $\lambda \in K$ *Nullstelle* eines Polynoms $f \in K[t]$, wenn $f(\lambda) = 0$. In Verbindung mit Satz 17.1 gilt:

Satz 17.3 *Sei $F \in \mathrm{End}(V)$ Endomorphismus eines endlich-dimensionalen Vektorraums. Dann sind die Nullstellen des charakteristischen Polynoms* $\det(F - t \cdot \mathrm{id}_V) \in K[t]$ *die Eigenwerte von F.* □

Wir erinnern an einige Eigenschaften, die wir für $K = \mathbb{C}$ im letzten Semester bewiesen haben, deren Beweis sich aber auf beliebige Körper überträgt. Aus der Eindeutigkeit der Division mit Rest ergab sich:

Satz 17.4 *Ist $\lambda \in K$ eine Nullstelle von $f \in K[t]$, so gibt es ein eindeutig bestimmtes Polynom $g \in K[t]$ mit*

i) $f = (t - \lambda) \cdot g$

ii) $\deg(g) = \deg(f) - 1$

Der Grad deg eines Polynoms $f = \sum_{i=0}^{n} a_i t^i \in K[t]$ war der höchste Exponent $\deg(f) = \max(i \ : \ a_i \neq 0)$, mit $\deg(f) = -\infty$ für $f = 0$. Ist $\lambda \in K$ Nullstelle von $f \in K[t]$ und sei $g \in K[t]$ durch $f = (t - \lambda) \cdot g$ definiert, dann kann das gleiche λ auch Nullstelle von g sein. Wir sagen, daß λ eine vielfache Nullstelle ist:

Definition 17.5 Ist $f \in K[t]$ verschieden vom Nullpolynom und $\lambda \in K$, so heißt

$$\mu(f; \lambda) := \max\{r \in \mathbb{N} \ : \ f = (t - \lambda)^r \cdot g \text{ für } g \in K[t]\}$$

die *Vielfachkeit* der Nullstelle λ.

Ist $f = (t - \lambda)^r \cdot g$ und $r = \mu(f; \lambda)$, so ist $g(\lambda) \neq 0$. Durch wiederholtes Abdividieren der Nullstellen läßt sich jedes Polynom $f \in K[t]$ darstellen als

$$f = (t-\lambda_1)^{r_1}(t-\lambda_2)^{r_2} \cdots (t-\lambda_k)^{r_k} \cdot g \ , \qquad \deg(g) = \deg(f) - r_1 - r_2 - \cdots - r_k \geq 0 \ ,$$

wobei g ein Polynom *ohne Nullstellen* ist. Insbesondere besitzt jedes Polynom n-ten Grades höchstens n mit Vielfachheit gezählte Nullstellen (damit auch höchstens n paarweise verschiedene Nullstellen). Ist $\deg(g) = 0$, dann sagen wir, daß f *in Linearfaktoren zerfällt.* Von größter Bedeutung ist

Theorem 17.6 (Fundamentalsatz der Algebra) *Jedes Polynom $f \in \mathbb{C}[t]$ zerfällt in Linearfaktoren, d.h. es gibt $a \in \mathbb{C}$ und (nicht notwendig verschiedene) $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ mit $n = \deg(f)$, so daß*

$$f(t) = a(t - \lambda_1) \cdots (t - \lambda_n) \ .$$

Im reellen Fall $K = \mathbb{R}$ zerfällt ein Polynom im allgemeinen *nicht* in Linearfaktoren. Das einfachste Beispiel ist $f(t) = t^2 + 1$, welches keine reelle Nullstelle besitzt. Man kann aber $\mathbb{R} \subset \mathbb{C}$ ausnutzen und somit für jedes reelle Polynom $f \in \mathbb{R}[t]$ mit $\deg(f) = n$ genau n *komplexe Nullstellen* $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ (gezählt mit Vielfachheit) finden. Es gibt zwei Möglichkeiten: Ist $\lambda_i \in \mathbb{R} \subset \mathbb{C}$, dann ist

λ_i auch eine Nullstelle von $f \in \mathbb{R}[t]$. Ist $\lambda_i \notin \mathbb{R}$, dann ist auch $\bar{\lambda}_i$ eine komplexe Nullstelle, was durch komplexe Konjugation folgt:

$$f(t) = \sum_{i=0}^{n} \alpha_i t^i \quad \text{mit } \alpha_i \in \mathbb{R} \text{ und } f(\lambda) = 0 \text{ für } \lambda \notin \mathbb{R}$$

$$\Rightarrow \quad \bar{f}(\bar{t}) = \sum_{i=0}^{n} \alpha_i \bar{t}^i = f(\bar{t}) \qquad \text{hat Nullstelle } \bar{t} = \lambda, \text{ also } t = \bar{\lambda} \text{ Nullstelle von } f\ .$$

Außerdem ist die Vielfachheit der komplexen Nullstellen $\lambda, \bar{\lambda}$ gleich, d.h.

$$\mu(f; \lambda) = \mu(f, \bar{\lambda}) \qquad \forall\ f \in \mathbb{R}[t]\ .$$

Angenommen, wir hätten $f(t) = (t-\lambda)^r (t-\bar{\lambda})^{r'} g(t)$ mit $r \neq r'$ und $\lambda, \bar{\lambda}$ sind keine Nullstellen von $g \in \mathbb{C}[t]$. Dann sind $\lambda, \bar{\lambda}$ auch keine Nullstellen von $\bar{g} \in \mathbb{C}[t]$, und es gilt

$$\overline{f(t)} = f(\bar{t}) = (\bar{t}-\bar{\lambda})^r (\bar{t}-\lambda)^{r'} \overline{g(t)}\ .$$

Durch Austausch der formalen Variable $t \mapsto \bar{t}$ folgt, daß nun $\mu(f; \bar{\lambda}) = r$ und $\mu(f; \lambda) = r'$, also $r = r'$. Aus

$$(t-\lambda)(t-\bar{\lambda}) = t^2 - 2\mathrm{Re}(\lambda) t + |\lambda|^2 \in \mathbb{R}[t]$$

folgt nun, daß für jedes reelle Polynom $f \in \mathbb{R}[t]$ gilt:

$$f = a(t-\lambda_1)\cdots(t-\lambda_{n-2r}) \cdot g_1 \cdots g_r\ , \qquad g_j = (t-\alpha_j)^2 + \beta_j^2 > 0\ ,$$
$$a, \lambda_i, \alpha_j, \beta_j \in \mathbb{R}\ , \quad \beta_j \neq 0\ , \qquad r, n-2r \geq 0\ .$$

Die analytische Berechnung (zunächst komplexer Nullstellen) ist im allgemeinen nur für Polynome vom Grad ≤ 4 möglich. Für Polynome mit höherem Grad ist man auf numerische Näherungsverfahren angewiesen. Der Fundamentalsatz der Algebra und für reelle Polynome zusätzlich die Gleichheit der Vielfachheit komplexer Nullstellen ist dann eine wichtige Kontrolle, ob man wirklich alle Nullstellen numerisch gefunden hat!

18 Diagonalisierbarkeit

Wir kommen nun zurück auf Eigenwerte und Diagonalisierbarkeit.

Satz 18.1 *Sei $F \in \mathrm{End}(V)$ und $A = M_{\mathcal{A}}^{\mathcal{A}}(F)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{A} = (w_1, \dots, w_n)$ von V. Dann gilt*

$$\mathrm{Eig}(F; \lambda) = \Phi_{\mathcal{A}}(\mathrm{L\ddot{o}s}(A - \lambda \cdot E_n, 0))\ .$$

Dabei ist

$$\Phi_{\mathcal{A}} : K^n \to V\ , \qquad \Phi_{\mathcal{A}}(e_i) = w_i\ ,$$

der Isomorphismus, der die Standardbasis des K^n in die Basis $\mathcal{A}$ überführt, und

$$\mathrm{L\ddot{o}s}(A - \lambda \cdot E_n, 0) = \{x \in K^n\ :\ (A - \lambda \cdot E_n)x = 0\}\ .$$

Beweis. Sei $v \in \mathrm{Eig}(F;\lambda)$, dann ist

$$F(v) = \lambda v \quad \Leftrightarrow \quad \Phi_{\mathcal{A}}^{-1} \circ F \circ \Phi_{\mathcal{A}} \circ \Phi_{\mathcal{A}}^{-1}(v) = \lambda \Phi_{\mathcal{A}}^{-1}(v) \ .$$

Wir setzen $x = \Phi_{\mathcal{A}}^{-1}(v) \in K^n$. Mit $A = M_{\mathcal{A}}^{\mathcal{A}}(F) := \Phi_{\mathcal{A}}^{-1} \circ F \circ \Phi_{\mathcal{A}} \in M(n,K)$ folgt

$$v \in \mathrm{Eig}(F;\lambda) \quad \Leftrightarrow \quad A \cdot x = \lambda \cdot x \quad \Leftrightarrow \quad (A - \lambda \cdot E_n)x = 0 \qquad \square$$

Damit haben wir ein Verfahren zur Bestimmung von Eigenwerten und zugehörigen Eigenräumen von $F \in \mathrm{End}(V)$ gefunden:

1. Wähle eine beliebige Basis $\mathcal{A} = (w_1, \dots, w_n)$ von V. Bestimme die darstellende Matrix $A = (a_{ij})$ durch Zerlegen der n Vektoren $F(v_j)$ nach der Basis $\mathcal{A}$,
$$F(w_j) = \sum_{i=1}^{n} a_{ij} \cdot w_i \ .$$
2. Berechne das charakteristische Polynom $P_A(t) = \det(A - t \cdot E_n) \in K[t]$ und bestimme seine Nullstellen $\lambda_1, \dots, \lambda_m \in K$ (paarweise verschieden) und ihre Vielfachheit $r_i := \mu(P_A; \lambda_i)$ aus der Darstellung
$$P_A(t) = (t - \lambda_1)^{r_1} \cdots (t - \lambda_m)^{r_m} \cdot g(t) \ ,$$
wobei $g \in K[t]$ keine Nullstelle in K besitzt. Die $\lambda_1, \dots, \lambda_m$ sind dann genau die Eigenwerte von F. Für $K = \mathbb{C}$ gilt $n = r_1 + \cdots + r_m$.
3. Löse zu jeder Nullstelle λ_i des charakteristischen Polynoms das lineare Gleichungssystem $(A - \lambda_i \cdot E_n)x^{(i)} = 0$. Der Lösungsraum ist ein s_i-dimensionaler Untervektorraum von K^n. Ist $x^{(i)} = \sum_{j=1}^{n} x_j^{(i)} e_j \in \mathrm{Lös}(A - \lambda_i \cdot E_n, 0)$, dann ist $v^{(i)} = \Phi_{\mathcal{A}}(x^{(i)}) = \sum_{j=1}^{n} x_j^{(i)} v_j \in \mathrm{Eig}(F;\lambda_i)$ mit $\dim(\mathrm{Eig}(F;\lambda_i)) = s_i$.

Beispiel 18.2 Es sei $F : \mathbb{R}^3 \to \mathbb{R}^3$ in der Standardbasis gegeben durch $F(x) = A \cdot x$ mit $A = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 2 & 1 \\ 2 & -2 & 1 \end{pmatrix}$. Schritt 1 entfällt.

Schritt 2. Es gilt

$$\begin{aligned}
&\det(A - tE_n) = \det \begin{pmatrix} 2-t & 2 & 3 \\ 1 & 2-t & 1 \\ 2 & -2 & 1-t \end{pmatrix} \\
&= (2-t) \cdot \det \begin{pmatrix} 2-t & 1 \\ -2 & 1-t \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 1 & 1 \\ 2 & 1-t \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 1 & 2-t \\ 2 & -2 \end{pmatrix} \\
&= (2-t) \cdot (t^2 - 3t + 4) + 2(t+1) + 3(2t - 6) = -t^3 + 5t^2 - 2t - 8 \\
&= -(t+1)(t^2 - 6t + 8) = -(t+1)(t-4)(t-2) \ .
\end{aligned}$$

Damit hat F die Eigenwerte $\lambda_1 = -1,\ \lambda_2 = 2,\ \lambda_3 = 4$. Insbesondere ist F diagonalisierbar.

Schritt 3. Die Eigenräume bestimmen sich (dank der Standardbasis) zu $\mathrm{Eig}(F;\lambda_i) = \ker(A - \lambda_i \cdot E_3)$, sie werden also als Lösung eines homogenen LGS erhalten. Wir bestimmen $\mathrm{Eig}(F;-1) = \{v_1 \in \mathbb{R}^3 \ :\ (A-(-1)E_3)v_1 = 0\}$ durch Zeilenumformungen zu

$$\begin{pmatrix} 3 & 2 & 3 \\ 1 & 3 & 1 \\ 2 & -2 & 2 \end{pmatrix} \xmapsto{IV_{12}(-\frac{1}{3}),\, IV_{13}(-\frac{2}{3})} \begin{pmatrix} 3 & 2 & 3 \\ 0 & \frac{7}{3} & 0 \\ 0 & -\frac{10}{3} & 0 \end{pmatrix} \xmapsto{IV_{23}(\frac{10}{7}),\, III_2(\frac{3}{7})} \begin{pmatrix} 3 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\xmapsto{IV_{21}(-2),\, III_1(\frac{1}{3})} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \Rightarrow \quad v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} t_1\ , \qquad t_1 \in \mathbb{R}\ .$$

Analog findet man $v_2 = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix} t_2$ und $v_3 = \begin{pmatrix} 8 \\ 5 \\ 2 \end{pmatrix} t_3$. Sämtliche Eigenräume sind eindimensional.

Für $t_i \neq 0$ bildet $\mathcal{B} = (v_1, v_2, v_3)$ eine Basis von $\mathbb{R}^3$. Aufgefaßt als Matrix vermittelt $S = (v_1, v_2, v_3) \in GL(3,\mathbb{R})$ die Transformation von der Standardbasis $\mathcal{A} = \mathcal{E}$ in die Eigenbasis $\mathcal{B}$ von F, d.h. es gilt $\Phi_{\mathcal{B}}(x) = S \cdot x$ in den Bezeichnungen von Satz 10.2. Nach Satz 13.1 (bzw. dem kommuativen Diagramm davor) gilt $M_{\mathcal{B}}^{\mathcal{B}}(F) = \Phi_{\mathcal{B}}^{-1} \circ F \circ \Phi_{\mathcal{B}} = S^{-1} \cdot A \cdot S$. Für $t_1 = t_2 = t_3 = 1$ ist

$$S = \begin{pmatrix} 1 & 2 & 8 \\ 0 & 3 & 5 \\ -1 & -2 & 2 \end{pmatrix} \qquad \Rightarrow \qquad S^{-1} = \frac{1}{30} \begin{pmatrix} 16 & -20 & -14 \\ -5 & 10 & -5 \\ 3 & 0 & 3 \end{pmatrix}$$

und

$$S^{-1} \cdot A \cdot S = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

$\triangleleft$

Sei $F \in \mathrm{End}(V)$ mit $\dim(V) = n$ und $A = M_{\mathcal{A}}^{\mathcal{A}}(F) \in M(n,K)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{A}$ von V. Wir wissen:

i) Ist F diagonalisierbar, dann ist $P_A(t) = (-1)^n(t-\lambda_1)^{r_1} \cdots (t-\lambda_m)^{r_m}$, d.h. das charakteristische Polynom zerfällt in Linearfaktoren (wähle eine Basis aus Eigenvektoren).

ii) Ist $P_A(t) = (-1)^n(t-\lambda_1)\cdots(t-\lambda_n)$ und alle λ_i sind paarweise verschieden, dann ist F diagonalisierbar.

Es verbleibt also zu untersuchen, wann F diagonalisierbar ist im Falle von Vielfachheiten der Eigenwerte.

Satz 18.3 *Sei $F \in \mathrm{End}(V)$ und A seine darstellende Matrix bezüglich einer beliebigen Basis von V. Ist λ ein Eigenwert von F, dann gilt*

$$1 \leq \dim(\mathrm{Eig}(F;\lambda)) \leq \mu(P_A;\lambda) \ .$$

Beweis. Sei $(v_1, \ldots, v_s)$ eine Basis von $\mathrm{Eig}(F;\lambda)$. Dann ist $s \geq 1$, da λ Eigenwert. Wir ergänzen $(v_1, \ldots, v_s)$ zu einer Basis $\mathcal{A} = (v_1, \ldots, v_s, v_{s+1}, \ldots, v_n)$ von V. Für die darstellende Matrix $A = M_{\mathcal{A}}^{\mathcal{A}}(F) = (a_{ij})$ gilt

$$A = \left(\begin{array}{ccc|c} \lambda & & 0 & \\ & \ddots & & B \\ 0 & & \lambda & \\ \hline & 0 & & C \end{array}\right) ,$$

wobei oben links der Block λE_s steht. Dann gilt für das charakteristische Polynom

$$P_A(t) := \det(A - tE_n) = (-1)^s (t-\lambda)^s \det(C - tE_{n-s}) \ .$$

Folglich ist $s \leq \mu(P_A;\lambda)$. □

Der Fall $\dim(\mathrm{Eig}(F;\lambda)) = \mu(P_A;\lambda)$ ist von besonderem Interesse:

Satz 18.4 *Sei V ein n-dimensionaler Vektorraum, $F \in \mathrm{End}(V)$ und $A = M_{\mathcal{A}}^{\mathcal{A}}(F) \in M(n,K)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{A}$ von V. Dann sind die folgenden Eigenschaften äquivalent:*

i) *F ist diagonalisierbar.*

ii) *Das charakteristische Polynom zerfällt in Linearfaktoren und es gilt $\dim(\mathrm{Eig}(F;\lambda)) = \mu(P_A;\lambda)$ für jeden Eigenwert λ von F.*

iii) *Sind $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Eigenwerte von F, dann gilt $V = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F;\lambda_k)$.*

Bemerkung: ii) liefert damit das entscheidende Kriterium für Diagonalisierbarkeit: Das charakteristische Polynom muß in Linearfaktoren zerfallen und die Dimensionen der Eigenräume müssen gleich den Vielfachheiten der Nullstellen sein.

Beweis. i)⇒ii) Ist F diagonalisierbar, so ordnen wir die zugehörige Basis aus Eigenvektoren wie folgt:

$$\mathcal{B} = (v_1^{(1)}, \ldots, v_{s_1}^{(1)}, v_1^{(2)}, \ldots, v_{s_2}^{(2)}, \ldots, v_1^{(k)}, \ldots, v_{s_k}^{(k)}) \ .$$

Dabei ist $(v_1^{(i)}, \ldots, v_{s_i}^{(i)})$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$, und insbesondere gilt $F(v_j^{(i)}) = \lambda_i v_j^{(i)}$ für $1 \leq j \leq s_i$. In dieser Basis gilt für das charakteristische Polynom

$$P_F(t) = (\lambda_1 - t)^{s_1} (\lambda_2 - t)^{s_2} \cdots (\lambda_k - t)^{s_k} \ ,$$

welches somit die Eigenschaften ii) besitzt.

ii)$\Rightarrow$iii) Durch $W = \mathrm{Eig}(F;\lambda_i) + \cdots + \mathrm{Eig}(F;\lambda_k)$ werde ein Untervektorraum von V definiert. Da Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind, gilt $W = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F;\lambda_k)$. Dann folgt $\dim(W) = s_1 + \cdots + s_k = n$ und somit $W = V$.

iii)$\Rightarrow$i) Sei $\mathcal{B}_i = (v_1^{(i)}, \ldots, v_{s_i}^{(i)})$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$. Dann ist

$$\mathcal{B} = (v_1^{(1)}, \ldots, v_{s_1}^{(1)}, v_1^{(2)}, \ldots, v_{s_2}^{(2)}, \ldots, v_1^{(k)}, \ldots, v_{s_k}^{(k)})$$

eine Basis von V. Wegen $F(v_j^{(i)}) = \lambda_i v_j^{(i)}$ für $1 \le j \le s_i$ ist $\mathcal{B}$ eine Basis aus Eigenvektoren von F, d.h. F ist diagonalisierbar. $\square$

Wir sehen uns ein Beispiel zur Diagonalisierung an:

Beispiel 18.5 Es sei $F \in \mathrm{End}(\mathbb{R}^3)$ gegeben durch

$$F(x,y,z) := (y - z,\ 3x + 2y - 3z,\ 2x + 2y - 3z)\ .$$

1. Schritt: Bestimmung der darstellenden Matrix bezüglich einer Basis. Die Vektoren $v = (x,y,z)$ und $w = (y - z,\ 3x + 2y - 3z,\ 2x + 2y - 3z)$ sind bereits in der Standardbasis $\mathcal{A} = (e_1, e_2, e_3)$ des $\mathbb{R}^3$ dargestellt. Daraus lesen wir

$$F(e_1) = 3e_2 + 2e_3\ , \qquad F(e_2) = e_1 + 2e_2 + 2e_3\ , \qquad F(e_3) = -e_1 - 3e_2 - 3e_3$$

ab. Dann ist die darstellende Matrix bezüglich der Standardbasis $A = (a_{ij}) = M_{\mathcal{A}}^{\mathcal{A}}(F)$ gegeben durch $F(e_j) = \sum_{i=1}^{3} a_{ij} \cdot e_i$. Wir lesen ab:

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

(Die Bilder der Basisvektoren ergeben die Spalten von A.)

2. Schritt: Berechnung des charakteristischen Polynoms. Zu berechnen ist $\det(A - t \cdot E_3)$, z.B. durch elementare Zeilenumformungen in eine obere Dreiecksmatrix:

$$\det(A - t\cdot E_3) = \det \begin{pmatrix} -t & 1 & -1 \\ 3 & 2-t & -3 \\ 2 & 2 & -3-t \end{pmatrix} \overset{P_{13}}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 3 & 2-t & -3 \\ -t & 1 & -1 \end{pmatrix}$$

$$\overset{Q_{12}(-\frac{3}{2})\cdot Q_{13}(\frac{t}{2})}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 0 & -1-t & \frac{3}{2} + \frac{3}{2}t \\ 0 & 1+t & -1 - \frac{3}{2}t - \frac{1}{2}t^2 \end{pmatrix}$$

$$\overset{Q_{23}(1)}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 0 & -1-t & \frac{3}{2} + \frac{3}{2}t \\ 0 & 0 & \frac{1}{2} - \frac{1}{2}t^2 \end{pmatrix} = -2(-1-t)(\tfrac{1}{2} - \tfrac{1}{2}t^2) = (1+t)(1-t^2)$$

$$= -(t-1)(t-(-1))^2 \, .$$

Damit zerfällt das charakteristische Polynom in Linearfaktoren. Seine Nullstellen sind $\lambda_1 = 1$ mit Vielfachheit $\mu(P_A, 1) = 1$ und $\lambda_2 = -1$ mit Vielfachheit $\mu(P_A, -1) = 2$.

3. Schritt: Bestimmen der Eigenräume. Zum Eigenwert $\lambda_1 = 1$ lösen wir das lineare Gleichungssystem $(A - \lambda_1 \cdot E_3)x = 0$ durch elementare Zeilenumformungen:

$$\begin{pmatrix} -1 & 1 & -1 \\ 3 & 1 & -3 \\ 2 & 2 & -4 \end{pmatrix} \xrightarrow{Q_{12}(3)\,Q_{13}(2)} \begin{pmatrix} -1 & 1 & -1 \\ 0 & 4 & -6 \\ 0 & 4 & -6 \end{pmatrix} \xrightarrow{Q_{23}(-1)} \begin{pmatrix} -1 & 1 & -1 \\ 0 & 4 & -6 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{Q_{12}(-\frac{1}{4})} \left(\begin{array}{cc|c} -1 & 0 & \frac{1}{2} \\ 0 & 4 & -6 \\ \hline 0 & 0 & 0 \end{array}\right) \xrightarrow{S_1(-1)\,S_2(\frac{1}{4})} \left(\begin{array}{cc|c} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{3}{2} \\ \hline 0 & 0 & 0 \end{array}\right) .$$

Damit gilt $\dim(\mathrm{Eig}(F; 1)) = 1$, genauer

$$\mathrm{Eig}(F; 1) = \mathbb{R}v_1^{(1)} \, , \quad v_1^{(1)} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} .$$

Zur Bestimmung des Eigenraums $\mathrm{Eig}(F; -1)$ ist das LGS $(A + E_3)x = 0$ zu lösen:

$$\begin{pmatrix} 1 & 1 & -1 \\ 3 & 3 & -3 \\ 2 & 2 & -2 \end{pmatrix} \xrightarrow{Q_{12}(-3)\,Q_{13}(-2)} \left(\begin{array}{c|cc} 1 & 1 & -1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right) .$$

Damit gilt $\dim(\mathrm{Eig}(F; 1)) = 2$, und die Lösungsvektoren sind

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \left(\begin{array}{cc} -1 & 1 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -w_1 + w_2 \\ w_1 \\ w_2 \end{pmatrix} .$$

Damit finden wir die beiden Basisvektoren

$$v_1^{(2)} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} , \qquad v_2^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} .$$

Also sind die Dimensionen der Eigenräume gleich der Vielfachheit der Nullstellen, und F ist diagonalisierbar. Eine Basis von V, welche F diagonalisiert, ist also

$$\mathcal{B} = \left(\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} , \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} , \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right) , \quad \Rightarrow \quad S = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} .$$

In dieser Basis gilt also

$$\Lambda := M_{\mathcal{B}}^{\mathcal{B}}(F) = S^{-1} \cdot A \cdot S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} .$$

$\triangleleft$

Über die Diagonalisierung lassen sich Polynome und sogar geeignete Potenzreihen von Matrizen bequem ausrechnen. Es sei $A = S \cdot \Lambda \cdot S^{-1} \in M(n, K)$ diagonalisierbar, wobei $\Lambda = (\lambda_{ij})$ mit $\lambda_{ij} = \delta_{ij}\lambda_i \in M(n, K)$ die aus den Eigenwerten gebildete Diagonalmatrix ist und $S \in GL(n, K)$ die Transformation in die Eigenbasis vermittelt, d.h. die i-te Spalte von S ist Eigenvektor zu λ_i. Dann gilt $A^k = S \cdot \Lambda^k \cdot S^{-1}$ mit $\Lambda^k = (\lambda'_{ij})$ und $\lambda_{ij'} = \lambda_i^k \delta_{ij}$. Insbesondere ist $A^0 = E_n$ zu setzen. In der Lösungstheorie für Differentialgleichungen besonders wichtig ist $\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$. Diese Matrix-Exponentialreihe konvergiert für beliebige $A \in M(n, K)$. Für diagonalisierbare Matrizen läßt sie sich ausrechnen zu $\exp(A) = S \cdot \exp(\Lambda) \cdot S^{-1}$.

Beispiel 18.6 Es sei $A = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 2 & 1 \\ 2 & -2 & 1 \end{pmatrix}$ aus Beispiel 18.2. Dann ist

$$\begin{aligned} \exp(A) &= \begin{pmatrix} 1 & 2 & 8 \\ 0 & 3 & 5 \\ -1 & -2 & 2 \end{pmatrix} \begin{pmatrix} e^{-1} & 0 & 0 \\ 0 & e^2 & 0 \\ 0 & 0 & e^4 \end{pmatrix} \frac{1}{30} \begin{pmatrix} 16 & -20 & -14 \\ -5 & 10 & -5 \\ 3 & 0 & 3 \end{pmatrix} \\ &= \frac{1}{30} \begin{pmatrix} 16e^{-1} - 10e^2 + 24e^4 & -20e^{-1} + 20e^2 & -14e^{-1} - 10e^2 + 24e^4 \\ -15e^2 + 15e^4 & 30e^2 & -15e^2 + 15e^4 \\ -16e^{-1} + 10e^2 + 6e^4 & 20e^{-1} - 20e^2 & 14e^{-1} + 10e^2 + 6e^4 \end{pmatrix} . \end{aligned}$$

$\triangleleft$

In Beispiel 18.5 ist $\Lambda^2 = E_3$ und damit auch $A^2 = (S\Lambda S^{-1}) \cdot (S\Lambda S^{-1}) = E_3$ oder

$$(A - \lambda_1 E_3)(A - \lambda_2 E_3) = 0 .$$

Ganz allgemein gilt, wenn man im charakteristischen Polynom $P_A(t)$ die formale Variable t durch A ersetzt, der

Satz 18.7 (Cayley-Hamilton) *Es sei $F \in \mathrm{End}(V)$ Endomorphismus eines endlich-dimensionalen Vektorraums und $P_F \in K[t]$ das charakteristische Polynom. Dann gilt $P_F(F) = 0 \in \mathrm{End}(V)$. Ausgedrückt durch Matrizen: Für beliebige $A \in M(n, K)$ mit charakteristischem Polynom $P_A \in K[t]$ gilt $P_A(A) = 0$.*

Der Beweis findet sich z.B. in G. Fischer, Lineare Algebra, §4.5. $\square$

In Beispiel 18.5 hatten wir $P_A(t) = (t - 1)(t + 1)^2$ und damit $(A - E_3)(A + E_3)^2 = 0$. Tatsächlich gilt bereits $(A - E_3)(A + E_3) = 0$ für ein Polynom kleineren

Grades. Allgemein läßt sich zu $F \in \mathrm{End}(V)$ ein eindeutiges *Minimalpolynom* $M_F[t]$ finden mit $M_F(F) = 0$ bzw. $M_A(A) = 0$. Das Minimalpolynom ist stets Teiler des charakteristischen Polynoms und ein wichtiges Hilfsmittel in Beweisen.

Zum Abschluß der Betrachtungen zur Diagonalisierbarkeit untersuchen wir folgendes Problem: Gegeben seien zwei diagonalisierbare Endomorphismen $F, G \in \mathrm{End}(V)$. Unter welchen Bedingungen sind F, G *simultan diagonalisierbar*, d.h. es gibt eine Basis von V aus Eigenvektoren von F und G gleichzeitig?

Satz 18.8 *Zwei diagonalisierbare Endomorphismen $F, G \in \mathrm{End}(V)$ sind genau dann simultan diagonalisierbar, wenn sie miteinander kommutieren, d.h. wenn $F \circ G = G \circ F$.*

Beweis. ($\Rightarrow$) Sind F, G simultan diagonalisierbar, dann existiert eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V mit $F(v_i) = \lambda_i v_i$ und $G(v_i) = \mu_i v_i$. Dann gilt für einen beliebigen Vektor $v = \sum_{i=1}^n \kappa_i v_i \in V$

$$(F \circ G)(v) = F(G(\sum_{i=1}^n \kappa_i v_i)) = \sum_{i=1}^n \lambda_i \kappa_i \mu_i v_i = (G \circ F)(v) \;.$$

($\Leftarrow$) 1) Wir zerlegen den Vektorraum V in die Eigenräume:

$$\begin{aligned} V &= \mathrm{Eig}(F; \lambda_1) \oplus \dots \oplus \mathrm{Eig}(F; \lambda_k) \\ &= \mathrm{Eig}(G; \mu_1) \oplus \dots \oplus \mathrm{Eig}(G; \mu_l) \;. \end{aligned}$$

Kommutieren F und G, dann gilt $F(\mathrm{Eig}(G; \mu_j)) \subset \mathrm{Eig}(G; \mu_j)$ für alle $1 \leq j \leq l$, denn für $w_j \in \mathrm{Eig}(G; \mu_j)$ folgt

$$G(F(w_j)) = F(G(w_j)) = F(\mu_j w_j) = \mu_j F(w_j) \;.$$

2) Sei nun $W_{ij} := \mathrm{Eig}(F; \lambda_i) \cap \mathrm{Eig}(G; \mu_j)$. Dann ist $W_{ij} \subset V$ ein Untervektorraum, und es gilt $\mathrm{Eig}(F; \lambda_i) = W_{i1} \oplus \dots \oplus W_{il}$. Denn sei $v_i \in \mathrm{Eig}(F; \lambda_i)$, dann gibt es $w_1 \in \mathrm{Eig}(G; \mu_1), \dots, w_l \in \mathrm{Eig}(G; \mu_l)$ mit $v_i = w_1 + \dots + w_l$ (verwende Basis von V aus Eigenvektoren von G). Anwenden von F liefert

$$F(v_i) = F(w_1) + \dots + F(w_l) = \lambda_i v_i = \lambda_i w_1 + \dots + \lambda_i w_l \;.$$

Nun ist $F(w_j) \subset \mathrm{Eig}(G; \mu_j)$, und da Vektoren w_j, w_j' linear unabhängig sind für $j \neq j'$, folgt $F(w_j) = \lambda_i w_j$ für alle $1 \leq j \leq l$. Damit ist $w_j \subset \mathrm{Eig}(F; \lambda_i)$, also $\mathrm{Eig}(F; \lambda_i) = W_{i1} + \dots + W_{il}$, und aus der linearen Unabhängigkeit folgt die Behauptung.

3) Sei $\mathcal{B}_{ij} = (v_1^{(ij)}, \dots v_{s_{ij}}^{(ij)})$ eine Basis von W_{ij}, dann ist

$$(\mathcal{B}_{11}, \dots, \mathcal{B}_{1l}, \mathcal{B}_{21}, \dots, \mathcal{B}_{2l}, \dots, \mathcal{B}_{k1}, \dots, \mathcal{B}_{kl})$$

eine Basis von V, in der F und G simultan diagonalisierbar sind, mit $F(v_r^{(ij)}) = \lambda_i v_r^{(ij)}$ und mit $G(v_r^{(ij)}) = \mu_j v_r^{(ij)}$. □

Simultane Diagonalisierbarkeit (in verallgemeinerter Form) ist wichtig in der Quantenmechanik, wo man in einem System zwei physikalische Größen nur dann gleichzeitig messen kann, wenn die entsprechenden Endomorphismen (des Hilbert-Raums) miteinander kommutieren. Im Wasserstoffatom sind das die Energie, der Gesamtdrehimpuls, eine Komponente des Drehimpulses (üblicherweise die z-Komponente) und der Spin. Entsprechend schreibt sich der Hilbert-Raum als direkte Summe von Eigenunterräumen der linearen Abbildungen, welche diesen physikalischen Größen entsprechen.

19 Trigonalisierung

Diagonalisierbarkeit eines Endomorphismus erfordert zwei Bedingungen:

(1) Das charakteristische Polynom zerfällt in Linearfaktoren.

(2) Die Vielfachheit der Nullstellen ist gleich der Dimension der Eigenräume.

Wir werden nun sehen, daß für Endomorphismen, die nur (1) erfüllen, eine Basis exisitiert, so daß die darstellende Matrix eine obere Dreiecksmatrix ist. Das genügt zur Lösung von Gleichungssystemen.

Definition 19.1 Sei $F \in \mathrm{End}(V)$. Ein Untervektorraum $W \subset V$ heißt F-*invariant*, wenn $F(W) \subset W$.

Offenbar sind die Eigenräume $\mathrm{Eig}(F;\lambda)$ automatisch F-invariant. Für die Triagonalisierung sind invariante Unterräume interessant, die keine Eigenräume sind.

Satz 19.2 *Sei $W \subset V$ ein F-invarianter Unterraum und $F|_W : W \to W$ die Einschränkung von F auf W. Dann ist das charakteristische Polynom $P_{F|_W}(t)$ ein Teiler von $P_F(t)$.*

Beweis. Sei $\dim(V) = n$ und $\dim(W) = r \leq n$. Wir ergänzen eine Basis $\mathcal{B}|_W$ von W zu einer Basis $\mathcal{B} = (\mathcal{B}|_W, \mathcal{B}')$ von V. Sei $A|_W := M_{\mathcal{B}|_W}^{\mathcal{B}|_W}(F|_W)$, dann gilt

$$A = M_{\mathcal{B}}^{\mathcal{B}}(F) = \begin{pmatrix} A|_W & * \\ 0 & A' \end{pmatrix} .$$

Damit ist $P_A(t) = \det(A - t \cdot E_n) = P_{A|_W}(t) \cdot \det(A' - t \cdot E_{n-r})$. □

Sei $F \in \mathrm{End}(K^n)$ in der Standardbasis durch eine obere Dreiecksmatrix $A \in M(n,K)$, d.h. $a_{ij} = 0$ für $i > j$, dargestellt. Definieren wir $W_i := \mathrm{span}(e_1, \ldots, e_r)$, dann gilt $F(W_r) \subset W_r$, d.h. alle W_r mit $1 \leq r \leq n$ sind F-invariant. Abstrakter formuliert:

Definition 19.3 Eine *Fahne* (V_r) in einem n-dimensionalen Vektorraum V ist eine Kette

$$\{0\} = V_0 \subset V_1 \subset V_2 \cdots \subset V_n = V$$

von Untervektorräumen mit $\dim(V_r) = r$. Ist $F \in \mathrm{End}(V)$, dann heißt die Fahne F-invariant, wenn $F(V_r) \subset V_r$ für alle $0 \leq r \leq n$.

Jede Basis von V definiert eine Fahne. Entscheidend ist, daß in einer F-invarianten Fahne gilt $F(V_1) \subset V_1$ mit $\dim(V_1) = 1$, so daß es einen Eigenvektor von F geben muß. Aus der Definition folgt direkt:

Satz 19.4 *Für $F \in \mathrm{End}(V)$ sind folgende Bedingungen äquivalent:*

i) *Es gibt eine F-invariante Fahne von V.*

ii) *Es gibt eine Basis $\mathcal{B}$ von V, so daß $M_{\mathcal{B}}^{\mathcal{B}}(F)$ eine obere Dreiecksmatrix ist.*

Ist das der Fall, dann heißt F trigonalisierbar.

Nun der entscheidende Satz:

Satz 19.5 *Für $F \in \mathrm{End}(V)$ mit $\dim(V) = n$ sind folgende Bedingungen äquivalent:*

i) *F ist trigonalisierbar.*

ii) *Das charakteristische Polynom von F zerfällt in Linearfaktoren, d.h.*

$$P_F(t) = (-1)^n (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n) , \qquad \lambda_1, \dots, \lambda_n \in K .$$

Insbesondere gilt: Jeder Endomorphismus eines endlich-dimensionalen komplexen Vektorraums ist trigonalisierbar.

Beweis. i)⇒ii) ist klar, denn ist die darstellende Matrix A von F bezüglich der Basis $\mathcal{B}$ aus Satz 19.4 eine obere Dreiecksmatrix, so gilt $T_F(t) = P_A(t) = \det(A - t \cdot E_n) = (a_{11} - t) \cdots (a_{nn} - t)$, d.h. $\lambda_i = a_{ii}$.

ii)⇒i) durch Induktion nach $n = \dim(V)$. Der Fall $n = 1$ ist klar. Sei also $n \geq 2$, dann wählen wir einen Eigenvektor v_1 zum Eigenwert (Nullstelle) λ_1 und ergänzen v_1 zu einer Basis $\mathcal{B} = (v_1, w_1, \dots, w_{n-1})$ von V. Dann ist

$$V = V_1 \oplus W \qquad \text{mit} \quad V_1 := \mathrm{span}(v_1) , \; W = \mathrm{span}(w_1, \dots, w_{n-1}) .$$

Nun ist V_1 ein F-invarianter Untervektorraum, W im allgemeinen aber nicht:

$$A = M_{\mathcal{B}}^{\mathcal{B}}(F) = \left(\begin{array}{c|ccc} \lambda_1 & a_1 & \dots & a_{n-1} \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right) .$$

Diese Darstellung definiert zwei lineare Abbildungen

$$
\begin{aligned}
G &: W \to W \,, & G(w_j) &= \sum_{i=1}^{n-1} B_{ij} \cdot w_i \,, \\
H &: W \to V_1 \,, & H(w_j) &= a_j \cdot v_1 \,.
\end{aligned}
$$

Nun gilt für das charakteristische Polynom $P_A(t) = -(t-\lambda_1)P_B(t)$. Da $P_A(t)$ nach Voraussetzung in Linearfaktoren zerfällt, zerfällt auch $P_B(t)$ in Linearfaktoren. Nach Induktionsvoraussetzung ist die durch B definierte lineare Abbildung $G \in \operatorname{End}(W)$ trigonalisierbar. Es gibt also eine G-invariante Fahne $\{0\} = W_0 \subset W_1 \subset \cdots \subset W_{n-1} = W$. Wir setzen $V_r := V_1 + W_{r-1}$ für $1 \leq r \leq n$. Ist $v = \mu v_1 + w \in V_r$, also $w \in W_{r-1}$, dann gilt

$$
F(v) = \mu F(v_1) + F(w) = \mu\lambda_1 v_1 + G(w) + H(w) \in V_1 + W_{r-1}
$$

wegen $H(w) \in V_1$ und $G(w) \in W_{r-1}$. Somit ist $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V$ eine F-invariante Fahne, und F ist trigonalisierbar. $\square$

Beispiel 19.6 Eine gedämpfte Schwingung wird durch die Differentialgleichung

$$
\ddot{x}\,(t) + 2\mu\dot{x}(t) + \omega^2 x(t) = 0 \,, \qquad x(0) = x_0 \,,\ \dot{x}(0) = v_0
$$

beschrieben. Wir setzen $y_1(t) = x(t)$, $y_2(t) = \dot{x}(t)$, dann ergibt sich ein System von zwei gekoppelten linearen Differentialgleichungen erster Ordnung

$$
\dot{y} = A \cdot y \qquad \text{mit} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \,, \quad A = \begin{pmatrix} 0 & 1 \\ -\omega^2 & -2\mu \end{pmatrix} \,, \quad y(0) = \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} \,.
$$

Da A unabhängig von t ist, ist die formale (und korrekte) Lösung gegeben durch

$$
y(t) = \exp(At) \cdot y(0) \quad \text{mit} \quad \exp(At) = E_2 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \underbrace{A \cdots A}_{n} \,,
$$

jedoch lassen sich diese Produkte so nicht leicht berechnen. Der Ausweg besteht in der Trigonalisierung von A.

Zunächst hat $P_A(t) = t^2 + 2\mu t + \omega^2$ die komplexen Nullstellen $\lambda_{1,2} = -\mu \pm \sqrt{\mu^2 - \omega^2}$. Für $\mu = \omega$ hat die Nullstelle $\lambda = -\mu$ die Vielfachheit 2, die uns näher interessieren wird. Für $\mu = \omega$ bestimmen wir den Eigenraum zu $\lambda = -\omega$:

$$
(A + \omega \cdot E_2) \cdot v = \begin{pmatrix} \omega & 1 \\ -\omega^2 & -\omega \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
$$

$$
\Rightarrow \begin{pmatrix} \omega & 1 \\ -\omega^2 & -\omega \end{pmatrix} \xrightarrow{IV_{12}(\omega)} \begin{pmatrix} \omega & 1 \\ 0 & 0 \end{pmatrix} \qquad \Rightarrow v = \begin{pmatrix} 1 \\ -\omega \end{pmatrix} \cdot t \,.
$$

Damit ist $\dim(\mathrm{Eig}(A;-\omega)) = 1$, aber die Nullstelle hat Vielfachheit 2. Folglich ist A nicht diagonalisierbar.

Zur Trigonalisierung wählen wir die Basis $\mathcal{B} = (v, e_2)$, dann ist

$$A = \underbrace{\begin{pmatrix} 1 & 0 \\ -\omega & 1 \end{pmatrix}}_{S} \begin{pmatrix} -\omega & 1 \\ 0 & -\omega \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 \\ \omega & 1 \end{pmatrix}}_{S^{-1}} .$$

Somit gilt

$$\begin{aligned} A^n &= S \begin{pmatrix} -\omega & 1 \\ 0 & -\omega \end{pmatrix}^n S^{-1} = S \begin{pmatrix} (-\omega)^n & n\omega^{n-1} \\ 0 & (-\omega)^n \end{pmatrix} S^{-1} \\ \Rightarrow \exp(At) &= S \begin{pmatrix} e^{-\omega t} & te^{\omega t} \\ 0 & e^{-\omega t} \end{pmatrix} S^{-1} \end{aligned}$$

und damit insbesondere

$$(S^{-1}y)(t) = \begin{pmatrix} e^{-\omega t} & te^{\omega t} \\ 0 & e^{-\omega t} \end{pmatrix} S^{-1}y(0)$$

Unter Verwendung von $S^{-1}y = \begin{pmatrix} x \\ \omega x + \dot{x} \end{pmatrix}$ erhalten wir für die 2. Komponente

$$\begin{aligned} & & \omega x + \dot{x} &= \mathrm{e}^{-\omega t}(\omega x_0 + v_0) \\ \Rightarrow & & \mathrm{e}^{-\omega t}\frac{d}{dt}(\mathrm{e}^{\omega t}x) &= \mathrm{e}^{-\omega t}(\omega x_0 + v_0) \\ \Rightarrow & & (\mathrm{e}^{\omega t}x) &= const + t(\omega x_0 + v_0)\ , \quad t = 0 \ \Rightarrow\ const = x_0 \\ \Rightarrow & & x(t) &= \mathrm{e}^{-\omega t}(x_0 + t\omega x_0 + tv_0)\ . \end{aligned}$$

Die Lösung beschreibt die Auslenkung $x(t)$ im *aperiodischen Grenzfall.* $\triangleleft$

Teil VI
Euklidische und unitäre Vektorräume

In reellen und komplexen Vektorräumen läßt sich als Zusatzstruktur ein *Skalarprodukt* einführen, welches dann insbesondere einen *Abstand* definiert. Damit lassen sich geometrische Größen wie Längen und Winkel berechnen, aber auch die Analysis vom $\mathbb{R}^1$ auf den $\mathbb{R}^n$ und sogar auf unendlich-dimensionale Vektorräume verallgemeinern. Letzteres ist Gegenstand der Funktionalanalysis. Im weiteren bezeichne $\mathbb{K}$ den Körper der reellen oder komplexen Zahlen.

20 Skalarprodukte

20.1 Definition und Beispiele. Cauchy-Schwarz

Definition 20.1 Es sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ und V ein Vektorraum über $\mathbb{K}$. Ein Skalarprodukt auf V ist eine Abbildung $\langle\ .\ \rangle : V \times V \to \mathbb{K}$ mit

(S1) $\langle\ .\ \rangle$ ist linear in der zweiten Variablen, d.h. $\langle w, \lambda_1 v_1 + \lambda_2 v_2\rangle = \lambda_1\langle w, v_1\rangle + \lambda_2\langle w, v_2\rangle$ für alle $v_1, v_2, w \in V$ und $\lambda_1, \lambda_2 \in \mathbb{K}$.

(S2) Es gilt $\langle w, v\rangle = \overline{\langle v, w\rangle}$ (komplexe Konjugation) für alle $v, w \in V$.

(S3) Es gilt $\langle v, v\rangle \geq 0$ für alle $v \in V$ und $\langle v, v\rangle = 0$ genau dann, wenn $v = 0$.

Ein reeller bzw. komplexer Vektorraum mit Skalarprodukt heißt *euklidischer bzw. unitärer Vektorraum* oder (in der Funktionalanalysis) *Prä-Hilbert-Raum*.

Achtung: Wir verwenden hier die Konvention der Physik. In der mathematischen Literatur wird das Skalarprodukt als *linear in der ersten Komponente definiert*. In unserer Konvention folgt aus (S1) und (S2) für die erste Komponente

$$\langle\lambda_1 w_1+\lambda_2 w_2, v\rangle = \overline{\langle v, \lambda_1 w_1+\lambda_2 w_2\rangle} = \overline{\lambda_1\langle v, w_1\rangle + \lambda_2\langle v, w_2\rangle} = \overline{\lambda_1}\langle w_1, v\rangle+\overline{\lambda_2}\langle w_2, v\rangle.$$

In reellen Vektorräumen ist das Skalarprodukt also auch in der ersten Variablen linear. Außerdem vereinfacht sich (S2) in reellen Vektorräumen zur Symmetrie $\langle v, w\rangle = \langle w, v\rangle$. Es sei bemerkt, daß aus (S2) folgt $\langle v, v\rangle \in \mathbb{R}$, so daß (S3) sinnvoll ist.

Beispiel 20.2 (Standardskalarprodukt in $\mathbb{R}^n$) Durch

$$\langle x, y\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n\ , \qquad \text{für} \quad \begin{array}{l} x = (x_1, \dots, x_n) \in \mathbb{R}^n\ , \\ y = (y_1, \dots, y_n) \in \mathbb{R}^n \end{array}$$

wird ein Skalarprodukt $\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ definiert. Dieses heißt das *kanonische Skalarprodukt* im $\mathbb{R}^n$. ◁

Beispiel 20.3 (Standardskalarprodukt in $\mathbb{C}^n$) Durch

$$\langle w,z\rangle := \overline{w_1}z_1 + \overline{w_2}z_2 + \cdots + \overline{w_n}z_n \;, \qquad \text{für} \quad \begin{array}{l} w=(w_1,\dots,w_n)\in\mathbb{C}^n\;, \\ z=(z_1,\dots,z_n)\in\mathbb{C}^n \end{array}$$

wird ein Skalarprodukt $\langle\ ,\ \rangle : \mathbb{C}^n\times\mathbb{C}^n\to\mathbb{C}$ definiert. Dieses heißt das *kanonische Skalarprodukt* im $\mathbb{C}^n$. ◁

Beispiel 20.4 Es sei $V=\mathbb{R}^2$. Dann definiert

$$\Big\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\Big\rangle = (2x_1+x_2)y_1 + (x_1+2x_2)y_2 = (2y_1+y_2)x_1 + (y_1+2y_2)x_2$$

ein Skalarprodukt. (S3) folgt aus $\langle x,x\rangle = 2x_1^2+2x_1x_2+2x_2^2 = (x_1+x_2)^2+x_1^2+x_2^2 \geq 0$ mit Gleichheit genau dann, wenn $x_1=x_2=0$. Dieses Skalarprodukt ist verschieden vom kanonischen. Tatsächlich gibt es unendlich viele verschiedene Skalarprodukte auf $\mathbb{K}^n$. ◁

Beispiel 20.5 Es sei

$$\ell^2(\mathbb{N}) := \{f:\mathbb{N}\to\mathbb{C}\ :\ \sum_{n=0}^{\infty}|f(n)|^2<\infty\}\;.$$

Dann definiert

$$\langle f,g\rangle := \sum_{n=0}^{\infty}\overline{f(n)}g(n)\;, \qquad f,g\in\ell^2(\mathbb{N})$$

ein Skalarprodukt auf $\ell^2(\mathbb{N})$. Dabei wird verwendet, daß wegen $|\overline{f(n)}g(n)| = |f(n)|\,|g(n)| \leq \frac{1}{2}(|f(n)|^2+|g(n)|^2)$ die unendliche Reihe existiert. ◁

Beispiel 20.6 Es sei $V=\mathcal{C}([a,b])$ der Vektorraum der stetigen komplexwertigen Funktionen über dem Intervall $[a,b]\subset\mathbb{R}$. Dann definiert

$$\langle f,g\rangle := \int_a^b dx\;\overline{f(x)}g(x)$$

ein Skalarprodukt auf V. Wegen der Stetigkeit von f folgt aus $\int_a^b dx\;|f(x)|^2=0$, daß $f(x)=0$ für alle $x\in[a,b]$. ◁

Wir kommen nun zur wichtigsten Ungleichung für Skalarprodukte.

Satz 20.7 (Cauchy-Schwarzsche Ungleichung) *Es sei V ein Vektorraum über $\mathbb{K}$ mit Skalarprodukt $\langle\ ,\ \rangle : V\times V\to\mathbb{K}$. Dann gilt*

$$\big|\langle w,v\rangle\big|^2 \leq \langle v,v\rangle\langle w,w\rangle \qquad \textit{für alle } v,w\in V$$

mit Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis. Sei $w \neq 0$ (für $w = 0$ ist die Ungleichung offensichtlich erfüllt), somit ist $\langle w, w\rangle \neq 0$. Dann gilt für beliebiges $\lambda \in \mathbb{C}$

$$\begin{aligned} 0 \leq \langle \lambda w - v, \lambda w - v\rangle &= \lambda\overline{\lambda}\langle w, w\rangle - \overline{\lambda}\langle w, v\rangle - \lambda\langle v, w\rangle + \langle v, v\rangle \\ &= \langle w, w\rangle \left|\lambda - \frac{\langle w, v\rangle}{\langle w, w\rangle}\right|^2 + \frac{1}{\langle w, w\rangle}\Big(\langle w, w\rangle\langle v, v\rangle - \big|\langle w, v\rangle\big|^2\Big) . \end{aligned}$$

Für $\lambda = \frac{\langle w,v\rangle}{\|w\|^2}$ wird direkt die Ungleichung von Cauchy-Schwarz erhalten.

Gilt das Gleichheitszeichen, so ist $\langle w, v\rangle = \mathrm{e}^{\mathrm{i}\alpha}\sqrt{\langle w, w\rangle}\sqrt{\langle v, v\rangle}$ für ein $\alpha \in [0, 2\pi]$. Setzen wir $\lambda = \mathrm{e}^{\mathrm{i}\alpha}\frac{\sqrt{\langle v,v\rangle}}{\sqrt{\langle w,w\rangle}}$, so folgt die Identität $\langle \lambda w - v, \lambda w - v\rangle = 0$, also $v = \lambda w$. □

Der Beweis benutzt, daß $\langle w, w\rangle$ eine nichtnegative reelle Zahl ist, aus der eine eindeutige Quadratwurzel $\sqrt{\langle w, w\rangle} \in \mathbb{R}_+$ gezogen werden kann. Damit wird der Prototyp einer Norm erhalten.

20.2 Normierte Vektorräume

Definition 20.8 Es sei V ein Vektorraum über $\mathbb{K}$. Eine *Norm* auf V ist eine Abbildung $\|\ \| : V \to \mathbb{R}_+$ mit

(N1) $\|v\| = 0 \quad \Leftrightarrow \quad v = 0$,

(N2) $\|\lambda v\| = |\lambda|\,\|v\|$ für alle $v \in V$, $\lambda \in \mathbb{C}$,

(N3) $\|v + w\| \leq \|v\| + \|w\|$ für alle $v, w \in V$ (Dreiecksungleichung)

Ein Vektorraum mit Norm heißt *normierter Vektorraum*.

Satz 20.9 *Jeder euklidische und unitäre Vektorraum ist auch ein normierter Vektorraum mit* $\|v\| := \sqrt{\langle v, v\rangle}$.

Beweis. (N1) und (N2) sind klar, (N3) folgt aus Cauchy-Schwarz:

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + 2\mathrm{Re}(\langle v, w\rangle) + \langle w, w\rangle \\ &\leq \langle v, v\rangle + 2|\langle v, w\rangle| + \langle w, w\rangle \\ &\leq \langle v, v\rangle + 2\sqrt{\langle v, v\rangle\langle w, w\rangle} + \langle w, w\rangle = (\|v\| + \|w\|)^2 . \end{aligned}$$

□

Damit liefern die Beispiele 20.2–20.6 auch Beispiele für normierte Vektorräume. Weitere Beispiele sind:

Beispiel 20.10 $V = \mathbb{K}^n$ ist mit $\|x\|_1 := \sum_{i=1}^n |x_i|$ für $x = (x_1, \dots, x_n) \in \mathbb{K}^n$ ein normierter Vektorraum $(\mathbb{K}^n, \|\ \|_1)$. (N1) und (N2) sind klar, (N3) folgt aus der Dreiecksungleichung in $\mathbb{R}$ bzw. $\mathbb{C}$ für jede Komponente.

Das kanonische Skalarprodukt auf $\mathbb{K}^n$ liefert die weitere Norm $\|x\|_2 := \sqrt{\sum_{i=1}^n |x_i|^2}$, die *Standardnorm.* Es läßt sich zeigen, daß $\|x\|_p := \big(\sum_{i=1}^n |x_i|^p\big)^{\frac{1}{p}}$ für jedes $1 \leq p < \infty$ eine Norm auf $\mathbb{K}^n$ ist.

Schließlich ist auch $\|x\|_\infty := \max_{i=1,\dots,n} |x_i|$ eine Norm auf $\mathbb{K}^n$. ◁

Beispiel 20.11 Es sei $V = \mathcal{C}([a,b])$ der Vektorraum der stetigen komplexwertigen Funktionen über $[a,b]$. Folgende Abbildungen $\|\ \|_p : V \to \mathbb{R}_+$ sind Normen (mit $f \in V$):

$$\begin{aligned}
\|f\|_1 &:= \int_a^b dx\, |f(x)| \\
\|f\|_2 &:= \sqrt{\int_a^b dx\, |f(x)|^2} = \sqrt{\langle f, f\rangle} \\
\|f\|_p &:= \Big(\int_a^b dx\, |f(x)|^p\Big)^{\frac{1}{p}}\ , \qquad 1 \leq p < \infty \\
\|f\|_\infty &:= \sup_{x\in[a,b]} |f(x)|\ .
\end{aligned}$$

Dabei heißt $\|f\|_\infty$ die Supremumsnorm einer Funktion f. Der Beweis, daß $\|\ \|_p$ eine Norm ist, erfordert die Höldersche Ungleichung. ◁

Wir sehen also, daß es auf einem Vektorraum verschiedene Normen geben kann, und in unendlich-diemnsionalen Vektorräumen werden die unterschiedlichen Normen wirklich gebraucht! Die aus einem Skalarprodukt folgende Norm ist ausgezeichnet durch

Satz 20.12 *Es sei $(V, \|\ \|)$ ein normierter Vektorraum über $\mathbb{K}$. Die Norm $\|\ \|$ geht genau dann aus einem Skalarprodukt hervor, wenn die* Parallelogrammgleichung *gilt,*

$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2 \qquad \textit{für alle } v,w \in V\ .$$

Beweisidee. Die Richtung ($\Rightarrow$) ist einfaches Nachrechnen. Für die Umkehrung ($\Leftarrow$) definiert man über die *Polarisationsformeln*

$$\begin{aligned}
\langle v,w\rangle &= \tfrac{1}{4}\big(\|v+w\|^2 - \|v-w\|^2 - \mathrm{i}\|v+\mathrm{i}w\|^2 + \mathrm{i}\|v-\mathrm{i}w\|^2\big) && \text{für } \mathbb{K} = \mathbb{C}\ , \\
\langle v,w\rangle &= \tfrac{1}{4}\big(\|v+w\|^2 - \|v-w\|^2\big) && \text{für } \mathbb{K} = \mathbb{R}
\end{aligned}$$

zunächst Abbildungen $\langle\ ,\ \rangle : V \times V \to \mathbb{K}$, für die man dann die Eigenschaften (S1)–(S3) zeigen muß, wobei die Parallelogrammgleichung eingeht. Für die Linearität ist das eine sehr mühsame Rechnung! □

20.3 Abstand und metrische Räume

Aus einer Norm läßt sich ein Abstand erhalten. Dieser ist auf allgemeineren Räumen definiert, insbesondere werden Vektorräume nicht vorausgesetzt.

Definition 20.13 Ein *metrischer Raum* ist eine Menge X zusammen mit einer Abbildung $d : X \times X \to \mathbb{R}_+$, dem *Abstand* oder der *Metrik*, wenn gilt:

(D1) $d(x, y) = 0 \quad \Leftrightarrow \quad x = y$,

(D2) $d(x, y) = d(y, x)$ (Symmetrie),

(D3) $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung).

Jeder normierte Vektorraum ist auch metrischer Raum mit Abstand $d(x, y) := \|x - y\|$. Für den $\mathbb{K}^n$ mit der aus dem Skalarprodukt erhaltenen Standardnorm gilt der *Satz des Pythagoras* $d(x, y) = \sqrt{|x_1 - y_1|^2 + \cdots + |x_n - y_n|}$.

21 Orthonormalsysteme

Es sei V wieder ein euklidischer bzw. unitärer Vektorraum. Wir haben gesehen, daß durch $d(v, w) = \|v - w\| = \sqrt{\langle v - w, v - w\rangle}$ ein Abstand zwischen Vektoren $v, w \in V$ erklärt ist. Außerdem läßt sich in *euklidischen Vektorräumen* der *Winkel* zwischen Vektoren erklären als

$$\cos \measuredangle(v, w) = \frac{\langle v, w\rangle}{\|v\| \, \|w\|} .$$

Wegen Cauchy-Schwarz ist die rechte Seite eine reelle Zahl aus $[-1, 1]$. Im $\mathbb{R}^n$ mit dem Standardskalarprodukt ist das die geometrische Winkelformel. Sie überträgt sich auf allgemeine euklidischen Vektorräume, ist aber nicht sinnvoll in unitären Vektorräumen. Jedoch übertägt sich der Begriff "senkrecht" auch auf den unitären Fall:

Definition 21.1 Sei V ein euklidischer bzw. unitärer Vektorraum.

i) Zwei Vektoren $v, w \in V$ heißen *orthogonal (bezüglich des Skalarprodukts)*, geschrieben $v \perp w$, wenn $\langle v, w\rangle = 0$.

ii) Zwei Untervektorräume $U, W \subset V$ heißen orthogonal, geschrieben $U \perp W$, wenn $u \perp w$ für alle $u \in U$ und $w \in W$.

iii) Ist $U \subset V$ ein Untervektorraum, dann heißt

$$U^\perp := \{v \in V \;:\; v \perp u \;\forall u \in U\}$$

das *orthogonale Komplement* von U in V.

iv) Eine Familie $(v_1, \dots, v_n)$ von Vektoren $v_i \in V$ heißt *orthogonal* oder *Orthogonalsystem*, wenn $v_i \perp v_j$ für alle $i \neq j$. Die Familie heißt *orthonormal* oder *Orthonormalsystem* (ONS), falls zusätzlich $\|v_i\| = 1$ für alle $1 \leq i \leq n$, und *Orthonormalbasis* (ONB), falls $(v_1, \dots, v_n)$ außerdem eine Basis von V ist[1].

Für Orthogonalsysteme gilt:

i) Ist in einer orthogonalen Familie $(v_1, \dots, v_n)$ der Nullvektor nicht enthalten, $v_i \neq 0$, so ist $(\frac{1}{\|v_1\|}v_1, \dots, \frac{1}{\|v_n\|}v_n)$ eine orthonormale Familie.

ii) Ist in einer orthogonalen Familie $(v_1, \dots, v_n)$ der Nullvektor nicht enthalten, $v_i \neq 0$, dann ist $(v_1, \dots, v_n)$ linear unabhängig: Sei $0 = \lambda_1 v_1 + \dots + \lambda_n v_n$, dann ergibt das Skalarprodukt mit sich selbst

$$0 = |\lambda_1|^2 \|v_1\|^2 + \dots |\lambda_n|^2 \|v_n\|^2 \qquad \Rightarrow \qquad \lambda_1 = \dots = \lambda_n = 0 \;.$$

iii) In einem ONS $(v_1, \dots, v_n)$ gilt $\langle v_i, v_j \rangle = \delta_{ij}$.

Einige wichtige Eigenschaften der Orthogonalität:

Satz 21.2 (Pythagoras) *Ist $(v_1, \dots, v_n)$ ein Orthogonalsystem, so gilt $\Big\| \sum_{i=1}^n v_i \Big\|^2 = \sum_{i=1}^n \|v_i\|^2$.*

Beweis. $\Big\| \sum_{i=1}^n v_i \Big\|^2 = \Big\langle \sum_{i=1}^n v_i \sum_{j=1}^n v_j \Big\rangle = \sum_{i,j=1}^n \langle v_i, v_j \rangle = \sum_{i=1}^n \langle v_i, v_i \rangle$. □

Satz 21.3 *Es sei $(V, \langle\;,\;\rangle)$ euklidischer oder unitärer Vektorraum, U Untervektorraum von V und $U^\perp$ das orthogonale Komplement. Dann ist $U^\perp$ ebenfalls Untervektorraum von V, und es gilt $U \cap U^\perp = \{0\}$.*

Beweis. Seien $w_1, w_2 \in U^\perp$ und $\lambda_1, \lambda_2 \in \mathbb{K}$, so gilt für beliebiges $u \in U$

$$\langle u, \lambda_1 w_1 + \lambda_2 w_2 \rangle = \lambda_1 \langle u, \lambda_1 \rangle + \lambda_2 \langle u, \lambda_2 \rangle = 0 \;,$$

also $\lambda_1 w_1 + \lambda_2 w_2 \in U^\perp$. Ist $u \in U \cap U^\perp$, so ist $0 = \langle u, u \rangle$, also $u = 0$. □

Satz 21.4 *Es sei $(V, \langle\;,\;\rangle)$ euklidischer oder unitärer Vektorraum, $U \subset V$ ein Untervektorraum und $v \in V$. Dann gibt es höchstens einen Vektor $P_U(v) \in U$ mit $v - P_U(v) \in U^\perp$. Falls existent, so heißt $P_U(v)$ die orthogonale Projektion von v auf U.*

[1]Orthonormalbasen gibt es auch in unendlich-dimensionalen euklidischen/unitären Vektorräumen. Die allgemeine Definition erfordert weitere Hilfsmittel aus der Analysis. Wir werden im nächsten Abschnitt mit den Fourierreihen einen wichtigen Fall behandeln.

Beweis. Seien $u_1, u_2 \in U$ mit $v - u_1 \in U^\perp$ und $v - u_2 \in U^\perp$. Dann ist

$$\underbrace{u_1 - u_2}_{\in U} = \underbrace{(v - u_2) - (v - u_1)}_{\in U^\perp} \in U \cap U^\perp \quad \Rightarrow \quad u_1 - u_2 = 0\,.$$

□

Satz 21.5 *Es sei* $(V, \langle\ ,\ \rangle)$ *euklidischer oder unitärer Vektorraum und* $(v_1, \ldots, v_n)$ *ein Orthonormalsystem. Sei* $U := \operatorname{span}(v_1, \ldots, v_n) \subset V$*. Dann gilt:*

i) $P_U(v) = \sum_{i=1}^{n} \langle v_i, v \rangle v_i$ *für alle* $V \in V$.

ii) $\|P_U(v)\|^2 = \sum_{i=1}^{n} |\langle v_i, v\rangle|^2 \leq \|v\|^2$ *für alle* $V \in V$ *(Besselsche Ungleichung)*

iii) *Ist* $v \in V$*, so gilt* $\|v - P_U(v)\| < \|v - u\|$ *für alle* $u \in U$ *mit* $u \neq P_U(v)$.

iv) $0 \leq \|v - P_U(v)\|^2 = \|v\|^2 - \|P_U(v)\|^2 \leq \|v\|^2$.

Beweis. i) Wir zeigen: $v - P_U(v) \in U^\perp$. Nach Satz 21.4 ist $P_U(v)$ dann eindeutig. Auf Grund der Linearität genügt es zu zeigen: $\langle v - P_U(v), v_j \rangle = 0$ für alle $j = 1, \ldots, n$:

$$\begin{aligned}\langle v_j, v - P_U(v) \rangle = \left\langle v_j, v - \sum_{i=1}^{n} \langle v_i, v \rangle v_i \right\rangle &= \langle v_j, v \rangle - \sum_{i=1}^{n} \langle v_i, v \rangle \langle v_j, v_i \rangle \\ &= \langle v_j, v \rangle - \langle v_j, v \rangle = 0\,.\end{aligned}$$

ii) und iv) Es gilt $v - P_U(v) \perp P_U(v)$ und damit nach Pythagoras

$$\|v\|^2 = \|(v - P_U(v)) + P_U(v)\|^2 = \|v - P_U(v)\|^2 + \|P_U(v)\|^2 \geq \|P_U(v)\|^2\,.$$

Andererseits folgt $\|v - P_U(v)\|^2 = \|v\|^2 - \|P_U(v)\|^2 \leq \|v\|^2$.

iii) Wegen $u \neq P_U(v)$ ist $\|P_U(v) - u\|^2 > 0$ und deshalb nach Pythagoras

$$\begin{aligned}\|v - u\|^2 = \|\underbrace{v - P_U(v)}_{\in U^\perp} + \underbrace{P_U(v) - u}_{\in U}\|^2 &= \|v - P_U(v)\|^2 + \|P_U(v) - u\|^2 \\ &> \|v - P_U(v)\|^2\,.\end{aligned}$$

□

Eigenschaft iii) aus Satz 21.5 hat folgende interessante Folgerung: Es sei $\varnothing \neq W \subset V$ eine Teilmenge. Definieren wir durch $d(v, W) := \inf_{w \in W} \|v - w\|$ den *Abstand* von v zu W, so gilt mit obigen Bezeichnungen $d(v, U) = \|v - P_U(v)\|$.

Der folgende Satz hebt die Bedeutung der Orthonormalbasen hervor:

Satz 21.6 (Fourier-Entwicklung) *Sei* $(V, \langle\ ,\ \rangle)$ *euklidischer oder unitärer Vektorraum und* $(v_1, \ldots, v_n)$ *ein Orthonormalsystem von* V*. Dann sind äquivalent:*

i) $(v_1, \ldots, v_n)$ *ist Orthonormalbasis.*

ii) *Für alle* $v \in V$ *gilt* $v = \sum_{i=1}^n \langle v_i, v \rangle v_i$ *(Fourier-Entwicklung)*

iii) *Für alle* $v, w \in V$ *gilt* $\langle v, w \rangle = \sum_{i=1}^n \langle v, v_i \rangle \langle v_i, w \rangle$ *(Parsevalsche Gleichung)*

iv) *Für alle* $v \in V$ *gilt* $\|v\|^2 = \sum_{i=1}^n |\langle v_i, v \rangle|^2$

v) *Ist* $v \in V$ *mit* $\langle v_i, v \rangle = 0$ *für alle* $i = 1, \ldots, n$, *so gilt* $v = 0$.

Beweis. i)⇒ii) Insbesondere ist $(v_1, \ldots, v_n)$ Basis, also gilt $v = \sum_{i=1}^n \lambda_i v_i$ mit eindeutig bestimmten $\lambda_i \in \mathbb{K}$. Dann ist $\langle v_j, v \rangle = \sum_{i=1}^n \lambda_i \langle v_j, v_i \rangle = \lambda_j$.

ii)⇒iii)

$$\langle v, w \rangle = \Big\langle \sum_{i=1}^n \langle v_i, v \rangle v_i, \sum_{j=1}^n \langle v_i, w \rangle v_i \Big\rangle = \sum_{i,j=1}^n \overline{\langle v_i, v \rangle} \langle v_i, w \rangle \langle v_i, v_j \rangle = \sum_{i=1}^n \langle v, v_i \rangle \langle v_i, w \rangle.$$

iii)⇒iv) Setze $w = v$.

iv)⇒v) Sei $\langle v_i, v \rangle = 0$ für alle $i = 1, \ldots, n$, so folgt $\|v\|^2 = \sum_{i=1}^n |\langle v_i, v \rangle| = 0$, also $v = 0$.

v)⇒i) Sei $U := \operatorname{span}(v_1, \ldots, v_n)$. Zu zeigen ist $V = U$. Angenommen, es gäbe ein $w \in V$ mit $w \notin U$. Also ist $w - P_U(w) \neq 0$. Wegen $w - P_U(w) \in U^\perp$ gilt $\langle w - P_u(w), v_j \rangle = 0$ für alle $i = 1, \ldots, n$ und somit $w - P_u(w) = 0$ nach v), Widerspruch. □

Satz 21.7 (Orthonormalisierungssatz) *Sei* V *ein endlich-dimensionaler euklidischer oder unitärer Vektorraum und sei* $W \subset V$ *ein Untervektorraum mit Orthonormalbasis* $(w_1, \ldots, w_m)$. *Dann gibt es eine Ergänzung zu einer Orthonormalbasis* $(w_1, \ldots, w_m, w_{m+1}, \ldots, w_n)$ *von* V.

Insbesondere besitzt jeder endlich-dimensionale euklidische oder unitäre Vektorraum eine Orthonormalbasis.

Beweis (Schmidtsches Orthonormalisierungsverfahren). Für $W = V$ ist alles klar. Andernfalls gibt es einen Vektor $v \in V$ mit $v \notin W$. Seine Projektion auf W ist

$$P_W(v) = \langle w_1, v \rangle w_1 + \cdots + \langle w_m, v \rangle w_m \ .$$

Damit ist $v - P_W(v) \in W^\perp$ und $v - P_W(v) \neq 0$ (sonst wäre $v \in W$). Setze $w_{m+1} := \frac{1}{\|v - P_W(v)\|}(v - P_W(v))$, dann ist die Familie $(w_1, \ldots, w_m, w_{m+1})$ orthonormal, und $W' := \operatorname{span}(w_1, \ldots, w_m, w_{m+1})$ ist $(m+1)$-dimensionaler Untervektorraum von V mit Orthonormalbasis $(w_1, \ldots, w_{m+1})$. Durch Wiederholung des Verfahrens erhält man eine Orthonormalbasis von V. □

Beispiel 21.8 Sei $V = \mathbb{R}^3$ und $U \subset V$ die von den Vektoren $u_1 = (2,1,2)$ und $u_2 = (1,2,7)$ aufgespannte Ebene. Gesucht ist die Projektion des Vektors $v = (1,4,9)$ auf U und der Abstand von v zu U.

Zunächst wird durch das Schmidtsche Orthonormalisierungsverfahren das System (u_1, u_2) in ein Orthonormalsystem überführt. Dazu ist $w_1 := \frac{u_1}{\|u_1\|} = \frac{1}{3}(2,1,2)$ und dann $P_{\mathbb{R}w_1}(u_2) = \langle w_1, u_2\rangle w_1 = \frac{1}{9}(2+2+14)\cdot(2,1,2) = (4,2,4)$. Es folgt $u_2 - P_{\mathbb{R}w_1}(u_2) = (-3,0,3)$ und damit $w_2 = \frac{1}{\sqrt{2}}(-1,0,1)$. Somit gilt

$$\begin{aligned} P_U(v) &= \langle w_1, v\rangle w_1 + \langle w_2, v\rangle w_2 \\ &= \tfrac{1}{9}(2+4+18)\cdot(2,1,2) + \tfrac{1}{2}(-1+0+9)\cdot(-1,0,1) \\ &= \left(\tfrac{16}{3}, \tfrac{8}{3}, \tfrac{16}{3}\right) + (-4,0,4) = \left(\tfrac{4}{3}, \tfrac{8}{3}, \tfrac{28}{3}\right). \end{aligned}$$

Schließlich ist

$$v - P_U(v) = (1,4,9) - \left(\tfrac{4}{3}, \tfrac{8}{3}, \tfrac{28}{3}\right) = \left(-\tfrac{1}{3}, \tfrac{4}{3}, -\tfrac{1}{3}\right)$$

und damit $d(v,U) = \|v - P_U(v)\| = \sqrt{2}$. $\triangleleft$

Schließlich können wir für endlich-dimensionale Untervektorräume das orthogonale Komplement genauer charakterisieren.

Satz 21.9 *Es sei $(V, \langle\ ,\ \rangle)$ ein euklidischer oder unitärer Vektorraum und $U \subset V$ ein endlich-dimensionaler Untervektorraum. Dann gilt:*

i) *$V = U \oplus U^\perp$. In diesem Fall heißt $U \oplus U^\perp$ die orthogonale direkte Summe.*

ii) *Ist $P_U : V \to U$ die orthogonale Projetion auf U, so ist $P_{U^\perp} = \mathrm{id}_V - P_U : V \to U^\perp$ die orthogonale Projetion auf $U^\perp$.*

Beweis. i) Nach Satz 21.7 besitzt U eine Orthonormalbasis $(w_1, \ldots, w_n)$. Nach Satz 21.5 ist dann die orthogonale Projektion eines beliebigen Vektors $v \in V$ auf U gegeben durch $P_U(v) = \sum_{i=1}^n \langle w_i, v\rangle w_i$. Damit ist $v = \underbrace{v - P_U(v)}_{\in U^\perp} + \underbrace{P_u(v)}_{\in U} \in U^\perp + U$, also $V = U + U^\perp$ und $U \cap U^\perp = \{0\}$ nach Satz 21.9.

ii) folgt aus $(\mathrm{id}_V - P_U)(v) = v - P_u(V)$ und i). $\square$

22 Fourier-Reihen

Definition 22.1 Eine Funktion $f : [0,T] \to \mathbb{C}$ heißt *stückweise stetig* auf $[0,T]$, wenn f auf $[0,T]$ beschränkt ist und es eine endliche Unterteilung $0 = t_0 < t_1 < \cdots < t_m = T$ gibt, so daß f auf jedem offenen Teilintervall $]t_{i-1}, t_i[$ stetig ist.

Die Summe zweier auf $[0,T]$ stückweise stetiger Funktionen f, g wird bezüglich der Vereinigung der Unterteilungen von f und von g definiert. Gleichheit von Funktionen $f, g \in \mathcal{C}_{sw}([0,T])$ wird erklärt als $f = g$ genau dann, wenn $(f -$

$g)(x) \neq 0$ für höchstens endlich viele Punkte $x \in \{t_0, \dots, t_m\}$. Diese Gleichheit ist eine Äquivalenzrelation, und mit $[f]$ werde die entsprechende Äquivalenzklasse zu einer stückweise stetigen Funktionen f bezeichnet. Die Äquivalenzklassenbildung überträgt sich auf Linearkombinationen, so daß

$$\mathcal{C}_{sw}([0,T]) = \{[f] \ : \ f \text{ ist stückweise stetig auf } [0,T]\}$$

ein Vektorraum ist. Für $[f],[g] \in \mathcal{C}_{sw}([0,T])$ mit gemeinsamer Unterteilung $0 = t_0 < t_1 < \dots < t_m = T$ definieren wir

$$\langle [f],[g]\rangle = \langle f,g\rangle = \frac{1}{T}\int_{[0,T]} dt\ \overline{f(t)}g(t) := \lim_{\epsilon\searrow 0}\sum_{i=1}^{m}\frac{1}{T}\int_{t_{i-1}+\epsilon}^{t_i-\epsilon} dt\ \overline{f(t)}g(t)\ . \qquad (*)$$

Dabei ist $2\epsilon < \min_i(t_i - t_{i-1})$ zu wählen. Die Definition ist unabhängig von der Wahl der Repräsentanten f,g der Äquivalenzklassen $[f],[g]$. Wegen der Stetigkeit der Funktionen auf beliebigen $[t_{i-1}+\epsilon, t_i-\epsilon]$ existiert das mehrfach uneigentliche Integral $\int_{[0,T]}$. Linearität von $\langle\ ,\ \rangle$ in der 2. Komponente und $\langle f,g\rangle = \overline{\langle g,f\rangle}$ sowie $\langle f,f\rangle \geq 0$ sind klar. Gleichheit $\langle f,f\rangle = 0$ genau für $f = [0] \in \mathcal{C}_{sw}([0,T])$ folgt aus der Stetigkeit in jedem Teilintervall $[t_{i-1}+\epsilon, t_i-\epsilon]$. Dabei bedeutet $f = [0]$, daß $f(x) \neq 0$ für höchstens endlich viele Punkte $x \in [0,T]$. Somit ist $\mathcal{C}_{sw}([0,T])$ mit $\langle\ ,\ \rangle$ ein unendlich-dimensionaler unitärer Vektorraum. Er enthält insbesondere alle stetigen Funktionen und alle Treppenfunktionen. Für stetige Funktionen ist $\int_{[0,T]} = \int_0^T$, und für Treppenfunktionen ist $\int_{[0,T]} = \int_0^T$ Im Sinne von Definition 19.2 aus dem letzten Semester.

Unser Ziel ist die Charakterisierung einer (unendlichen) Orthonormalbasis in $(\mathcal{C}_{sw}([0,T]), \langle\ ,\ \rangle)$.

Satz 22.2 *Für $n \in \mathbb{Z}$ werden durch $e_n(t) := e^{\mathrm{i}n\omega_T t}$ mit $\omega_T := \frac{2\pi}{T}$ stetige Funktionen $e_n \in \mathcal{C}([0,T]) \subset \mathcal{C}_{sw}([0,T])$ definiert. Dann ist $(e_n)_{n\in\mathbb{Z}}$ ein unendliches Orthonormalsystem für $(\mathcal{C}_{sw}([0,T]), \langle\ ,\ \rangle)$.*

Beweis. Wegen der Stetigkeit der e_k ist das Skalarprodukt $(*)$ durch das Riemannsche Integral über $[0,T]$ gegeben:

$$\begin{aligned}\langle e_k, e_l\rangle &= \frac{1}{T}\int_0^T dt\ e^{\mathrm{i}(l-k)\omega_T t}\\ &= \begin{cases} \dfrac{1}{T}\displaystyle\int_0^T dt\ 1 = 1 & \text{für } k = l\\ \dfrac{1}{\mathrm{i}T(l-k)\omega_T} e^{\mathrm{i}(l-k)\omega_T t}\Big|_0^T = \dfrac{\left(e^{\mathrm{i}(l-k)\cdot 2\pi} - 1\right)}{\mathrm{i}T(l-k)\omega_T} = 0 & \text{für } k \neq l\end{cases}\end{aligned}$$

□

Wir verallgemeinern nun die Definition von Orthonormalbasen auf unendlich-dimensionale Vektorräume mit Skalarprodukt. Dazu wird Konvergenz der unendlichen Reihen benötigt.

Definition 22.3 Es sei $(V, \| \ \|)$ ein normierter Vektorraum. Eine Folge $(v_n)_{n\in\mathbb{N}}$ von Vektoren $v_n \in V$ heißt *konvergent gegen $v \in V$ bezüglich der Norm $\| \ \|$*, wenn $\lim_{n\to\infty} \|v_n - v\| = 0$.

Wir hatten in Beispiel 20.11 auf $\mathcal{C}([a,b])$ verschiedene Normen eingeführt. Diese führen auf verschiedene Konvergenzbegriffe. So heißt eine Folge (f_n) stetiger Funktionen $f_n \in \mathcal{C}([a,b])$

- *gleichmäßig konvergent* gegen f, wenn
$\lim\limits_{n\to\infty} \|f_n - f\|_\infty = \lim\limits_{n\to\infty} \sup\limits_{x\in[a,b]} |f_n(x) - f(x)| = 0$,
- und *im quadratischen Mittel* konvergent, wenn
$\lim\limits_{n\to\infty} \|f_n - f\|_2 = \lim\limits_{n\to\infty} \displaystyle\int_a^b dx\, |f_n(x) - f(x)|^2 = 0$.

Aus gleichmäßiger Konvergenz folgt Konvergenz im quadratischen Mittel, die Umkehrung gilt aber nicht.

Definition 22.4 Es sei $f \in \mathcal{C}_{sw}([0,T])$. Für $N \in \mathbb{N}$ und $x \in [0,T]$ heißt

$$S_N f(x) := \sum_{k=-N}^{N} \langle e_k, f\rangle e_k(x) = \sum_{k=-N}^{N} \frac{1}{T}\Big(\int_{[0,T]} dt\; e^{-\mathrm{i}k\omega_T t} f(t)\Big) e^{\mathrm{i}k\omega_T x}$$

das *N-te Fourier-Polynom* von f. Wir sagen, die Reihe $S_N f = \sum_{k=-N}^{N} \langle e_k, f\rangle e_k$ konvergiert

- *punktweise* gegen f, falls $\lim_{N\to\infty} S_N f(x) = f(x)$ für alle $x \in [0,T]$,
- *gleichmäßig* gegen f, falls $\lim_{N\to\infty} \|S_N f - f\|_\infty = 0$,
- *im quadratischen Mittel* gegen f, falls $f \lim_{N\to\infty} \|S_N f - f\|_2 = 0$.

Aus gleichmäßiger Konvergenz folgt punktweise Konvergenz und Konvergenz im quadratischen Mittel, andere Implikationen gelten i.a. nicht.

Jedes Fourier-Polynom $S_N f$ ist stetig. Wegen $\overline{e_k} = e_{-k}$ gilt

$$\overline{S_N f} = \overline{\sum_{k=-N}^{N} \langle e_k, f\rangle e_k} = \sum_{k=-N}^{N} \langle f, e_k\rangle e_{-k} = \sum_{k=-N}^{N} \langle e_{-k}, \overline{f}\rangle e_{-k} = S_N \bar{f}\ .$$

Insbesondere ist für reellwertige Funktionen f auch $S_N f$ eine reellwertige Funktion.

Satz 22.5 (Fourier-Entwicklung) *Es seien $f, g \in \mathcal{C}_{sw}([a,b])$. Dann gilt:*

i) *Die Fourier-Reihe $Sf := \sum_{n=-\infty}^{\infty} \langle e_n, f\rangle e_n$ von f konvergiert im quadratischen Mittel gegen f.*

ii) $\langle f, g\rangle = \displaystyle\sum_{n=-\infty}^{\infty} \langle f, e_n\rangle\langle e_n, g\rangle$ *(Parsevalsche Gleichung).*

iii) $\displaystyle \frac{1}{T}\int_{[0,T]} dt\, \|f(t)\|^2 = \|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\langle e_n, f\rangle|^2\ .$

iv) *Ist* $\langle e_k, f\rangle = 0$ *für alle* $k \in \mathbb{Z}$, *so folgt* $f = [0]$.

Da es sich um analoge Aussagen zu Satz 21.6 handelt, sagt man, $(e_k)_{k\in\mathbb{Z}}$ ist Orthonormalbasis von $\mathcal{C}_{sw}([a,b])$.

Beweis. Der Hauptschritt ist der Beweis von i), den wir später führen.

i)⇒ii) Nach der Parsevalschen Gleichung Satz 21.6.iii) für den endlichdimensionalen Vektorraum $U = \operatorname{span}\{e_i \;:\; -N \leq i \leq N\}$ ist $\langle S_N f, S_N g\rangle = \sum_{n=-N}^{N} |\langle e_n, f\rangle|^2$. Andererseits erhalten wir mit Cauchy-Schwarz

$$\begin{aligned}
|\langle f, g\rangle - \langle S_N f, S_N g\rangle| &= |\langle f - S_N f, g\rangle + \langle S_N f, g - S_N g\rangle| \\
&\leq |\langle f - S_N f, g\rangle| + |\langle S_N f, g - S_N g\rangle| \\
&\leq \|f - S_N f\|_2\, \|g\|_2 + \|S_N f\|_2\, \|g - S_N g\|_2\ .
\end{aligned}$$

Das Quadrat stückweise stetiger Funktionen ist stückweise stetig und damit uneigentlich Riemann-integrierbar, also sind $\|g\|_2$ und $\|S_N f\|_2$ beschränkt, und nach i) gilt $\lim_{N\to\infty} \|f - S_N f\|_2 = 0$ und $\lim_{N\to\infty} \|g - S_N g\|_2 = 0$. Damit konvergiert die Summe in ii) gegen $\langle f, g\rangle$. Setzt man $g = f$, so folgt iii). Ist $\langle e_k, f\rangle = 0$, so folgt $\|f\| = 0$ und damit $f = [0]$, also iv).

Verbleibt der Beweis von i). Es genügt, reellwertige stückweise stetige Funktionen f (die nach Definition auch beschränkt sind) zu betrachten, für die das N-te Fourier-Polynom $S_N f$ ebenfalls reell ist. Da für $f = 0$ (als Äquivalenzklasse) die Aussage klar ist, können wir annehmen, daß $f \neq 0$ auf einer offenen Teilmenge von $[0,T]$ gilt. Dann ist auch $0 < \|f\|_\infty$, und durch Skalieren mit $\frac{1}{\|f\|_\infty}$ können wir uns dann auf reellwertige stückweise stetige Funktionen mit $\|f\|_\infty \leq 1$ beschränken. Diese sind Riemann-integrierbar. Folglich gibt es zu jedem $\epsilon > 0$ Treppenfunktionen ϕ, ψ auf $[0,T]$ mit $-1 \leq \phi \leq f \leq \psi \leq 1$ und $\int_{[0,T]} dt (\psi - \phi)(t) < \frac{T}{2}\epsilon^2$. Folglich gilt

$$\begin{aligned}
\|f - S_N f\|_2 &= \|(f-\phi) - S_N(f-\phi) + (\phi - S_N\phi)\|_2 \\
&\leq \|(f-\phi) - S_N(f-\phi)\|_2 + \|\phi - S_N\phi\|_2\ .
\end{aligned}$$

Nach Satz 21.5.iv) und wegen $-1 \leq \phi \leq f \leq \psi \leq 1$ gilt

$$\begin{aligned}
\|(f-\phi) - S_N(f-\phi)\|_2^2 &= \|f-\phi\|_2^2 - \|S_N(f-\phi)\|_2^2 \leq \|f-\phi\|_2^2 \\
&= \frac{1}{T}\int_{[0,T]} dt\, |(f-\phi)(t)|^2 \leq \frac{1}{T}\int_{[0,T]} dt\, |(\psi-\phi)(t)|^2 \\
&\leq \frac{1}{T}\int_{[0,T]} dt\, 2(\psi-\phi)(t) < \epsilon^2\ ,
\end{aligned}$$

also $\|(f-\phi) - S_N(f-\phi)\|_2 < \epsilon$. Das nächste Lemma liefert die Abschätzung $\|\phi - S_N\phi\|_2 < \epsilon$ für genügend großes N, womit die Behauptung bewiesen ist. □

Lemma 22.6 *Es sei ϕ eine Treppenfunktion auf $[0,T]$. Dann gibt es zu jedem $\epsilon > 0$ ein $N_0 \in \mathbb{N}$, so daß $\|\phi - S_N\phi\|_2 < \epsilon$ für alle $N \geq N_0$.*

Beweis. Wir führen zunächst den Beweis für die Treppenfunktion $\phi = \begin{cases} 1 & \text{für } 0 \leq x \leq a \\ 0 & \text{für } a < x \leq T \end{cases}$ mit $0 < a < T$. Dann gilt

$$\langle e_k, \phi\rangle = \frac{1}{T}\int_0^a dt\; e^{-\mathrm{i}k\omega_T t} = \begin{cases} \frac{a}{T} & \text{für } k = 0 \\ \frac{1}{-\mathrm{i}2\pi k}(e^{-\mathrm{i}k\omega_T a} - 1) & \text{für } k \neq 0 \end{cases}$$

und folglich

$$|\langle e_k, \phi\rangle|^2 = \begin{cases} \frac{a^2}{T^2} & \text{für } k = 0 \\ \frac{1}{2\pi^2 k^2}(1 - \cos(k\omega_T a)) & \text{für } k \neq 0 \end{cases}$$

(Verwende $(\cos x - \mathrm{i}\sin x - 1)(\cos x + \mathrm{i}\sin x - 1) = 2 - 2\cos x$). Nach Satz 21.5.iv) gilt

$$\begin{aligned} \|\phi - S_N\phi\|_2^2 = \|\phi\|_2^2 - \|S_N\phi\|_2^2 &= \frac{1}{T}\int_0^T dt\; (\phi(t))^2 - \sum_{k=-N}^{N} |\langle e_k, \phi\rangle|^2 \\ &= \frac{a}{T} - \frac{a^2}{T^2} - \frac{1}{\pi^2}\sum_{k=1}^{N}\frac{1}{k^2} + \frac{1}{\pi^2}\sum_{k=1}^{N}\frac{\cos(k\omega_T a)}{k^2}\;. \end{aligned}$$

In Beispiel 22.12 aus dem letzten Semester hatten wir

$$\sum_{k=1}^{\infty}\frac{\cos(kx)}{k^2} = \Big(\frac{\pi - x}{2}\Big)^2 - \frac{\pi^2}{12} \quad \text{für alle } x \in [0, 2\pi]\;, \qquad \sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6}$$

bewiesen. Setzt man $x = \frac{2\pi a}{T} \in [0, 2\pi]$, so ergibt sich

$$\lim_{N\to\infty}\Big(\frac{1}{\pi^2}\sum_{k=1}^{N}\frac{\cos(k\omega_T a)}{k^2} - \frac{1}{\pi^2}\sum_{k=1}^{N}\frac{1}{k^2}\Big) = \frac{a^2}{T^2} - \frac{a}{T}\;.$$

Folglich gibt es zu jedem $\epsilon' > 0$ ein $N_0 \in \mathbb{N}$ mit $\|\phi - \phi_N\|_2^2 < (\epsilon')^2$ für alle $N \geq N_0$ und beliebiges $a \in [0, T]$.

Eine beliebige Treppenfunktion läßt sich, bis auf die beim Integral unwesentlichen Werte an den Sprungstellen $0 = x_0 < x_1 \cdots < x_m = T$, darstellen als Linearkombination $\phi = \sum_{i=1}^m c_i\phi_i$, wobei $\phi_i(x) = \begin{cases} 1 & \text{für } 0 \leq x \leq x_i \\ 0 & \text{für } x_i < x \leq T \end{cases}$ Treppenfunktion von der obigen Form ist. Dann ist nach Dreiecksungleichung $\|\phi - \phi_N\|_2 \leq \sum_{i=1}^m |c_i|\|\phi_i - (\phi_i)_N\|_2 < (1 + \sum_{i=1}^m |c_i|)\epsilon'$ für alle $N \geq N_0$. Dann folgt die Behauptung für die Wahl $\epsilon' = \frac{\epsilon}{1+\sum_{i=1}^m |c_i|}$. □

Definition 22.7 Die Abbildung

$$\widehat{\ }: \mathcal{C}_{sw}([0,T]) \to \ell^2(\mathbb{Z}) \ , \quad f \mapsto \hat{f} = (\hat{f}(k))_{k\in\mathbb{Z}} \text{ mit } \hat{f}(k) := \langle e_k, f\rangle$$

heißt *Fourier-Transformation*. Dabei ist $\ell^2(\mathbb{Z}) = \{a : \mathbb{Z} \to \mathbb{C} \ : \ \sum_{k=-\infty}^{\infty} |a(k)|^2 < \infty\}$ der Hilbertsche Folgenraum mit dem zugehörigen Skalarprodukt $\langle a, b\rangle_{\ell^2(\mathbb{Z})} = \sum_{k=-\infty}^{\infty} \overline{a(k)}b(k)$,

Nach Satz 22.5.ii) ist die Fourier-Transformation eine *Isometrie*, d.h. $\langle f, g\rangle = \langle \hat{f}, \hat{g}\rangle_{\ell^2(\mathbb{Z})}$.

Definition 22.8 Eine Funktion $f \in \mathcal{C}([0,T])$ heißt stückweise stetig differenzierbar, wenn es eine Unterteilung $0{=}t_0{<}\ldots{<}t_m{=}T$ gibt, so daß f auf der offenen Teilmenge $[0, 2\pi] \setminus \{t_0, \ldots, t_m\} =]t_0, t_1[\cup]t_1, t_2[\cup \cdots \cup]t_{m-1}, t_m[$ stetig differenzierbar ist.

Satz 22.9 *Es sei f eine auf $[0,T]$ stückweise stetig differenzierbare Funktion mit $f(0) = f(T)$ (d.h f sei T-periodisch). Dann gilt*

i) $\widehat{f'}(k) = (\mathrm{i}k\omega_T)\hat{f}(k)$ *für alle* $k \in \mathbb{Z}$.

ii) $\lim\limits_{N\to\infty} \|S_N f - f\|_\infty = 0$, *d.h.* $(S_N f)_{N\in\mathbb{N}}$ *konvergiert gleichmäßig gegen* f.

Beweis. i) Es gilt $f' \in \mathcal{C}_{sw}([0,T])$ und damit nach partieller Integration

$$\begin{aligned}
\widehat{f'}(k) &= \frac{1}{T}\int_{[0,T]} dx\ f'(x)e^{-ik\omega_T x} = \lim_{\epsilon\searrow 0}\frac{1}{T}\sum_{l=1}^{m}\int_{t_{l-1}+\epsilon}^{t_l-\epsilon} dx\ f'(x)e^{-ik\omega_T x} \\
&= \lim_{\epsilon\searrow 0}\Big(\frac{1}{T}\sum_{l=1}^{m}\Big(f(t_l-\epsilon)e^{-ik\omega_T(t_l-\epsilon)} - f(t_{l-1}+\epsilon)e^{-ik\omega_T(t_{l-1}+\epsilon)} \\
&\qquad\qquad - \int_{t_{l-1}+\epsilon}^{t_l-\epsilon} dx\ f(x)(e^{-ik\omega_T x})'\Big)\Big) \\
&= \mathrm{i}k\omega_\mathrm{T}\lim_{\epsilon\searrow 0}\Big(\frac{1}{T}\sum_{l=1}^{m}\int_{t_{l-1}+\epsilon}^{t_l-\epsilon} dx\ f(x)\,e^{-ik\omega_T x}\Big) = ik\omega_T\hat{f}(k)\ ,
\end{aligned}$$

wegen der Stetigkeit von f und der T-Periodizität $f(t_0) = f(0) = f(T) = f(t_m)$.

ii) Nach Cauchy-Schwarz im unitären Vektorraum $\ell^2(\mathbb{N})$ gilt

$$\sum_{k=1}^{\infty}|\hat{f}(k)| = \sum_{k=1}^{\infty}\Big|\frac{1}{ik\omega_T}\widehat{f'}(k)\Big| \leq \underbrace{\sum_{k=1}^{\infty}\Big|\frac{1}{ik\omega_T}\Big|^2}_{=\frac{\pi^2}{6\omega_T^2}}\ \underbrace{\sum_{l=1}^{\infty}\big|\widehat{f'}(l)\big|^2}_{\hat{f}\in\ell^2(\mathbb{Z})} < \infty\ .$$

Analog für $\hat{f}(-k)$. Also folgt $\sum_{k\in\mathbb{Z}} |\hat{f}(k)| < \infty$, d.h. zu $\epsilon > 0$ gibt es ein $N(\epsilon)$, so daß für alle $N \geq N' \geq N(\epsilon)$ gilt

$$\begin{aligned}\big\|S_N f - S_{N'} f\big\|_\infty &:= \sup_{x\in[0,T]} \Big| \sum_{k=-N}^{N} \hat{f}(k) e^{ik\omega_T x} - \sum_{k=-N'}^{N'} \hat{f}(k) e^{ik\omega_T x}\Big| \\ &= \sup_{x\in[0,T]} \Big| \sum_{N'\leq|k|\leq N} \hat{f}(k) e^{ik\omega_T x}\Big| \leq \sum_{N'\leq|k|\leq N} \big|\hat{f}(k)\big| < \epsilon\ .\end{aligned}$$

Damit ist in jedem Punkt $x \in [0,T]$ die Folge $((S_N f)(x))_{N\in\mathbb{N}}$ komplexer Zahlen eine Cauchy-Folge, welche *punktweise* gegen eine Grenzfunktion $(S_\infty f)(x) = \lim_{N\to\infty}(S_N f)(x))$ konvergiert. Damit kann in obiger Abschätzung $N' \to \infty$ genommen werden, so daß

$$\sup_{x\in[0,T]} \big\|S_N f - S_\infty f\big\|_\infty < \epsilon \qquad \text{für alle } N \geq N(\epsilon)\ .$$

Somit ist $(S_N f)_{N\in\mathbb{N}}$ eine Folge stetiger Funktionen, die *gleichmäßig* gegen die Grenzfunktion $S_\infty f$ konvergiert. Nach Satz 12.12 aus dem letzten Semester ist $S_\infty f$ dann stetig. Aus gleichmäßiger Konvergenz folgt Konvergenz im quadratischen Mittel, damit konvergiert $(S_N f)_{N\in\mathbb{N}}$ auch im quadratischen Mittel gegen $S_\infty f$. Andererseits konvergiert $(S_N f)_{N\in\mathbb{N}}$ nach Satz 22.5 im quadratischen Mittel gegen f. Nach Dreiecksungleichung gilt dann $\|f - S_\infty f\|_2 = 0$, und da f und $S_\infty f$ stetig sind, folgt $f = S_\infty f = \sum_{N\to\infty} S_N f$. □

Durch Induktion folgt $\widehat{f^{(n)}}(k) = (\mathrm{i}k\omega_T)^n \hat{f}(k)$ für eine n-mal stückweise stetig differenzierbare Funktion f, die $(n-1)$-mal stetig differenzierbar ist mit $f^{(p)}(0) = f^{(p)}(T)$ für alle $p = 0, \dots, n-1$.

Die Fourier-Entwicklung stückweise stetig differenzierbarer Funktionen erlaubt oft die Bestimmung des Grenzwertes von Reihen.

Beispiel 22.10 Gesucht ist die Fourier-Entwicklung der Funktion $f(x) = \cos(ax)$, für $a \notin \mathbb{Z}$, im Intervall $[-\pi, \pi]$. Wir setzen sie (2π)-periodisch auf $\mathbb{R}$ fort. Dann kann im Fourier-Integral über ein beliebiges Intervall der Länge $T = 2\pi$ integriert werden, z.B. über $[-\pi, \pi]$. Also haben wir für $a \notin \mathbb{Z}$

$$\begin{aligned}\hat{f}(k) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} dt\ e^{-\mathrm{i}kt}\cos(at) = \frac{1}{4\pi}\int_{-\pi}^{\pi} dt\ (e^{-\mathrm{i}(k-a)t} + e^{-\mathrm{i}(k+a)t}) \\ &= \frac{1}{-4\pi\mathrm{i}(k-a)}(e^{-\mathrm{i}(k-a)\pi} - e^{\mathrm{i}(k-a)\pi}) + \frac{1}{-4\pi\mathrm{i}(k+a)}(e^{-\mathrm{i}(k+a)\pi} - e^{\mathrm{i}(k+a)\pi}) \\ &= \frac{1}{2\pi}(-1)^k \sin(a\pi)\Big(\frac{1}{(a-k)} + \frac{1}{a+k}\Big)\ .\end{aligned}$$

Somit gilt wegen $f(k) = f(-k)$

$$f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{ikx} = \hat{f}(0) + 2\sum_{k=1}^{\infty} \hat{f}(k)\cos(kx)$$
$$= \frac{\sin(a\pi)}{a\pi} + \sum_{k=1}^{\infty} \frac{(-1)^k \sin(a\pi)}{\pi} \frac{2a}{a^2-k^2}\cos(kx)$$

also

$$\cos(ax) = \frac{2a\sin(a\pi)}{\pi}\Big(\frac{1}{2a^2} + \sum_{k=1}^{\infty} \frac{(-1)^k}{a^2-k^2}\cos(kx)\Big)\ .$$

Setzt man $x = \pi$, so folgt die Partialbruchzerlegung des Cotangens

$$\cot(a\pi) = \frac{1}{a\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty}\Big(\frac{1}{a-k} + \frac{1}{a+k}\Big)\ .$$

oder die Summenformel $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2-a^2} = \frac{\pi}{2a}\Big(\frac{1}{a\pi} - \cot(a\pi)\Big)$ für $a \notin \mathbb{Z}$. $\triangleleft$

Bemerkungen.

i) In Satz 22.5 wird nicht behauptet, daß stets $Sf \in \mathcal{C}_{sw}([0,T])$ gilt. Der Grund ist, daß $\mathcal{C}_{sw}([0,T])$ nicht vollständig ist. Es gibt also Folgen (f_n) stückweise stetiger Funktionen, die zwar im quadratischen Mittel konvergent sind, für deren Grenzwert aber $\lim_{n\to\infty} f_n \notin \mathcal{C}_{sw}([0,T])$ gilt. Tatsächlich stellt sich heraus, daß Satz 22.5 sogar für eine deutlich größere Klasse von Funktionen gilt, nämlich für die quadratisch Lebesgue-integrierbaren Funktion $L^2([0,T])$, die wir im 3. Semester einführen.

ii) Da $\mathcal{C}([0,T])$ in gewisser Weise die "falsche" Klasse von Funktionen für die Fourier-Reihe ist, gibt es einige Überraschungen: Es gibt Beispiele stetiger Funktionen $f \in \mathcal{C}([0,T])$, für die die Fourierreihe $Sf := \sum_{k=-\infty}^{\infty}\langle e_k, f\rangle e_k$ *nicht in jedem Punkt* $x \in [0,T]$ *konvergent ist*, und selbst wenn sie konvergiert, muß der Grenzwert nicht mit $f(x)$ übereinstimmen. Das folgt schon aus der Tatsache, daß wegen $e_k(0) = e_k(T)$ im Fall der Konvergenz stets $(Sf)(0) = (Sf)(T)$ gilt, während $f(0) \neq f(T)$ sein kann.

iii) Nach Satz 22.9 haben wir punktweise (sogar gleichmäßige) Konvergenz der Fourier-Reihe für stückweise stetig differenzierbare Funktionen. Man kann etwas allgemeiner folgendes zeigen: Existiert für eine Funktion $f \in \mathcal{C}([0,T])$ in $x \in [0,T]$ sowohl die linksseitige als auch die rechtsseitige Ableitung, dann konvergiert $(Sf)(x)$ gegen den arithmetischen Mittelwert $\frac{1}{2}(f(x_+)+f(x_-)) := \lim_{\epsilon\searrow 0}\frac{1}{2}(f(x+\epsilon)+f(x-\epsilon))$, mit $T_+ \mapsto 0_+$ und $0_- \mapsto T_-$.

iv) Oft definiert man die Fourier-Reihe für *T-periodische (stetige) Funktionen*, also (stetige) Funktionen $f : \mathbb{R} \to \mathbb{C}$ mit $f(x+T) = f(x)$ für alle $x \in \mathbb{R}$. Dann kann man statt $\int_0^T$ über ein beliebiges anderes Intervall der Länge T integrieren, $\int_a^{a+T}$. Insbesondere ist $f(0) = f(T)$.

v) Statt $(e_k)_{k\in\mathbb{Z}}$ kann auch das reelle Orthogonalsystem $((c_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}\setminus\{0\}})$ benutzt werden, mit $c_n(x) = \cos(n\omega_T x)$ und $s_n(x) = \sin(n\omega_T x)$. Zu beachten ist dann eine um einen Faktor 2 unterschiedliche Normierung von c_0 zu c_n, s_n mit $n > 0$!

23 Selbstadjungierte und unitäre Endomorphismen

Definition 23.1 Sei V ein euklidischer bzw. unitärer Vektorraum und $F \in \mathrm{End}(V)$. Ein Endomorphismus $F^* \in \mathrm{End}(V)$ heißt zu F *adjungiert*, wenn

$$\langle v, F(w)\rangle = \langle F^*(v), w\rangle \qquad \forall v, w \in V \ .$$

Satz 23.2 *Sei V ein euklidischer bzw. unitärer Vektorraum. Falls der zu F adjungierte Endomorphismus $F^* \in \mathrm{End}(V)$ existiert, so gilt:*

i) *F^* ist eindeutig.*

ii) $\ker(F^*) = (\mathrm{im}(F))^\perp$

iii) $\mathrm{im}(F^*) = (\ker(F))^\perp$

Beweis. i) Gäbe es zwei Lösungen F_1^* und F_2^* mit $\langle v, F(w)\rangle = \langle F_1^*(v), w\rangle = \langle F_2^*(v), w\rangle$ für alle $v, w \in V$, so folgt

$$0 = \langle F_1^*(v), w\rangle - \langle F_2^*(v), w\rangle = \langle F_1^*(v) - F_2^*(v), w\rangle$$

für alle v, w, insbesondere für $w = F_1^*(v) - F_2^*(v)$. Damit ist $F_1^* = F_2^*$.

ii) Sei $v \in \big(\mathrm{im}(F)\big)^\perp$, so ist $0 = \langle v, F(w)\rangle = \langle F^*(v), w\rangle$ für alle $w \in V$, und damit $\ker(F^*) = \big(\mathrm{im}(F)\big)^\perp$.

iii) Sei $w \in \ker(F)$, dann ist $0 = \langle v, F(w)\rangle = \langle F^*(v), w\rangle$ für alle $v \in V$. Das bedeutet $\mathrm{im}(F^*) = (\ker(F))^\perp$. □

Satz 23.3 *Sei V ein endlich-dimensionaler euklidischen bzw. unitären Vektorraums Dann gibt es zu jedem $F \in \mathrm{End}(V)$ den adjungierten Endomorphismus F^*, und bezüglich einer Orthonormalbasis $\mathcal{B}$ von V gilt $M^{\mathcal{B}}_{\mathcal{B}}(F^*) = \big(M^{\mathcal{B}}_{\mathcal{B}}(F)\big)^* := \overline{\big(M^{\mathcal{B}}_{\mathcal{B}}(F)\big)^t}$.*

Beweis. Nach Satz 21.7 besitzt V eine Orthonormalbasis $\mathcal{B} = (v_1, \dots, v_n)$. In dieser Basis sei $F(v_j) = \sum_{i=1}^n a_{ij} v_i$, d.h. $M^{\mathcal{B}}_{\mathcal{B}}(F) = (a_{ij})$. Dann gilt

$$\langle v_k, F(v_j)\rangle = \sum_{i=1}^n a_{ij}\langle v_k, v_i\rangle = a_{kj} = \sum_{i=1}^n \langle \overline{a_{ki}} v_i, v_j\rangle \ .$$

Damit hat $F^*(v_k) := \sum_{i=1}^n b_{ik} v_i$ mit $b_{ik} = \overline{a_{ki}}$ die Eigenschaft eines adjungierten Endomorphismus, und $M_{\mathcal{B}}^{\mathcal{B}}(F^*) = (b_{ij}) = \overline{(a_{ij})^t}$. □

Definition 23.4 Es sei $(V, \langle\ ,\ \rangle)$ ein endlich-dimensionaler euklidischer oder unitärer Vektorraum. Ein Endomorphismus $F \in \mathrm{End}(V)$ heißt

i) *selbstadjungiert* (für $\mathbb{K} = \mathbb{R}$ auch *symmetrisch*), wenn $F = F^*$ gilt, d.h. $\langle v, F(w)\rangle = \langle F(v), w\rangle$ für alle $v, w \in V$,

ii) *orthogonal* bzw. *unitär*, wenn $F \circ F^* = F^* \circ F = \mathrm{id}_V$ gilt, d.h. $\langle F(v), F(w)\rangle = \langle v, w\rangle$ für alle $v, w \in V$,

iii) *normal*, falls $F \circ F^* = F^* \circ F$ gilt.

Orthogonale, unitäre und selbstadjungierte Endomorphismen sind folglich auch normal. Die Definition überträgt sich wegen Satz 23.3 auf Matrizen: Eine Matrix $A \in M(n, K)$ heißt *selbstadjungiert* oder *hermitesch*, wenn $A = A^* = \overline{A^t}$ gilt, und *orthogonal* bzw. *unitär*, wenn $A^* = A^{-1}$ ist. Es folgt $|\det A| = 1$ für orthogonale bzw. unitäre Matrizen. Die Bedingungen $A^t \cdot A = E_n$ bzw. $A^* \cdot A = E_n$ bedeuten, daß die Spalten von A eine Orthonormalbasis von $\mathbb{K}^n$ bilden. Analog bedeutet $A \cdot A^t = E_n$ bzw. $A \cdot A^* = E_n$, daß die Zeilen von A eine Orthonormalbasis von $\mathbb{K}^n$ bilden.

Satz 23.5

- $O(n) := \{A \in GL(n;\mathbb{R}) : A^{-1} = A^t\}$ *ist Untergruppe der* $GL(n;\mathbb{R})$ *(orthogonale Gruppe).*
- $SO(n) := \{A \in O(n) : \det A = 1\}$ *ist Untergruppe der* $GL(n;\mathbb{R})$ *(spezielle orthogonale Gruppe).*
- $U(n) := \{A \in GL(n;\mathbb{C}) : A^{-1} = A^*\}$ *ist Untergruppe der* $GL(n;\mathbb{C})$ *(unitäre Gruppe).*
- $SU(n) := \{A \in U(n) : \det A = 1\}$ *ist Untergruppe der* $GL(n;\mathbb{C})$ *(spezielle unitäre Gruppe).*

Beweis. Für $A, B \in O(n)$ ist $(AB)^{-1} = B^{-1}A^{-1} = B^tA^t = (AB)^t$ sowie $(A^{-1})^{-1} = A = (A^{-1})^t$, damit ist $O(n)$ Untergruppe. Für $A, B \in SO(n)$ ist $\det(AB) = (\det A)(\det B) = 1$, damit ist $SO(n)$ Untergruppe. Analog für $U(n)$ und $SU(n)$. □

Beispiel 23.6 Im $\mathbb{R}^2$ mit dem kanonischen Skalarprodukt ist die Standardbasis eine Orthonormalbasis. Die Bedingung für orthogonale Matrizen lautet

$$O(2) := \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \begin{array}{l} a, b, c, d \in \mathbb{R}\ ,\ A^t \cdot A = E_2 \\ \Leftrightarrow \quad a^2 + b^2 = c^2 + d^2 = 1 \text{ und } ac + bd = 0 \end{array} \right\} .$$

Setzen wir $a = \cos\phi$, $b = \sin\phi$, $c = \sin\theta$, $d = \cos\theta$, so verbleibt $\sin(\phi + \theta) = 0$ mit den beiden verschiedenen Lösungen $\phi = -\theta$ und $\phi = \pi - \theta$. Somit gilt:

$$O(2) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \ : \ \theta \in [0, 2\pi[\right\} \cup \left\{ \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} \ : \ \phi \in [0, 2\pi[\right\}.$$

Die erste Menge beschreibt Drehungen im $\mathbb{R}^2$ und die zweite eine Kombination aus Drehungen und einer Spiegelung. Das sind gerade die Transformationen im $\mathbb{R}^2$, welche Winkel und Abstände erhalten. Die Gruppe $SO(2)$ besteht nur aus der ersten durch θ parametrisierten Menge. $\triangleleft$

Satz 23.7 *Ist $F \in \mathrm{End}(V)$ orthogonal bzw. unitär, so gilt*

i) $\|F(v)\| = \|v\| \qquad \forall v \in V$.

ii) *Ist umgekehrt $\|F(v)\| = \|v\|$ für alle $v \in V$, so ist F orthogonal bzw. unitär.*

iii) $v \perp w \quad \Rightarrow \quad F(v) \perp F(w)$.

iv) *F ist Isomorphismus und F^{-1} ist ebenfalls orthogonal bzw. unitär.*

v) *Ist $\lambda \in \mathbb{K}$ Eigenwert von F, so ist $|\lambda| = 1$.*

Beweis. i) und iii) sind klar. Die Umkehrung ii) von i) folgt aus der Polarisationsformel.

iv) Aus i) folgt, daß F injektiv ist. Wegen $F(F^*(v)) = v$ ist F surjektiv, also bijektiv. Die Ersetzung $v \mapsto F^{-1}(v)$ liefert Orthogonalität/Unitarität von F^{-1}.

v) Ist v Eigenvektor von F zum Eigenwert λ, so ist

$$\|v\| = \|F(v)\| = \|\lambda v\| = |\lambda|\|v\| \qquad \Rightarrow \quad |\lambda| = 1 \text{ wegen } v \neq 0\,. \qquad \square$$

Damit erhalten orthogonale/unitäre Endomorphismen die Normen, Abstände und Winkel und haben somit eine wichtige geometrische Interpretation als *Isometrien.*

Satz 23.8 *Ist $F \in \mathrm{End}(V)$ selbstadjungiert, dann sind alle Eigenwerte von F reell. Insbesondere hat eine hermitesche Matrix nur reelle Eigenwerte.*

Beweis. Ist $F(v) = \lambda v$ mit $v \neq 0$, so gilt

$$\lambda\langle v, v\rangle = \langle v, F(v)\rangle = \langle F(v), v\rangle = \overline{\lambda}\langle v, v\rangle$$

und damit $\lambda = \overline{\lambda}$. $\square$

Satz 23.9 *Es sei $(V, \langle\ .\ \rangle)$ ein endlich-dimensionaler euklidischer oder unitärer Vektorraum und $F \in \mathrm{End}(V)$ normal. Dann gilt:*

i) $\ker F = \ker F^*$

ii) *Ist* λ *Eigenwert von* F*, dann ist* $\bar{\lambda}$ *Eigenwert von* F^**, und es gilt* $\mathrm{Eig}(F;\lambda) = \mathrm{Eig}(F^*;\bar{\lambda})$*.*

iii) *Sind* λ_1, λ_2 *Eigenwerte von* F *mit* $\lambda_1 \neq \lambda_2$*, so gilt* $\mathrm{Eig}(F;\lambda_1) \perp \mathrm{Eig}(F;\lambda_2)$*.*

Beweis. i) Sei $v \in \ker F$, so folgt

$$0 = \langle F(v), F(v)\rangle = \langle F^* \circ F(v), v\rangle = \langle F \circ F^*(v), v\rangle = \langle F^*(v), F^*(v)\rangle \ ,$$

also auch $v \in \ker F^*$. Ebenso $\ker F^* \subset \ker F$.

ii) Es gilt $(F - \lambda \mathrm{id}_V)^* = F^* - \overline{\lambda}\mathrm{id}_V$. Mit F ist auch $F - \lambda\mathrm{id}_V$ normal. Dann folgt aus i)

$$\mathrm{Eig}(F;\lambda) = \ker(F - \lambda\,\mathrm{id}_V) = \ker(F - \lambda\,\mathrm{id}_V)^* = \ker(F^* - \overline{\lambda}\,\mathrm{id}_V) = \mathrm{Eig}(F^*;\bar{\lambda}) \ .$$

Sei $v \in \mathrm{Eig}(F;\lambda_1)$, $w \in \mathrm{Eig}(F;\lambda_2)$, so folgt

$$\lambda_2\langle v, w\rangle = \langle v, \lambda_2 w\rangle = \langle v, F(w)\rangle = \langle F^*(v), w\rangle = \langle \overline{\lambda_1} v, w\rangle = \lambda_1\langle v, w\rangle$$

also $0 = (\lambda_2 - \lambda_1)\langle v, w\rangle$. □

Theorem 23.10 *Es sei* $(V, \langle\ .\ \rangle)$ *ein endlich-dimensionaler* <u>*unitärer*</u> *Vektorraum und* $F \in \mathrm{End}(V)$ *normal. Dann besitzt* V *eine Orthonormalbasis* $\mathcal{B} = (v_1, \dots, v_n)$ *aus Eigenvektoren von* F*. Insbesondere ist* F *diagonalisierbar.*

Es seien $\lambda_1, \dots, \lambda_k$ die paarweise verschiedenen Eigenwerte von F und $\mathrm{Eig}(F;\lambda_i)$, $i = 1, \dots, k$ die zugehörigen paarweise orthogonalen Eigenräume. Nach Satz 21.7 besitzt $\mathrm{Eig}(F;\lambda_i)$ eine Orthonormalbasis $\mathcal{B}_i = (v_{i1}, \dots, v_{in_i})$. Nach Satz 23.9 ist $\mathcal{B}' = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ ein Orthonormalsystem aus Eigenvektoren von F. Wir zeigen: $\mathcal{B}'$ ist Orthonormalbasis von V.

Sei $W := \bigoplus_{i=1}^k \mathrm{Eig}(F;\lambda_i)$. Nach Satz 21.9 ist $V = W \oplus W^\perp$. Angenommen, es gäbe ein $0 \neq v \in W^\perp$. Dann gilt für beliebiges $w \in W$ die Identität $\langle F(v), w\rangle = \langle v, F^*(w)\rangle = 0$, da $F^*(W) \subset W$. Somit ist $F\big|_{W^\perp} \in \mathrm{End}(W^\perp)$ und damit nach Satz 19.5 in $W^\perp$ trigonalisierbar. Insbesondere besitzt $F\big|_{W^\perp}$ einen Eigenvektor (Definition 19.3). Dieser ist aber auch Eigenvektor von $F \in \mathrm{End}(V)$ und damit bereits in W enthalten, Widerspruch. □

Wir können nicht erwarten, daß sich dieses Theorem auf den reellen Fall überträgt, da Endomorphismen eines reellen Vektorraums nicht triagonalisierbar sein müssen. Insbesondere lassen sich die Matrizen aus $O(2)$ nicht diagonalisieren. Es gilt aber:

Satz 23.11 *Jeder* <u>*selbstadjungierte*</u> *Endomorphismus eines endlich-dimensionalen euklidischen Vektorraums* V *ist diagonalisierbar.*

Beweis. Wie im vorigen Theorem bis zur Konstruktion von $F \in \mathrm{End}(W^\perp)$. Einziger Unterschied ist, daß $W^\perp = V$ sein könnte. Offenbar ist $F\big|_{W^\perp}$ auch auf $W^\perp$ selbstadjungiert. Das charakteristische Polynom zerfällt zunächst über $\mathbb{C}$ in komplexe Linearfaktoren. Nach Satz 23.8 sind alle Eigenwerte reell, d.h. das charakteristische Polynom zerfällt sogar in reelle Linearfaktoren. Damit ist $F\big|_{W^\perp} \in \mathrm{End}(W^\perp)$ in $W^\perp$ reell trigonalisierbar, und es gibt einen Eigenvektor. □

Satz 23.12 *Ist $A \in M(n, \mathbb{K})$ eine reelle symmetrische bzw. hermitesche Matrix, dann gibt es eine orthogonale bzw. unitäre Matrix $T \in O(n)$ bzw. $T \in U(n)$ mit*

$$A = T\Lambda T^* \,, \qquad \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} ,$$

wobei $\lambda_i \in \mathbb{R}$ die Eigenwerte von A sind.

Beweis. Die Matrix $A = \overline{A^t} = (a_{ij})$, also $a_{ji} = \overline{a_{ij}}$, definiert einen selbstadjungierten Endomorphismus $F = F^* \in \mathrm{End}(\mathbb{K}^n)$ durch $F(e_j) = \sum_{i=1}^n a_{ij} e_i$, wobei $(e_1, \dots, e_n)$ die Standardbasis ist. Sei $(v_1, \dots, v_n)$ eine orthonormale Basis von $\mathbb{K}^n$ aus Eigenvektoren von F, d.h. $F(v_i) = \lambda_i v_i$. Wir zerlegen v_i nach der Standardbasis: $v_i = \sum_{k=1}^n t_{ki} e_k$. Aus der Orthonormalität folgt $\langle v_i, v_j \rangle = \sum_{k=1}^n \overline{t_{ki}} t_{kj} = \delta_{ij}$, und damit $T^*T = E_n$ für $T = (t_{ij})$. Andererseits gilt nach der Parsevalschen Gleichung

$$\delta_{kj} = \langle e_k, e_j \rangle = \sum_{i=1}^n \langle e_k, v_i \rangle \langle v_i, e_j \rangle = \sum_{i=1}^n t_{ki} \overline{t_{ji}} \,,$$

also $TT^* = E_n$ und damit $T \in O(n)$ bzw. $T \in U(n)$. Nun gilt

$$F(v_k) = \sum_{j=1}^n t_{jk} F(e_j) = \sum_{i,j=1}^n t_{jk} a_{ij} e_i = \lambda_k v_k = \sum_{l=1}^n \lambda_k t_{lk} e_l \,.$$

Da die e_i linear unabhängig sind, folgt $\sum_{j=1}^n a_{ij} t_{jk} = t_{ik} \lambda_k$ für alle k. Wir multiplizieren mit $\overline{t_{lk}}$ und summieren über k:

$$\sum_{k,j=1}^n a_{ij} t_{jk} \overline{t_{lk}} = a_{il} = \sum_{k=1}^n t_{ik} \lambda_k \overline{t_{lk}} = (T\Lambda T^*)_{il} \,. \qquad \square$$

Ist $(V, \langle\ ,\ \rangle)$ ein euklidischer bzw. unitärer Vektorraum, dann definiert jeder Endomorphismus $F \in \mathrm{End}(V)$ eine Abbildung $L : V \times V \to \mathbb{K}$ durch $L(v, w) := \langle v, F(w) \rangle$, die linear in der 2. Komponente (Axiom (S1) aus Definition 20.1) und semilinear in der 1. Komponente ist. Abbildungen $L : V \times V \to \mathbb{K}$ mit diesen beiden Eigenschaften heißen *Bilinearformen* bzw. *Sesquilinearformen.* Gilt zusätzlich (S2) aus Definition 20.1, so heißt die Bilinearform *symmetrisch* bzw. die Sesquilinearform *hermitesch*. Wegen

$$\langle v, F(w)\rangle = \langle F^*(v), w\rangle = \overline{\langle w, F^*(v)\rangle}$$

(falls F^* existiert) gilt (S2) für $L(v,w) := \langle v, F(w)\rangle$ genau dann, wenn F selbstadjungiert ist. Schließlich kann man sich fragen, wann (S3) gilt:

Definition 23.13 Sei V ein Vektorraum über $\mathbb{K}$. Eine symmetrische Bilinearform bzw. hermitesche Sesquilinearform $L : V \times V \to \mathbb{K}$ heißt

- *positiv semidefinit*, wenn $L(v,v) \geq 0$ für alle $v \in V$,
- *nicht ausgeartet*, wenn $L(v,v) = 0 \quad \Leftrightarrow \quad v = 0$,
- *positiv definit*, wenn L positiv semidefinit und nicht-ausgeartet ist.

Ein Skalarprodukt auf V ist damit eine positiv definite symmetrische Bilinearform bzw. positiv definite hermitesche Sesquilinearform.

Wir interessieren uns nun für die Frage, wie man entscheiden kann, ob eine symmetrische Bilinearform positiv definit ist (und somit ein Skalarprodukt definiert).

Satz 23.14 *Es sei $(V, \langle\ ,\ \rangle)$ ein euklidischer bzw. unitärer Vektorraum. Die durch einen selbstadjungierten Endomorphismus $F = F^* \in \mathrm{End}(V)$ über $L(v,w) = \langle v, F(w)\rangle$ definierte Bilinearform bzw. Sesquilinearform L ist genau dann positiv definit, wenn alle Eigenwerte von F positiv sind.*

Beweis. Nach Theorem 23.10 bzw. Satz 23.11 besitzt V eine Orthonormalbasis $\mathcal{B} = (v_1, \ldots, v_n)$ aus Eigenvektoren von F. Es sei $F(v_i) = \lambda_i v_i$ mit $\lambda_i \in \mathbb{R}$ nach Satz 23.8. Dann gilt für einen beliebigen Vektor $v = \mu_1 v_1 + \cdots + \mu_n v_n \in V$

$$\begin{aligned} L(v,v) = \langle v, F(v)\rangle &= \Big\langle \sum_{j=1}^n \mu_j v_j, \sum_{j=1}^n \mu_i F(v_i) \Big\rangle = \sum_{i,j=1}^n \overline{\mu_j}\mu_i\lambda_i \langle v_i, v_j\rangle \\ &= \lambda_1|\mu_1|^2 + \cdots + \lambda_n|\mu_n|^2\,. \end{aligned}$$

Daraus folgt die Behauptung. □

Allerdings ist die Bestimmung der Eigenwerte meist nur numerisch möglich. Ein einfacheres Kriterium im reellen Fall kann wie folgt erhalten werden

Satz 23.15 *Sei $A \in M(n, \mathbb{R})$ eine symmetrische Matrix und $A_k \in M(k, \mathbb{R})$ die linke obere Teilmatrix von A aus k Zeilen und Spalten. Dann ist A genau dann positiv definit, wenn* $\det A_k > 0$ *für alle* $1 \leq k \leq n$.

Beweis. ($\Rightarrow$) Ist $A = A^t \in M(n, \mathbb{R})$ positiv definit, dann ist $\det A = \det(T\Lambda T^t) = \det\Lambda(\det T)^2 = \lambda_1 \cdot \cdots \cdot \lambda_n > 0$. Die Matrix A_k beschreibt die Einschränkung der durch A definierten Bilinearform auf den Untervektorraum

$$V_k := \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ x_{k+1} = \cdots = x_n = 0\} \, .$$

Ist A positiv definit, dann ist auch $\langle x, Ax\rangle > 0$ für alle $x \in V_k$ mit $x \neq 0$. Folglich ist $\det A_k > 0$.

($\Leftarrow$) wird durch Induktion nach n bewiesen. Sei in Blockdarstellung

$$A = A^t = \left(\begin{array}{c|c} B & b \\ \hline b^t & c \end{array}\right) \in M(n, \mathbb{R}) \qquad B = B^t \in M(n-1, \mathbb{R}) \, , \quad c \in \mathbb{R}$$
$$b \in \mathbb{R}^{n-1} \text{ (Spaltenvektor)} \, .$$

Nach Induktionsvoraussetzung folgt aus $\det A_k = \det B_k > 0$ für $1 \leq k \leq n-1$, daß B positiv definit ist. Damit gilt $B = T\Lambda T^t$, wobei $T \in O(n-1)$ und $\Lambda \in M(n-1, \mathbb{R})$ eine Diagonalmatrix mit *positiven Eigenwerten* $\lambda_1, \ldots, \lambda_{n-1}$ von B auf der Diagonale ist. Dann ist

$$\begin{aligned} A &= \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} \Lambda & T^t b \\ \hline b^t T & c \end{array}\right) \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right)^t \\ &= \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} E_{n-1} & 0 \\ \hline b^t T\Lambda^{-1} & 1 \end{array}\right) \left(\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & c - b^t T\Lambda^{-1}T^t b \end{array}\right) \left(\begin{array}{c|c} E_{n-1} & \Lambda^{-1}T^t b \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right)^t . \end{aligned}$$

Nun ist $\det A = (\det T)^2(\det\Lambda)(c - b^t T\Lambda^{-1}T^t b)$ genau dann positiv, wenn $c - b^t T\Lambda^{-1}T^t b > 0$. Andererseits ist $x^t A x > 0$ für alle $x \in \mathbb{R}^n$ genau dann, wenn $y^t \left(\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & c - b^t T\Lambda^{-1}T^t b \end{array}\right) y > 0$ für alle $y \in \mathbb{R}^n$. □

Achtung: In obiger Darstellung für A ist $\left(\begin{array}{c|c} E_{n-1} & \Lambda^{-1}T^t b \\ \hline 0 & 1 \end{array}\right)$ keine orthogonale Matrix für $b \neq 0$, so daß $\lambda_1, \ldots, \lambda_{n+1}, c - b^t T\Lambda^{-1}T^t b$ *nicht* die Eigenwerte von A sind!

Teil VII
Vollständigkeit

24 Topologie metrischer Räume

Es sei erinnert an die Definition 20.13 eines metrischen Raums. Über die Metrik führen wir den zentralen Begriff der *offenen Teilmengen* ein:

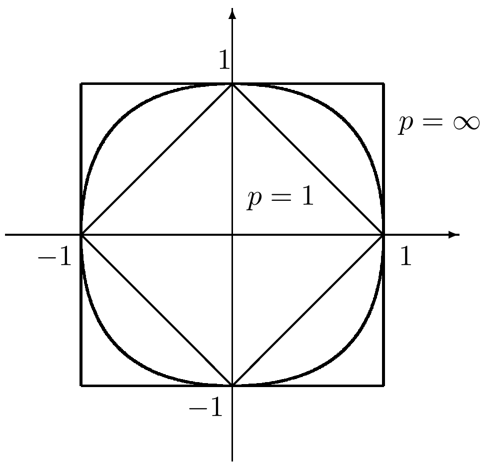
Definition 24.1 Es sei (X, d) ein metrischer Raum, $a \in X$ und $r > 0$. Dann heißt die Teilmenge

$$K_r(a) := \{x \in X \; : \; d(a, x) < r\} \subset X$$

die *offene Kugel* in (X, d) mit Mittelpunkt a und Radius r.

Diese Kugeln können je nach Metrik verschiedene Formen haben:

Beispiel 24.2 Betrachtet werde $X = \mathbb{R}^2$ mit den Metriken $d_p(x, y) = \|x - y\|_p$, die aus den Normen $\| \; \|_p$, $1 \leq p \leq \infty$, aus Beispiel 20.10 induziert werden. Dann haben die Einheitskugeln mit $r = 1$ um $a = 0$ folgende Gestalt:



Definition 24.3 Es sei (X, d) ein metrischer Raum.

i) Eine Teilmenge $U \subset X$ heißt *Umgebung* eines Punktes $a \in X$, wenn es eine offene Kugel $K_\epsilon(a) \subset U$ gibt. Speziell heißt $K_\epsilon(a)$ die *ϵ-Umgebung* von a.

ii) Eine Teilmenge $U \subset X$ heißt *offen*, wenn jeder Punkt $x \in U$ eine in U enthaltene ϵ-Umgebung besitzt, d.h. $\forall x \in U \; \exists \epsilon > 0 \; : \; K_\epsilon(x) \subset U$.
Außerdem wird die leere Menge $\varnothing \subset X$ als offen erklärt.

iii) Die Gesamtheit aller offenen Teilmengen von (X, d) heißt die *von d erzeugte Topologie auf X*.

iv) Eine Teilmenge $A \subset X$ heißt *abgeschlossen*, wenn $X \setminus A$ offen ist.

v) Sei $Y \subset X$ eine Teilmenge. Ein Punkt $y \in X$ heißt *Randpunkt* von Y, wenn es in jeder Umgebung von y sowohl Punkte aus Y als auch aus $X \setminus Y$ gibt. Die Menge aller Randpunkte von Y heißt der *Rand* von Y und wird mit ∂Y bezeichnet.

Damit sind X und $\varnothing$ sowohl offen als auch abgeschlossen in X.

Satz 24.4 *In einem metrischen Raum (X, d) gilt:*

i) *Der Durchschnitt endlich vieler offener Mengen ist offen.*

ii) *Die Vereinigung beliebig vieler offener Mengen ist offen.*

iii) *Der Durchschnitt beliebig vieler abgeschlossener Mengen ist abgeschlossen.*

iv) *Die Vereinigung endlich vieler abgeschlossener Mengen ist abgeschlossen.*

Außerdem gilt (Hausdorffsches Trennungsaxiom):

v) *Zu je zwei Punkten $x, y \in X$ mit $x \neq y$ gibt es disjunkte offene Umgebungen U von x und V von y, d.h. $U \cap V = \varnothing$.*

Satz 24.5 *Sei (X, d) ein metrischer Raum und $Y \subset X$. Dann gilt:*

i) *$Y \setminus \partial Y$ ist offen in X.*

ii) *$Y \cup \partial Y$ ist abgeschlossen in X.*

iii) *∂Y ist abgeschlossen in X.*

Definition 24.6 Für eine Teilmenge $Y \subset X$ eines metrischen Raumes (X, d) wird der *Durchmesser* definiert als $\operatorname{diam}(Y) := \sup_{x,y \in Y} d(x, y)$. Eine Teilmenge $Y \subset X$ heißt *beschränkt*, wenn $\operatorname{diam}(Y) < \infty$.

Beispiel 24.7 Die n-dimensionale Einheitsvollkugel $B^n := \{x \in \mathbb{R}^n \ : \ \|x\|_2 \leq 1\}$ ist abgeschlossen im $\mathbb{R}^n$. Ihr Rand ist die Einheitssphäre $S^{n-1} := \partial B^n = \{x \in \mathbb{R}^n \ : \ \|x\|_2 = 1\}$, sie ist ebenfalls abgeschlossen im $\mathbb{R}^n$. Es ist dann sinnvoll, offene Teilmengen von S^{n-1} zu betrachten. Jeder Durchschnitt der S^{n-1} mit einer offenen Teilmenge des $\mathbb{R}^n$ definiert eine offene Teilmenge von S^{n-1}. Im Sinne dieser Teilmengen ist S^{n-1} dann offen und abgeschlossen zugleich. Es gilt $\operatorname{diam}(B^n) = 2$. $\triangleleft$

Verschiedene Metriken (und Normen) könnten verschiedene Topologien erzeugen. In unendlich-dimensionalen Vektorräumen ist das in der Tat der Fall, im endlich-dimensionalen Fall aber nicht. Wir betrachten hier nur Topologien, die von einer Norm erzeugt werden:

Definition 24.8 Zwei Normen $\|\ \|_a$ und $\|\ \|_b$ auf einem Vektorraum V heißen *äquivalent*, wenn es positive Zahlen $c, C > 0$ gibt, so daß für beliebige $x \in V$ gilt $c\|x\|_a \leq \|x\|_b \leq C\|x\|_a$.

Im Äquivalenzfall enthält jede offene Kugel bezüglich $\| \ \|_a$ eine offene Kugel bezüglich $\| \ \|_b$, und umgekehrt.

Satz 24.9 *Je zwei Normen auf einem n-dimensionalen Vektorraum V über $\mathbb{K}$ sind äquivalent.*

Beweis. Zunächst zeigen wir für $V = \mathbb{R}^n$, daß eine beliebige Norm $\| \ \|$ äquivalent ist zur Standardnorm $\| \ \|_2 = \sqrt{\langle \ , \ \rangle}$. Es sei $(e_1, \dots, e_n)$ die Standardbasis. Dann gilt für $x = \sum_{i=1}^n x_i e_i$ nach Dreiecksungleichung und Cauchy-Schwarz

$$\|x\| \leq \sum_{i=1}^n |x_i| \|e_i\| \leq \sqrt{\sum_{i=1}^n |x_i|^2} \underbrace{\sqrt{\sum_{i=1}^n \|e_i\|^2}}_{C} = C\|x\|_2 \ .$$

Zur Umkehrung betrachten wir die Einheitssphäre $S^{n-1} := \{x \in \mathbb{R}^n \ : \ \|x\| = 1\} \subset \mathbb{R}^n$. Sei $c := \inf_{x \in S^{n-1}} \|x\|$. Wir zeigen: $c > 0$, dann ist $x = \frac{1}{\|y\|_2} y \in S^{n-1}$ für beliebige $y \neq 0$, und damit $c \leq \|x\| = \frac{\|y\|}{\|y\|_2}$. (Für $y = 0$ ist die Äquivalenz klar.) Angenommen, es wäre $c = 0$. Dann gibt es eine Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten $x_k \in S^{n-1}$, so daß die Folge $(\|x_k\|)_{k \in \mathbb{N}}$ reeller Zahlen gegen 0 konvergiert. Wir betrachten die Komponenten x_{ki} von $x_k = (x_{k1}, \dots, x_{kn})$. Die Folge $(x_{k1})_{k \in \mathbb{N}}$ ist beschränkt und enthält nach dem Satz von Bolzano-Weierstraß eine gegen a_1 konvergente Teilfolge $(x_{k_m 1})_{m \in \mathbb{N}}$. Analog konstruiert man eine gegen a_2 konvergente Teilfolge von $(x_{k_m 2})_{m \in \mathbb{N}}$, usw., bis schließlich eine Teilfolge von (x_k) erhalten ist, die komponentenweise gegen $a = (a_1, \dots, a_n)$ konvergiert. Diese konvergente Teilfolge sei wieder mit $(x_k)_{k \in \mathbb{N}}$ bezeichnet. Dann gilt $a_1^2 + \dots + a_n^2 = \lim_{k \to \infty}(x_{k1}^2 + \dots + x_{kn}^2) = 1$, d.h. $a \in S^{n-1}$. Andererseits gilt nach Dreiecksungleichung

$$\|a\| \leq \|a - x_k\| + \|x_k\| \leq C\|a - x_k\|_2 + \|x_k\| \quad \text{für alle } k \in \mathbb{N} \ .$$

Nun ist $\lim_{k \to \infty} \|a - x_k\|_2 = \lim_{k \to \infty} \sqrt{|a_1 - x_{k1}|^2 + \dots |a_n - x_{kn}|^2} = 0$ und $\lim_{k \to \infty} \|x_k\| = 0$, also $\|a\| = 0$ und damit $a = 0$, im Widerspruch zu $a \in S^{n-1}$.

Sei nun V ein beliebiger n-dimensionaler Vektorraum. Dann gibt es eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V. Es sei $\Phi_{\mathcal{B}} : \mathbb{R}^n \to V$ der kanonische Isomorphismus. Damit induzieren die Normen $\| \ \|_a, \| \ \|_b$ auf V entsprechende Normen auf $\mathbb{R}^n$ durch $\|x\|_{a,\mathcal{B}} := \|\Phi_{\mathcal{B}}(x)\|_a$ und $\|x\|_{b,\mathcal{B}} := \|\Phi_{\mathcal{B}}(x)\|_b$. Die Äquivalenz zwischen $\| \ \|_{a,\mathcal{B}}$ und $\| \ \|_{b,\mathcal{B}}$ überträgt sich damit auf $\| \ \|_a, \| \ \|_b$. □

25 Konvergenz in metrischen Räumen. Banach-Räume

Definition 25.1 Sei (X, d) ein metrischer Raum und $(x_k)_{k \in \mathbb{N}}$ eine Folge von Punkten aus X. Die Folge (x_k) heißt *konvergent*, wenn es ein $a \in X$ gibt mit folgenden äquivalenten Eigenschaften:

i) Zu jedem $\epsilon > 0$ gibt es ein $n \in \mathbb{N}$ mit $d(a, x_k) < \epsilon$ für alle $k \geq n$.

ii) Zu jeder Umgebung U von a gibt es ein $n \in \mathbb{N}$ mit $x_k \in U$ für alle $k \geq n$.

iii) Es gilt $\lim_{k\to\infty} d(x_k, a) = 0$.

Im Konvergenzfall heißt $a \in X$ der *Grenzwert* der Folge $(x_k)_{k\in\mathbb{N}}$, und man schreibt $\lim_{k\to\infty} x_k = a$.

Existiert der Grenzwert, dann ist er nach Dreiecksungleichung eindeutig bestimmt: Wären $d(x_k, a_1) < \epsilon$ und $d(x_k, a_2) < \epsilon$ für alle $k \geq n$, so ist auch $d(a_1, a_2) \leq d(a_1, x_k) + d(x_k, a_2) < 2\epsilon$ für beliebige $\epsilon > 0$, und damit $d(a_1, a_2) = 0$ und $a_1 = a_2$.

Die Definition überträgt sich auf Konvergenz und Grenzwert von Folgen in normierten Vektorräumen $(V, \| \ \|)$ mit der durch die Norm induzierten Metrik. Nach Satz 24.9 ist im endlich-dimensionalen Fall der Konvergenzbegriff unabhängig von der Wahl der Norm. Insbesondere konvergiert im $\mathbb{R}^n$ eine Punktfolge $(x_k)_{k\in\mathbb{N}}$ mit $x_k = (x_{k1}, \ldots, x_{kn})$ genau dann gegen $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, wenn für jede Koordinate $1 \leq i \leq n$ die Folge x_{ki} gegen a_i konvergiert.

Satz 25.2 *Sei (X, d) ein metrischer Raum. Eine Teilmenge $Y \subset X$ ist genau dann abgeschlossen in X, wenn für jede in X konvergente Folge $(y_k)_{k\in\mathbb{N}}$ mit $y_k \in Y$ gilt $\lim_{k\to\infty} y_k = x \in Y$.*

Beweis. ($\Rightarrow$) Sei $Y \subset X$ abgeschlossen und $x = \lim_{k\to\infty} y_k$ mit $y_k \in Y$. Angenommen, $x \notin Y$, also $x \in X \setminus Y$. Da $X \setminus Y$ offen ist, gibt es in $X \setminus Y$ eine ϵ-Umgebung U_ϵ um x. Da x der Grenzwert der Folge (y_k) ist, gibt es ein $n \in \mathbb{N}$, so daß $y_k \in U_\epsilon \subset X \setminus Y$ für alle $k \geq n$, Widerspruch.

($\Leftarrow$) Die Grenzwerte aller konvergenten Folgen von Punkten aus Y liegen in Y. Wir zeigen: $X \setminus Y$ ist offen. Angenommen, es gibt einen Punkt $\tilde{x} \in X \setminus Y$, der keine ϵ-Umgebung in $X \setminus Y$ besitzt. Dann gibt es zu jedem $k \in \mathbb{N}$ ein $y_k \in Y$ mit $d(\tilde{x}, y_k) < \epsilon_k := \frac{1}{k+1}$. Auf diese Weise wird eine Folge $(y_k)_{k\in\mathbb{N}}$ von Punkten aus Y konstruiert, die gegen $\tilde{x}$ konvergiert. Nach Voraussetzung ist $\tilde{x} \in Y$, Widerspruch.
□

Definition 25.3 Sei (X, d) ein metrischer Raum. Eine Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus X heißt *Cauchy-Folge*, wenn zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$ existiert, so daß $d(x_n, x_m) < \epsilon$ für alle $m, n \geq k$.

Die Definition überträgt sich auf Cauchy-Folgen in normierten Vektorräumen $(V, \| \ \|)$ mit der durch die Norm induzierten Metrik.

Satz 25.4 *Jede konvergente Folge in einem metrischen Raum ist eine Cauchy-Folge.*

Beweis. Sei $(x_k)_{k\in\mathbb{N}}$ konvergent in (X,d) mit $\lim_{k\to\infty} x_k = x$. Dann gibt es zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$, so daß $d(x_n, x) < \frac{\epsilon}{2}$ und $d(x_m, x) < \frac{\epsilon}{2}$ für alle $m, n \geq k$. Dann folgt aus der Dreiecksungleichung $d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon$. □

Wie in der eindimensionalen Analysis besteht der Sinn der Cauchy-Folge ist, daß man den Grenzwert nicht kennen muß. Das wirft die Frage auf, wann eine Cauchy-Folge auch konvergent ist.

Definition 25.5 Ein metrischer Raum (X,d) heißt *vollständig*, wenn jede Cauchy-Folge in ihm konvergiert.

Ein vollständiger normierter Vektorraum $(V, \| \ \|)$ heißt *Banach-Raum*.

Ein euklidischer oder unitärer Vektorraum $(V, \langle\ ,\ \rangle)$, der bezüglich $\|v\| = \sqrt{\langle v, v\rangle}$ vollständig ist, heißt *Hilbert-Raum*.

Satz 25.6 *Jeder endlich-dimensionale normierte Vektorraum ist vollständig.*

Beweis. Nach Satz 24.9 genügt es in endlich-dimensionalen Vektorräumen, die Vollständigkeit bezüglich einer Norm nachzuweisen. Über den kanonischen Isomorphismus zu einer Basis können wir uns dann auf $\mathbb{K}^n$ beschränken. Sei also $(x_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge bezüglich der $\| \ \|_1$-Norm von Punkten $x_k = (x_{k1}, \dots, x_{kn}) \in \mathbb{K}^n$. Dann gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$, so daß für alle $k, l \geq N$ gilt

$$\|x_k - x_l\|_1 = |x_{k1} - x_{l1}| + \dots + |x_{kn} - x_{ln}| < \epsilon \ .$$

Also ist $(x_{ki})_{k\in\mathbb{N}}$ eine Cauchy-Folge in $\mathbb{K}$ für jedes $i = 1, \dots, n$. Wegen der Vollständigkeit von $\mathbb{K}$ konvergiert $(x_{ki})_{k\in\mathbb{N}}$ gegen ein $a_i \in \mathbb{K}$. Damit konvergiert (x_k) gegen $a = (a_1, \dots, a_n)$. □

Satz 25.7 *Der normierte Vektorraum* $(\mathcal{C}([a,b]), \| \ \|_\infty)$ *der stetigen komplexwertigen Funktionen über dem Intervall* $[a,b] \subset \mathbb{R}$ *zusammen mit der Supremums-Norm* $\|f\|_\infty := \sup_{x\in[a,b]} \|f(x)\|$ *ist ein Banach-Raum.*

Beweis. Sei $(f_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge stetiger Funktionen $f_k \in \mathcal{C}([a,b])$. Dann gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$, so daß für alle $k, l \geq N$ gilt $\|f_k - f_l\|_\infty < \epsilon$. Insbesondere ist $|f_k(x) - f_l(x)| < \epsilon$ für alle $x \in [a,b]$. Damit ist die Folge komplexer Zahlen $(f_k(x))_{k\in\mathbb{N}}$ eine Cauchy-Folge, die gegen ein $f(x) = \lim_{k\to\infty} f_k(x) \in \mathbb{C}$ konvergiert. Dadurch wird punktweise eine Grenzfunktion $f : [a,b] \to \mathbb{C}$ erhalten. Damit gilt $\|f_k - f\|_\infty < \epsilon$ für alle $k \geq N$, d.h. $(f_k)_{k\in\mathbb{N}}$ konvergiert gleichmäßig gegen die Grenzfunktion. Nach Satz 12.12 aus dem letzten Semester ist f stetig. □

Dagegen ist $(\mathcal{C}([a,b]), \| \ \|_p)$ für $1 \leq p < \infty$ *kein* Banachraum. Ähnlich wie bei der Vervollständigung von $\mathbb{Q}$ zu $\mathbb{R}$ lassen sich nicht vollständige normierte Vektorräume vervollständigen. Die Vervollständigung von $(\mathcal{C}([a,b]), \| \ \|_p)$ ist der Vektorraum $L^p([a,b])$ der über $[a,b]$ in p-ter Potenz Lebesgue-integrierbaren Funktionen, die wir im nächsten Semester einführen werden.

Definition 25.8 Es sei $(V, \| \ \|)$ ein normierter Vektorraum über $\mathbb{K}$. Eine lineare Abbildung $f : V \to \mathbb{K}$ heißt *beschränkt*, wenn es ein $0 \leq C < \infty$ gibt, so daß für alle $v \in V$ gilt $|f(v)| \leq C\|v\|$.

Satz 25.9 *Es sei* $(V, \| \ \|)$ *ein normierter Vektorraum über* $\mathbb{K}$. *Die Menge*

$$V' = \mathcal{B}(V,\mathbb{K}) := \{ f : V \to \mathbb{K} \ : \ f \textit{ linear und beschränkt} \}$$

aller linearen beschränkten Abbildungen von V *nach* $\mathbb{K}$ *ist ein Banach-Raum bezüglich der* Operator-Norm

$$\|f\|_{op} := \sup_{v \in V \ , \ \|v\| \leq 1} |f(v)| \ .$$

Dieser Banach-Raum V' *heißt der* Dualraum *von* V.

Beweis. Betrachtet man nur die Linearität, dann ist $V' = \mathrm{Hom}(V,\mathbb{K})$ als Homomorphismus von Vektorräumen selbst wieder ein Vektorraum. Zu zeigen ist jedoch, daß aus $f, g \in V'$ auch $f + g \in V'$ folgt. Das ergibt sich jedoch aus dem Beweis, daß $\| \ \|_{op}$ eine Norm ist. $\|\lambda f\|_{op} = |\lambda| \|f\|_{op}$ ist klar. Die Dreiecksungleichung folgt aus

$$\begin{aligned} \|f+g\|_{op} &= \sup_{v \in V \ , \ \|v\| \leq 1} |f(v) + g(v)| \leq \sup_{v \in V \ , \ \|v\| \leq 1} (|f(v)| + |g(v)|) \\ &\leq \sup_{v \in V \ , \ \|v\| \leq 1} |f(v)| + \sup_{w \in V \ , \ \|w\| \leq 1} |g(v)| = \|f\|_{op} + \|g\|_{op} \ . \end{aligned}$$

Schließlich bedeutet $\|f\|_{op} = 0$, daß $f(v) = 0$ für alle $v \in V$, also ist f die Nullabbildung.

Verbleibt der Beweis der Vollständigkeit. Es sei $(f_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in V'. Dann folgt für alle $v \in V \setminus \{0\}$ (für $v = 0$ ist nichts zu zeigen)

$$\begin{aligned} |f_n(v) - f_m(v)| &= |(f_n - f_m)(v)| = \Big| \|v\| (f_n - f_m)\Big(\frac{1}{\|v\|} v\Big) \Big| \\ &= \|v\| \Big| (f_n - f_m)\Big(\frac{v}{\|v\|}\Big) \Big| \leq \|f_n - f_m\|_{op} \|v\| \overset{n,m \to \infty}{\to} 0 \ . \end{aligned}$$

Also ist auch $(f_n(v))_{n \in \mathbb{N}}$ eine Cauchy-Folge in $\mathbb{K}$, die wegen der Vollständigkeit konvergiert gegen ein $f(v) := \lim_{n \to \infty} f_n(v) \in \mathbb{K}$. Weiter gilt wegen der Linearität der f_n

$$f(\lambda v + \mu w) = \lim_{n \to \infty} f_n(\lambda v + \mu w) = \lambda \lim_{n \to \infty} f_n(v) + \mu \lim_{n \to \infty} f_n(w) = \lambda f(v) + \mu f(w) \ ,$$

d.h. f ist linear.

Zu jedem $\epsilon > 0$ existiert ein $N \in \mathbb{N}$, so daß $\|f_n - f_m\|_{op} < \epsilon$ für alle $m, n \geq N$. Dann ist für alle $v \in V$ mit $\|v\| \leq 1$ und $n \geq N$

$$|f(v) - f_n(v)| = \lim_{m\to\infty} |f_m(v) - f_n(v)| \leq \sup_{v\in V\,,\,\|v\|\leq 1} \{|f_m(v) - f_n(v)| \;:\; m \geq N\} < \epsilon\,.$$

Das ergibt $\|f - f_n\|_{op} := \sup_{v\in V\,,\,\|v\|\leq 1} |f(v) - f_n(v)| < \epsilon$, d.h. f_n konvergiert in der Operatornorm gegen f. Insbesondere ist $f - f_n$ beschränkt und somit auch $f = (f - f_n) + f_n$. □

Satz 25.10 *Sei $(H, \langle\ ,\ \rangle)$ ein Hilbert-Raum und $U \subset H$ ein abgeschlossener Untervektorraum. Dann gilt:*

i) *Zu jedem $v \in H$ existiert genau ein $w \in U$ mit $\|v - w\| = \min_{u\in U} \|v - u\| =: d(v, U)$.*

ii) *Es gilt $v - w \in U^\perp$, so daß $P_U(v) := w$ die orthogonale Projektion von v auf U definiert.*

iii) *$H = U \oplus U^\perp$.*

Beweis. i) Sei $\gamma := \inf_{u\in U} \|v - u\|$. Dann gibt es eine Folge $(u_n)_{n\in\mathbb{N}}$ von Vektoren $u_n \in U$, so daß die Folge $(\gamma_n)_{n\in\mathbb{N}}$ mit $\gamma_n := \|v - u_n\|$ gegen γ konvergiert. Wir zeigen mit der Parallelogrammgleichung, daß $(u_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist. Wegen der Vollständigkeit konvergiert diese gegen ein $w \in H$, und wegen der Abgeschlossenheit von U gilt sogar $w \in U$.

$$\begin{aligned}\|u_m - u_n\|^2 &= \|(v - u_n) - (v - u_m)\|^2 \\ &= 2\|v - u_n\|^2 + 2\|v - u_m\|^2 - 4\|v - \tfrac{1}{2}(u_n + u_m)\|^2\,.\end{aligned}$$

Wegen $\frac{1}{2}(u_n + u_m) \in U$ ist $\gamma \leq \|v - \frac{1}{2}(u_n + u_m)\| \leq \frac{1}{2}(\|v - u_n\| + \|v - u_m\|)$. Da $\|v - u_n\|$ und $\|v - u_m\|$ gegen γ konvergieren, konvergiert auch $\|v - \frac{1}{2}(u_n + u_m)\|^2$ gegen γ und $\|u_m - u_n\|^2$ gegen 0 für $m, n \to \infty$.

Wir zeigen anschließend $v - w \in U^\perp$, dann folgt die Eindeutigkeit von w wie in Satz 21.4.

ii) Sei also $0 \neq u \in U$ beliebig und $\lambda := \frac{\langle v-w,u\rangle}{\|u\|^2} \in \mathbb{K}$. Dann bedeutet die Minimalität $\|v - w\|^2 \leq \|v - w - \lambda u\|^2$. Andererseits ist

$$\begin{aligned}\|v - w - \lambda u\|^2 &= \|v - w\|^2 + |\lambda|^2\|u\|^2 - \lambda\langle v - w, u\rangle - \overline{\lambda}\langle u, v - w\rangle \\ &= \|v - w\|^2 - |\lambda|^2\|u\|^2\,,\end{aligned}$$

also $\lambda = 0$ und damit $v - w \in U^\perp$.

iii) Somit ist $v = (v - P_U(v)) + P_U(v) \in U^\perp + U$, also $H = U + U^\perp$ und $U \cap U^\perp = \{0\}$ nach Satz 21.9. □

Satz 25.11 (Riesz) *Sei H ein Hilbert-Raum, $(H', \|\ \|_{op})$ sein Dualraum und $j : H \to H'$ definiert durch $j(v)(w) := \langle v, w\rangle$. Dann gilt:*

i) *j ist antilinear, d.h. $j(\lambda v + \mu w) = \overline{\lambda} j(v) + \overline{\mu} j(w)$.*

ii) *j ist isometrisch, also injektiv.*

iii) *j ist surjektiv, also ein Anti-Isomorphismus. (gilt nicht für Prä-Hilbert-Räume!)*

Beweis. i) folgt aus der Definition.

ii) Zu zeigen ist $\|j(v)\|_{op} = \|v\|$ für alle $v \in H$. Das ist klar für $v = 0$. Ansonsten ($v \neq 0$) betrachten wir

$$\|j(v)\|_{op} = \sup_{w \in H\,,\ \|w\| \leq 1} |j(v)(w)| = \sup_{w \in H\,,\ \|w\| \leq 1} |\langle v, w\rangle| \leq \|v\|\ ,$$

also $\|j(v)\|_{op} \leq \|v\|$. Andererseits ist für die spezielle Wahl $j(v)\big(\frac{v}{\|v\|}\big) = \|v\|$ und damit $\|j(v)\|_{op} \geq \|v\|$.

iii) Sei $0 \neq f \in H'$ und sei $U := \ker f \subset H$. Wir zeigen, daß der Untervektorraum $U \subset H$ abgeschlossen bezüglich der durch $\| \; \|$ induzierten Topologie ist. Sei $(v_n)_{n\in\mathbb{N}}$ eine in H konvergente Folge von Vektoren $v_n \in U$ mit $v := \lim_{n\to\infty} v_n \in H$, dann ist $f(v_n) = 0$ für alle $n \in \mathbb{N}$. Da f beschränkt ist, folgt mit

$$|f(v)| \leq |f(v) - f(v_n)| + |f(v_n)| = |f(v - v_n)| \leq \|f\|_{op}\|v - v_n\| \to 0 \text{ für } n \to \infty$$

daß $f(v) = 0$ gilt. Damit ist $v \in \ker f$, und U ist abgeschlossen.

Da $f \neq 0$, gibt es Vektoren $0 \neq v \in H$ mit $f(v) \neq 0$, insbesondere ist $U \neq H$. Wegen $H = U \oplus U^\perp$ gibt es $0 \neq v \in U^\perp$ mit $\|v\| = 1$. Sei $\lambda := f(v)$. Dann gilt für alle $w \in H$

$$f\big(f(w)v - f(v)w\big) = f(w)f(v) - f(v)f(w) = 0\ ,$$

also $f(w)v - f(v)w \in U$ und weiter

$$0 = \langle \underbrace{v}_{\in U^\perp}, \underbrace{f(w)v - f(v)w}_{\in U}\rangle = f(w)\|v\|^2 - f(v)\langle v, w\rangle\ ,$$

also $f(w) = \lambda\langle v, w\rangle = \langle \overline{\lambda} v, w\rangle$ und damit $f = j(\overline{\lambda} v)$. □

Insbesondere gilt in H' die Parallelogrammgleichung, so daß H' mit dem Skalarprodukt aus der Polarisationsformel selbst ein Hilbert-Raum ist.

Wir nutzen den Rieszschen Darstellungssatz, um die Existenz des *adjungierten Operators* zu beweisen.

Definition 25.12 Es sei H ein Hilbert-Raum. Ein linearer beschränkter Operator A auf H ist eine lineare Abbildung $A : H \to H$ mit $\|A\|_{op} := \sup_{v\in H\,,\ \|v\|\leq 1} \|Av\| < \infty$. Mit $\mathcal{B}(H)$ wird die Menge aller linearen beschränkten Operatoren auf H bezeichnet.

Analog zu Satz 25.9 beweist man, daß $\mathcal{B}(H)$ ein Banach-Raum ist.

Satz 25.13 *Es sei H ein Hilbert-Raum und $A \in \mathcal{B}(H)$ ein linearer beschränkter Operator. Dann existiert genau ein linearer beschränkter Operator $A^* \in \mathcal{B}(H)$, der zu A adjungierte Operator, mit $\langle A^*v, w\rangle = \langle v, Aw\rangle$ für alle $v, w \in H$. Es gilt $\|A^*\|_{op} = \|A\|_{op}$.*

Beweis. Für jedes $v \in H$ ist das lineare Funktional $f_v : H \to \mathbb{K}$ mit $f_v(w) := \langle v, Aw\rangle$ beschränkt wegen

$$|f_v(w)| = |\langle v, Aw\rangle| \leq \|v\|\|Aw\| \leq \|A\|_{op}\|v\|\|w\| \;.$$

Nach dem Rieszschen Darstellungssatz existiert genau ein Vektor $A^*v \in H$ mit $f_v = j(A^*v)$, d.h. $f_v(w) = \langle v, Aw\rangle = \langle A^*v, w\rangle$. Die so konstruierte Abbildung $A^* : H \to H$ ist linear:

$$\begin{aligned}\langle A^*(\lambda_1 v_1 + \lambda_2 v_2), w\rangle &= \langle(\lambda_1 v_1 + \lambda_2 v_2), Aw\rangle = \overline{\lambda_1}\langle v_1, Aw\rangle + \overline{\lambda_2}\langle v_2, Aw\rangle \\ &= \overline{\lambda_1}\langle A^*(v_1), w\rangle + \overline{\lambda_2}\langle A^*(v_2), w\rangle = \langle \lambda_1 A^*(v_1) + \lambda_2 A^*(v_2), w\rangle,\end{aligned}$$

also $\langle A^*(\lambda_1 v_1 + \lambda_2 v_2) - (\lambda_1 A^*(v_1) + \lambda_2 A^*(v_2)), w\rangle = 0$ für alle $w \in H$, und daraus $A^*(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 A^*(v_1) + \lambda_2 A^*(v_2)$. Für die Operatornormen sei $v \in H$ mit $\|v\| \leq 1$ gegeben, dann gilt mit Cauchy-Schwarz

$$\|A^*v\|^2 = \langle A^*v, A^*v\rangle = \langle v, AA^*v\rangle \leq \|v\|\|AA^*v\| \leq \|A\|_{op}\|A^*v\| \;,$$

also $\|A^*\|_{op} \leq \|A\|_{op}$. Durch analoge Abschätzung von $\|Aw\|$ beweist man die Umkehrung $\|A^*\|_{op} \leq \|A\|_{op}$. Daraus folgt $A^* \in \mathcal{B}(H)$. □

Somit lassen sich in Hilbert-Räumen selbstadjungierte, unitäre und normale Operatoren definieren und untersuchen. Insbesondere gibt es Analogien zu den Eigenwertproblemen im endlich-dimensionalen Fall.