

# Lineare Algebra

Hans-Peter Beise

Sommersemester 2023

Literatur:

*Strang, Gilbert. Linear Algebra For Everyone. Wellesley-Cambridge*

*Jänich, Klaus. Lineare Algebra. Springer*

*Axler, Sheldon. Linear Algebra Done Right. Springer*

## Contents

1	Der Vektorraum $\mathbb{R}^n$	3
2	Matrizen	7
3	Das Skalarprodukt	18
4	Vektorräume und Lineare Abbildungen	26
5	Inverse Matrix, lineare Gleichungssysteme	38
6	Zerlegung von Vektorräumen und Orthogonalität	51
7	Kleinste-Quadrate-Problem und Lineare Regression	63
8	Erweiterung auf komplexe Vektorräume	66
9	Determinante	70
10	Eigenwerte	77
A	Zusammenfassung Matrix Faktorisierung	90
A.1	LR-Zerlegung (LU-Decomposition) . . . . .	90
A.2	QR-Zerlegung . . . . .	90
A.3	Singulärwertzerlegung . . . . .	91
B	Vektorräume	93
C	Mengen und Abbildungen	98
D	Komplexe Zahlen	108

# 1 Der Vektorraum $\mathbb{R}^n$

In diesem Abschnitt wird eine Einführung in grundlegenden Eigenschaften des (Vektorraums)  $\mathbb{R}^n$  gegeben.

Wir betrachten folgende Mengen:

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\},$$

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1, x_2, x_3 \in \mathbb{R} \right\}$$

oder allgemein die Menge der  $n$ -Tupel ( $n \in \mathbb{N}$ )

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_k \in \mathbb{R}, k = 1, \dots, n \right\}$$

Die Elemente dieser Mengen heißen **Vektoren**. Typischerweise haben Vektoren in dieser Veranstaltung die Variablennamen  $u, v, w, x, y, z$ . An vielen Stellen wird aber auch die **Komponentenschreibweise** benötigt:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Hierbei heißen die  $x_1, \dots, x_n$  die **Komponenten** von  $x$ . An einigen Stellen wird auch eine endliche Folge von Vektoren  $v_1, \dots, v_m \in \mathbb{R}^n$  betrachtet werden. Hier werden die Indizes zur Unterscheidung der verschiedenen Vektoren, und nicht zur Benennung einzelner Komponenten eines Vektors verwendet. Die Bedeutung der Indizes sollte stets aus dem Zusammenhang erkennbar sein. Skalare (Elemente aus  $\mathbb{R}$ ) werden meist durch kleine griechische Buchstaben  $\alpha, \beta, \lambda, \mu \dots$  bezeichnet.

Der **Nullvektor** ist definiert durch:

$$0 := \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Somit wird das Symbol 0 gleichzeitig für die skalare Null und den Nullvektor verwendet. Die jeweilige Bedeutung erschließt sich stets aus dem Zusammenhang. Die elementaren Operationen (Addition, Skalarmultiplikation) sind wie folgt definiert.

**Definition 1.1** Es seien  $x, y \in \mathbb{R}^n$  und  $\lambda \in \mathbb{R}$ , dann ist die **Addition** definiert durch

$$x + y := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

und die **skalare Multiplikation** definiert durch

$$\lambda x := \lambda \cdot x := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Der Kürze halber schreiben wir auch:

$$(-1)x = -x = \begin{pmatrix} -x_1 \\ \vdots \\ -x_n \end{pmatrix}$$

und

$$x + (-y) = x - y.$$

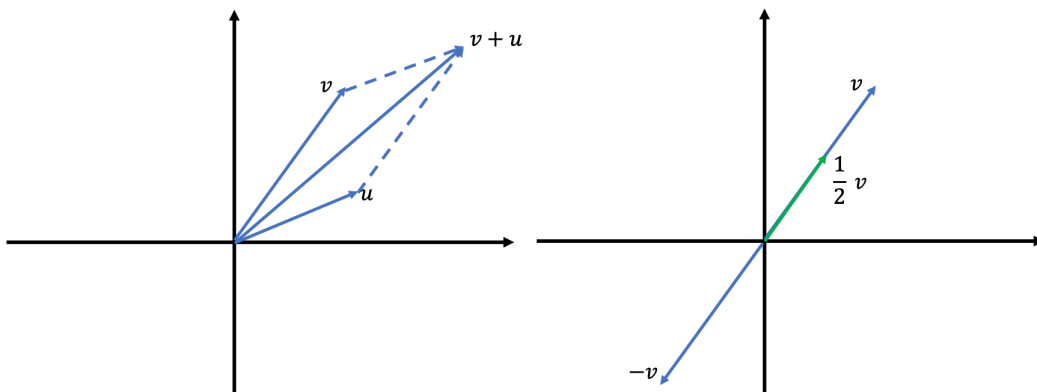


Figure 1: (Vektor-) Addition und skalare Multiplikation im  $\mathbb{R}^2$

Der folgende Satz fasst die grundlegenden Rechenregeln der Addition und skalaren Multiplikation aus Definition 1.1 zusammen.

**Satz 1.2** *Es seien  $v, w, z \in \mathbb{R}^n$  und  $\lambda, \mu \in \mathbb{R}$ . Dann gelten:*

1. Für alle  $(v + w) + z = v + (w + z)$  (Assoziativgesetz),
2.  $v + w = w + v$  (Kommutativgesetz),
3.  $\lambda(\mu v) = (\lambda\mu)v$ ,
4.  $\lambda(v + w) = \lambda v + \lambda w$ ,
5.  $(\lambda + \mu)v = \lambda v + \mu v$ .

**Beweis.** Der Beweis folgt unmittelbar durch komponentenweises Anwenden des Assoziativ- und Kommutativgesetzes der reellen Zahlen.  $\square$

Aus den oben eingeführten Operationen ergibt sich der für die lineare Algebra sehr zentrale Begriff der Linearkombination.

**Definition 1.3** Es seien  $v_1, \dots, v_m \in \mathbb{R}^n$  und  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  ( $m \in \mathbb{N}$ ). Dann heißt eine Summe der Form

$$\sum_{j=1}^m \lambda_j v_j = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_m v_m$$

**Linearkombination** der  $v_1, \dots, v_m$ .

**Beispiel 1.4** Es seien  $\lambda_1 = 2$ ,  $\lambda_2 = 3$ , und

$$v_1 = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}.$$

Dann gilt

$$\lambda_1 v_1 + \lambda_2 v_2 = \begin{pmatrix} 1 \\ 12 \\ -2 \end{pmatrix}.$$

**Definition 1.5** Es seien  $v_1, \dots, v_m \in \mathbb{R}^n$ . Dann definieren wir

$$\text{Span}(v_1, \dots, v_m) := \left\{ v \in \mathbb{R}^n : v = \sum_{j=1}^m \lambda_j v_j \text{ mit } \lambda_1, \dots, \lambda_m \in \mathbb{R} \right\},$$

und nennen  $\text{Span}(v_1, \dots, v_m)$  den von  $v_1, \dots, v_m$  **aufgespannten Raum** oder auch die **lineare Hülle** der  $v_1, \dots, v_m$ .

Im weiteren Verlauf der Vorlesung werden Mengen dieses Typs an vielen Stellen eine zentrale Rolle spielen. Mengen dieses Typs werden später allgemein als Untervektorräume verstanden.

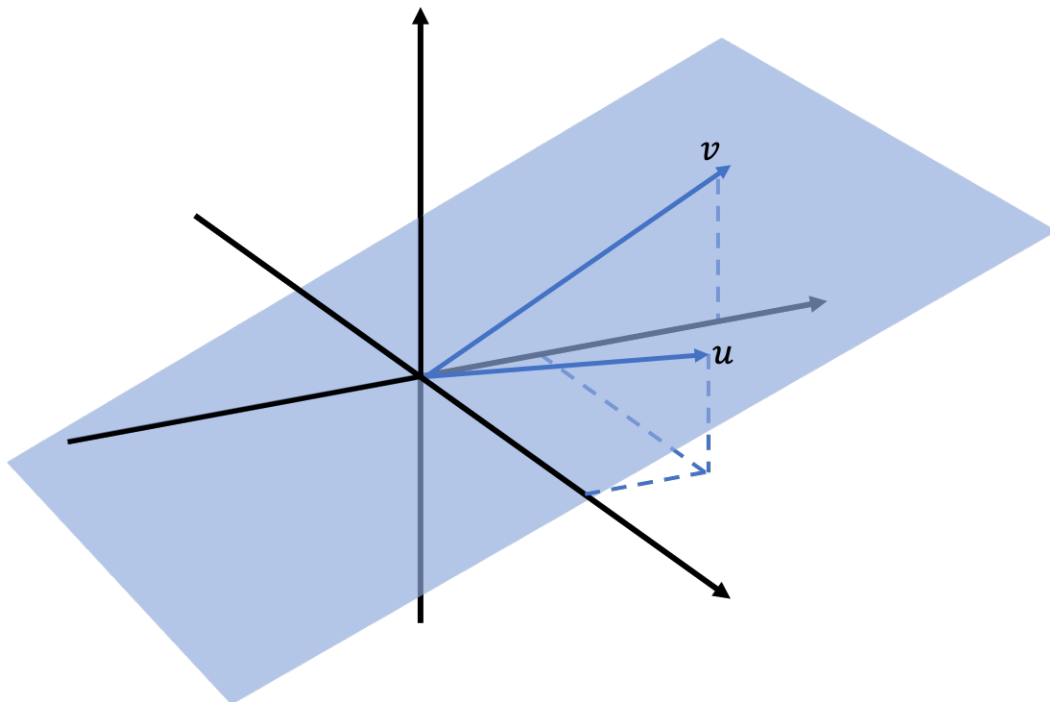


Figure 2:  $\text{Span}(v, u)$

## 2 Matrizen

Neben Vektoren sind Matrizen die zentralen mathematischen Objekte in der (anwendungsorientierten) linearen Algebra.

Matrizen und Vektoren sind grundlegende Bausteine in nahezu jeder Anwendung, die mehrdimensionale mathematische Objekte verwendet:

- Computer-Grafik
- Maschinelles Lernen
- Medizinische Bildgebung
- Statistik - Big Data
- Signalverarbeitung
- Simulation physikalischer, chemischer oder biologischer Vorgänge
- ...

**Definition 2.1** Es seien  $a_{j,k} \in \mathbb{R}$  für  $j = 1, \dots, m$ ,  $k = 1, \dots, n$ , dann heißt das rechteckige Schema

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

eine  $m \times n$  **Matrix** und wir bezeichnen die Menge aller  $m \times n$  **Matrizen** mit  $\mathbb{R}^{m \times n}$ . Die Elemente in  $\mathbb{R}^{n \times n}$  heißen **quadratische** Matrizen

**Bemerkung 2.2** Für eine Matrix  $A \in \mathbb{R}^{m \times n}$  wie in Definition 2.1 schreiben wir auch

$$[A]_{jk} = a_{jk}$$

$j = 1, \dots, m$ ,  $k = 1, \dots, n$  für einzelne Einträge und für Vektoren  $x \in \mathbb{R}^n$

$$[x]_k = x_k$$

$k = 1, \dots, n$ , wobei  $x_k$  die  $k$ -te Komponente des Vektors bezeichnet.

**Definition 2.3** Es seien  $A \in \mathbb{R}^{m \times n}$  wie in Definition 2.1, dann heißen

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

die **Spaltenvektoren** von  $A$ , und

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \end{pmatrix}, \dots, \begin{pmatrix} a_{m1} \\ \vdots \\ a_{mn} \end{pmatrix}$$

die **Zeilenvektoren** von  $A$ . Sind die Spalten Vektoren von  $A$  geben durch  $a_1, \dots, a_n \in \mathbb{R}^m$  so schreiben wir auch

$$A = (a_1 \dots a_n).$$

Es ist oftmals hilfreich näher Informationen zur Struktur einer Matrix zu haben. Folgendes zeigt typische Fälle, die dabei auftreten können: Eine Matrix  $D \in \mathbb{R}^{n \times n}$  der Form

$$D = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}. \quad (2.1)$$

heißt **Diagonalmatrix**. Wir schreiben auch  $D := \text{diag}(a_{11}, \dots, a_{nn})$ .

Eine Matrix  $U \in \mathbb{R}^{n \times n}$  der Form

$$U = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \ddots & a_{2n} \\ 0 & 0 & a_{33} & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}. \quad (2.2)$$

heißt **obere Dreiecksmatrix**,



und eine Matrix  $L \in \mathbb{R}^{n \times n}$  der Form

$$L = \begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \ddots & 0 \\ a_{31} & a_{32} & a_{33} & \ddots & \dots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}. \quad (2.3)$$

heißt **unter Dreiecksmatrix**.

**Definition 2.4** Die Summe zweier  $m \times n$  Matrizen ist definiert als

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ b_{21} & \dots & b_{2n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & \dots & a_{2n} + b_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix},$$

und das Produkt eines Skalars  $\lambda \in \mathbb{R}$  mit einer  $m \times n$  Matrix ist definiert als

$$\lambda \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \dots & \lambda a_{2n} \\ \vdots & \ddots & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{pmatrix}.$$

Nach unserer bisherigen Konvention sind  $x \in \mathbb{R}^n$  Spaltenvektoren

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Wir könnten hier aber auch schreiben  $x \in \mathbb{R}^{n \times 1}$ , und könnten so anhand der Definition zu den Zeilenvektoren  $x = (x_1, \dots, x_n)$  unterscheiden, welche dann Vektoren des  $\mathbb{R}^{1 \times n}$  sind.

Die Unterscheidung zwischen Spalten- und Zeilenvektor ist bei der folgenden Definition des Matrix-Vektor-Produkts wichtig.

**Definition 2.5** Für

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

ist das **Matrix-Vektor-Produkt**  $A \cdot x = Ax$  wie folgt definiert

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^m (\in \mathbb{R}^{m \times 1}).$$

**Beispiel 2.6** Es gilt

$$\begin{pmatrix} 1 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}.$$

**Beispiel 2.7** Wichtige Beispiele zur Anwendung des Matrix-Vektor-Produkts im  $\mathbb{R}^2$  und  $\mathbb{R}^3$  (zum Beispiel in der Computergraphik) sind Drehungen. Sei dazu

$$D_\alpha := \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \in \mathbb{R}^{2 \times 2},$$

dann bewirkt das Matrix-Vektor-Produkt

$$D_\alpha x$$

geometrisch eine Drehung des Vektors  $x$  am Ursprung um den Winkel  $\alpha$  (Bogenmaß  $[0, 2\pi)$ ) gegen den Uhrzeigersinn (mathematisch positive Richtung). Dabei bleibt die Länge erhalten.

Wesentlicher Gegenstand dieser Vorlesung ist die Untersuchung der Abbildungen, die sich durch das Matrix-Vektor-Produkt ergeben: Für  $A \in \mathbb{R}^{m \times n}$  ist durch

$$\boxed{x \mapsto Ax}$$

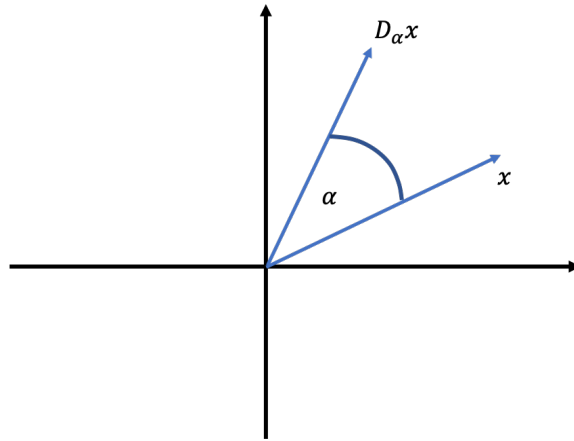


Figure 3: Drehung eines Vektors  $x \in \mathbb{R}^2$  um den Winkel  $\alpha$

eine Abbildung von  $\mathbb{R}^n$  nach  $\mathbb{R}^m$  definiert.

**Satz 2.8** (*Linearität des Matrix-Vektor-Produkts*) Es sei  $A \in \mathbb{R}^{m \times n}$ ,  $x, y \in \mathbb{R}^n$  und  $\lambda \in \mathbb{R}$ , dann gilt:

1.  $A(x + y) = Ax + Ay$
2.  $A\lambda x = \lambda Ax$ .

Diese wichtige Eigenschaft des Matrix-Vektor-Produkts wird uns später in einem allgemeineren Rahmen auf den Begriff der linearen Abbildung führen.

**Beweis.** Wir nutzen die in (2.2) eingeführte Notation. Zu 1. Für  $k \in \{1, \dots, m\}$  gilt

$$[A(x + y)]_k = \sum_{j=1}^n [A]_{kj}([x]_j + [y]_j) = \sum_{j=1}^n [A]_{kj}[x]_j + \sum_{j=1}^n [A]_{kj}[y]_j = [Ax]_k + [Ay]_k.$$

Da  $k$  beliebig war folgt  $A(x + y) = Ax + Ay$ .

Teil 2. kann als Übung gezeigt werden. □

Das iterative Anwenden der Eigenschaft aus Satz 2.8 führt sofort zu Folgendem:

**Satz 2.9** Es sei  $A \in \mathbb{R}^{m \times n}$ ,  $v_1, \dots, v_k \in \mathbb{R}^n$  und  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ , dann gilt:

$$A \sum_{j=1}^k \lambda_j v_j = \sum_{j=1}^k \lambda_j A v_j.$$

Ein weitere Anwendung des Matrix-Vektor-Produkts ist eine elegante Schreibweise linearer Gleichungssysteme. Dazu das folgende Beispiel:

**Beispiel** Das lineare Gleichungssystem

$$\begin{aligned} 4x_1 + 5x_2 &= 10 \\ 8x_1 + 9x_2 &= -2 \end{aligned}$$

lässt sich schreiben als

$$\begin{pmatrix} 4 & 5 \\ 8 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10 \\ -2 \end{pmatrix}$$

Allgemein ist ein lineares Gleichungssystem (LGS) geben durch

$$Ax = b,$$

wobei  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  und eine Lösung  $x \in \mathbb{R}^n$  gesucht ist. Ausgeschrieben heißt das

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = b,$$

wobei  $a_{jk} = [A]_{jk}$ ,  $x_k = [x]_k$ ,  $b_j = [b]_j$ . Das Lösen, die Lösbarkeit und die Lösungsmenge solcher LGS ist ein wichtiges Anwendungsgebiet der linearen Algebra. Dieses Thema wird im späteren Teil der Vorlesung wieder aufgegriffen.

Eine nützliche Interpretation des Matrix-Vektor-Produkts liefert die folgende Betrachtung

**Bemerkung 2.10** Es seien  $A \in \mathbb{R}^{m \times n}$  und  $A_1, \dots, A_n \in \mathbb{R}^m$  die Spaltenvektoren von  $A$ , und

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n,$$

dann ist

$$Ax = \sum_{j=1}^n x_j A_j = x_1 A_1 + x_2 A_2 + \dots + x_n A_n,$$

und somit eine Linearkombination der Spaltenvektoren der Matrix  $A$ . Damit folgt sofort mit Definition 1.5

$$Ax \in \text{Span}(A_1, \dots, A_n).$$

**Beispiel 2.11** Als Fortsetzung von Beispiel 2.7 betrachten wir nun die  $3 \times 3$  Drehmatrizen:

$$D_{1,\alpha} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} \in \mathbb{R}^{3 \times 3},$$

$$D_{2,\alpha} := \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix} \in \mathbb{R}^{3 \times 3},$$

$$D_{3,\alpha} := \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 3}.$$

Dann bewirkt die zugehörige Abbildung

$$x \mapsto D_{k,\alpha} x,$$

mit  $k = 1, 2, 3$ , geometrisch eine Drehung des Vektors  $x$  um die  $k$ -te Koordinatenachse mit Winkel  $\alpha$  (Bogenmaß  $[0, 2\pi)$ ) unter Erhalt der Länge des Vektors.

Es stellt sich nun die Frage, ob man die Matrizen aus Beispiel 2.11 zur Berechnung allgemeiner Drehungen um den Ursprung verwenden kann. Die Antwort dazu liefert das Matrix-Matrix-Produkt, welches im Folgenden eingeführt wird.

**Definition 2.12** Für

$$B = \begin{pmatrix} b_{11} & \dots & b_{1m} \\ b_{21} & \dots & b_{2m} \\ \vdots & \ddots & \vdots \\ b_{k1} & \dots & b_{km} \end{pmatrix} \in \mathbb{R}^{k \times m}, \quad A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

ist das **Matrix-Matrix-Produkt**

$$B \cdot A = B A = C = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ c_{21} & \dots & c_{2n} \\ \vdots & \ddots & \vdots \\ c_{k1} & \dots & c_{kn} \end{pmatrix} \in \mathbb{R}^{k \times n}$$

definiert durch

$$c_{j,i} := \sum_{l=1}^m b_{j,l} a_{l,i}, \quad j = 1, \dots, k \text{ und } i = 1, \dots, n.$$

Für quadratische Matrizen  $A \in \mathbb{R}^{n \times n}$  und  $k \in \mathbb{N}$  ist weiter

$$A^k := \underbrace{A \cdot A \cdot \dots \cdot A}_{k\text{-mal}}$$

und

$$A^0 = E$$

wobei  $E$  die Einheitsmatrix ist, vgl. (2.4).

**Bemerkung 2.13** Es sei  $A \in \mathbb{R}^{m \times n}$  und  $x \in \mathbb{R}^n$ , dann ist das Matrix-Vektor-Produkt  $Ax = AX$ , wobei  $X \in \mathbb{R}^{n \times 1}$  identisch mit  $x$  in der Interpretation als  $\mathbb{R}^{n \times 1}$  Matrix ist.

Ein zentrale quadratische Matrix ist gegeben durch

$$E = E_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}, \quad (2.4)$$

diese wird **Einheitsmatrix** genannt (dabei wird die Notation  $E_n$  anstatt  $E$  verwendet, falls Uneindeutigkeit der Dimension des zugehörigen Raums besteht). Die Einheitsmatrix ist das neutrale Element bezüglich des Matrix-Matrix-Produkts, denn es gilt, wie mal leicht nachprüft, für  $A \in \mathbb{R}^{m \times n}$ :

$$AE_n = A$$

und

$$E_m A = A.$$

Das folgende Ergebnis fasst wichtige Rechenregeln der Matrixrechnung zusammen. Mit Bemerkung 2.13 erkennt man, dass dieses Resultat die Inhalte aus Satz 2.8 verallgemeinert.

**Satz 2.14** *Es gelten die folgenden Rechenregeln für Matrizen  $A, B, C$ :*

1.  $A + B = B + A$
2.  $(A + B)C = AC + BC$
3.  $A(B + C) = AB + AC$
4.  $A(BC) = (AB)C$

*vorausgesetzt, dass  $A, B, C$  jeweils passende Dimension haben (vgl. Übung)*

**Beweis.** Wir zeigen nur (2.), die Übrigen dienen als Übung. Seien  $A, B \in \mathbb{R}^{m \times n}$ ,  $C \in \mathbb{R}^{n \times k}$ . Mit der Schreibweise aus (2.2) gilt für  $\nu \in \{1, \dots, m\}$ ,  $l \in \{1, \dots, k\}$ :

$$[(A+B)C]_{\nu,l} = \sum_{j=1}^n ([A]_{\nu,j} + [B]_{\nu,j})[C]_{j,l} = \sum_{j=1}^n [A]_{\nu,j}[C]_{j,l} + \sum_{j=1}^n [B]_{\nu,j}[C]_{j,l} = [AC]_{\nu,l} + [BC]_{\nu,l}.$$

Da  $\nu, l$  beliebig waren folgt die Behauptung.  $\square$

Eine erste Anwendung des Matrix-Matrix-Produkts liefert das folgende Beispiel.

**Beispiel** Es seien  $D_{1,\alpha}$ ,  $D_{2,\beta}$  aus Beispiel 2.11,  $x \in \mathbb{R}^3$  und  $A = D_{2,\beta}D_{1,\alpha}$ , dann bewirkt

$$Ax = D_{2,\beta}D_{1,\alpha}x$$

eine Drehung um den Winkel  $\alpha$  um die erste Koordinatenachse und eine Drehung um den Winkel  $\beta$  um die zweite Koordinatenachse. Durch das Matrix-Matrix-Produkt werden also die beiden Drehungen von  $D_{1,\alpha}$ ,  $D_{2,\beta}$  in einer neuen Matrix  $A$  zusammengefasst.

Am Beispiel der Drehmatrizen könnte man leicht zu der Auffassung gelangen, dass die Reihenfolge der Matrizen im Matrix-Matrix-Produkt quadratischer Matrizen unerheblich ist. Dies ist jedoch nicht der Fall:

**Bemerkung 2.15** Das Matrix-Matrix-Produkt ist nicht kommutativ, d.h. für zwei Matrizen  $A, B \in \mathbb{R}^{n \times n}$  ist allgemein

$$AB \neq BA.$$

Zum Ende dieses Abschnitts führen wir noch eine grundlegende Operation auf Matrizen ein, die den Übergang zwischen Spalten- und Zeilenvektoren auf Matrizen verallgemeinert.

**Definition 2.16** Es sei  $A \in \mathbb{R}^{m \times n}$  bestehend aus den Einträgen  $a_{jk}$ ,  $j = 1, \dots, m$  und  $k = 1, \dots, n$ . Dann ist die Matrix  $B \in \mathbb{R}^{n \times m}$  bestehend aus den Einträgen

$$b_{jk} := a_{kj} \text{ für } j = 1, \dots, n \text{ und } k = 1, \dots, m$$

oder gleichbedeutend

$$[B]_{jk} = [A]_{kj} \text{ für } j = 1, \dots, n \text{ und } k = 1, \dots, m,$$

die **Transponierte** oder die **transponierte Matrix** von  $A$ , und wir schreiben  $A^T := B$ .

Die Transponierte ergibt sich also wie folgt:



$$A = \begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & & \\ a_{m1} & a_{m2} \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}, \quad A^T = \begin{pmatrix} a_{11} & a_{21} \dots & a_{m1} \\ a_{12} & a_{22} \dots & a_{m2} \\ \vdots & & \\ a_{1n} & a_{2n} \dots & a_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}.$$

**Definition 2.17** Es sei  $A \in \mathbb{R}^{n \times n}$ , dann heißt  $A$  **symmetrisch**, wenn  $A = A^T$  gilt.

Viele quadratische Matrizen in Anwendungen sind symmetrisch:

- Kovarianz-Matrix,
- Hesse Matrix („zweite Ableitung“ höhere Dimension),
- Gram Matrix
- Adjazenzmatrizen ungerichteter Graphen,
- ... .

Dies begründet sich häufig aus der Kommutativität der zugrundeliegenden mathematischen Operation.

Wir fassen grundlegende Eigenschaften der transponierten Matrix zusammen.

**Satz 2.18** Es sei  $A \in \mathbb{R}^{m \times n}$ , dann gilt

1.  $(A^T)^T = A$ ,
2.  $(AB)^T = B^T A^T$  mit  $B \in \mathbb{R}^{n \times k}$ ,
3.  $(\lambda A)^T = \lambda A^T$  mit  $\lambda \in \mathbb{R}$ ,
4.  $(A + B)^T = A^T + B^T$  mit  $B \in \mathbb{R}^{m \times n}$ ,

**Bemerkung 2.19** Insbesondere folgt aus Satz 2.18 (2) für  $x \in \mathbb{R}^n$ , also  $x \in \mathbb{R}^{n \times 1}$ ,

$$x^T \in \mathbb{R}^{1 \times n},$$

und

$$(Ax)^T = x^T A^T.$$

### 3 Das Skalarprodukt

Seien in diesem Kapitel die Komponenten von Vektoren  $x, y, z$  wie folgt:

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}.$$

**Definition 3.1** Eine Abbildung  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  heißt **Skalarprodukt** falls für beliebige  $x, y, z \in \mathbb{R}^n$  und  $\lambda \in \mathbb{R}$ , gilt:

1.  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ ,
2.  $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ ,
3.  $\langle x, y \rangle = \langle y, x \rangle$ ,
4.  $\langle x, x \rangle \geq 0$ , wobei  $\langle x, x \rangle = 0$  genau, dann wenn  $x = 0$ .

**Satz und Definition 3.2** Für  $x, y \in \mathbb{R}^n$  setzen wir

$$\langle x, y \rangle := \langle x, y \rangle_e := \sum_{k=1}^n x_k y_k.$$

Diese Abbildung  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  ist ein Skalarprodukt, und heißt **kanonisches Skalarprodukt** oder **euklidisches Skalarprodukt**.

In dieser Veranstaltung betrachten im Wesentlichen das euklidische Skalarprodukt, welches dabei der Kürze halber *Skalarprodukt* genannt, was auch die vereinfachte Notation  $\langle x, y \rangle := \langle x, y \rangle_e$  begründet. In Ausnahmefällen wird explizit darauf hingewiesen.

**Beweis.** Wir zeigen nur, dass 1. aus Definition 3.1 gilt. Die Übrigen folgen auf die gleiche Weise und können als Übung nachvollzogen werden. Es gilt

$$\langle x + y, z \rangle = \sum_{j=1}^n (x_j + y_j) z_j = \sum_{j=1}^n x_j z_j + \sum_{j=1}^n y_j z_j = \langle x, z \rangle + \langle y, z \rangle.$$

□

Das euklidische Skalarprodukt lässt sich als Matrix-Matrix-Produkt auffassen und ist somit formal schon weiter oben eingeführt.

### Beispiel

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \end{pmatrix} \right\rangle = 7, \quad \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\rangle = 0, \quad \left\langle \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \begin{pmatrix} 10 \\ 5 \end{pmatrix} \right\rangle = -25$$

**Bemerkung 3.3** Es sei  $A \in \mathbb{R}^{m \times n}$  und  $A_1, A_2, \dots, A_m \in \mathbb{R}^n$  die Zeilenvektoren von  $A$ , dann ist

$$Ax = \begin{pmatrix} \langle A_1, x \rangle \\ \langle A_2, x \rangle \\ \vdots \\ \langle A_m, x \rangle \end{pmatrix}.$$

Allgemeiner gilt für  $B \in \mathbb{R}^{n \times k}$  bestehend aus den Spaltenvektoren  $B_1, \dots, B_k \in \mathbb{R}^n$

$$AB = \begin{pmatrix} \langle A_1, B_1 \rangle & \langle A_1, B_2 \rangle & \dots & \langle A_1, B_k \rangle \\ \langle A_2, B_1 \rangle & \langle A_2, B_2 \rangle & \dots & \langle A_2, B_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle A_m, B_1 \rangle & \langle A_m, B_2 \rangle & \dots & \langle A_m, B_k \rangle \end{pmatrix}.$$

**Satz 3.4** Es seien  $A \in \mathbb{R}^{m \times n}$ ,  $x \in \mathbb{R}^n$  und  $y \in \mathbb{R}^m$ , dann gilt

$$\langle y, Ax \rangle = \langle A^T y, x \rangle.$$

**Beweis.** Übungsaufgabe. □

Die Namensgebung „euklidisch“ ist dadurch begründet, dass das Skalarprodukt direkt zur euklidischen Norm (euklidische Länge) führt.

**Definition 3.5** Es sei  $x \in \mathbb{R}^n$ , dann heißt

$$\|x\| := \|x\|_2 := \sqrt{\langle x, x \rangle} = \sqrt{\sum_{j=1}^n x_j^2}$$

die **euklidische Norm** von  $x$ .

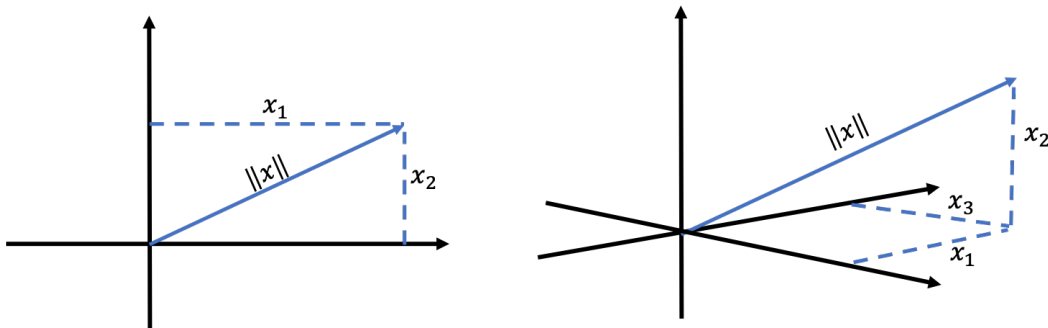


Figure 4: Euklidische Norm in  $\mathbb{R}^2$  und  $\mathbb{R}^3$

**Satz 3.6** Es seien  $x, y \in \mathbb{R}^n$  und  $\lambda \in \mathbb{R}$ , dann gilt für die euklidische Norm das Folgende:

1.  $\|x\| \geq 0$ , wobei  $\|x\| = 0 \Leftrightarrow x = 0$
2.  $\|\lambda x\| = |\lambda| \|x\|$
3.  $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$
4.  $\|x + y\| \leq \|x\| + \|y\|$  (Dreiecksungleichung)

**Beweis.** Punkte 1. und 2. ergeben sich unmittelbar und können als Übung nachvollzogen werden. 3. folgt durch Anwendung der Regeln in Definition 3.1:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x + y \rangle + \langle y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &= \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle. \end{aligned}$$

Punkt 4. Übungsaufgabe. □

**Satz 3.7** (Cauchy-Schwarzsche Ungleichung) Es seien  $x, y \in \mathbb{R}^n$ , dann gilt

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

**Beweis.** Falls  $y = 0$  gilt ist die Aussage trivial. Seien also  $y \neq 0$  und

$$\lambda = \frac{\langle x, y \rangle}{\langle y, y \rangle}.$$

Damit folgt

$$\begin{aligned} 0 &\leq \|x - \lambda y\|^2 = \langle x - \lambda y, x - \lambda y \rangle \\ &= \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \\ &= \langle x, x \rangle - 2 \frac{\langle x, y \rangle}{\langle y, y \rangle} \langle x, y \rangle + \frac{\langle x, y \rangle^2}{\langle y, y \rangle^2} \langle y, y \rangle \\ &= \langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle}. \end{aligned}$$

Es ergibt sich

$$\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle = \|x\|^2 \|y\|^2$$

woraus die Behauptung folgt. □

Damit ist für  $x, y \neq 0$  stets

$$\frac{|\langle x, y \rangle|}{\|x\| \|y\|} \leq 1.$$

**Definition 3.8** Für  $x, y \in \mathbb{R}^n$ ,  $x, y \neq 0$  ist der **Winkel** zwischen diesen Vektoren definiert als das kleinste nicht negative  $\alpha$ , für das

$$\cos(\alpha) = \frac{\langle x, y \rangle}{\|x\| \|y\|} \tag{3.5}$$

gilt. Zwei Vektoren  $x, y \in \mathbb{R}^n$  heißen **orthogonal**, wenn  $\langle x, y \rangle = 0$  gilt.

Für  $\mathbb{R}^2$  und  $\mathbb{R}^3$  ist diese Definition konsistent mit der geometrischen Definition des von  $x, y$  eingeschlossenen Winkels.

Das Skalarprodukt liefert also Informationen über die geometrische Ausrichtung zweier Vektoren zueinander. Dies lässt sich im  $\mathbb{R}^2$  und  $\mathbb{R}^3$  leicht veranschaulichen.

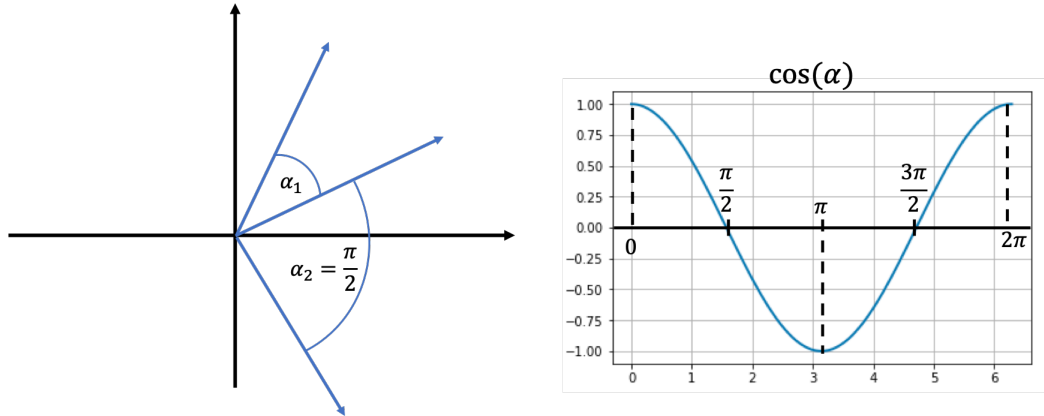


Figure 5: Winkel im Fall von Vektoren im  $\mathbb{R}^2$

Das Skalarprodukt wird auch bei der Verarbeitung hochdimensionaler Vektoren (z.B. digitale Signale oder andere Zeitreihen) sehr häufig eingesetzt. Aus dem Skalarprodukt werden dann beispielsweise Größen wie die Korrelation, Faltung und Kovarianz abgeleitet. Weiter lassen sich mit Hilfe des Skalarprodukts in einfacher Weise Geraden und Ebenen beschreiben.

**Bemerkung 3.9** 1. Es sei  $x \in \mathbb{R}^2$ ,  $x \neq 0$ , dann ist durch

$$H_x := \{y \in \mathbb{R}^2 : \langle x, y \rangle = 0\}$$

eine Ursprungsgerade im  $\mathbb{R}^2$  definiert. Diese Ursprungsgerade teilt den Raum  $\mathbb{R}^2$  in zwei disjunkte Teilmengen

$$H_x^+ := \{y \in \mathbb{R}^2 : \langle x, y \rangle > 0\}, \quad H_x^- := \{y \in \mathbb{R}^2 : \langle x, y \rangle < 0\}.$$

Weiter lässt sich eine solche Gerade in einfacher Weise im  $\mathbb{R}^2$  verschieben. Sei dazu  $u \in \mathbb{R}^2$  und  $b = \langle u, x \rangle$ , dann definiert

$$H_x := \{y + u \in \mathbb{R}^2 : \langle x, y \rangle = 0\}$$

eine Gerade parallel zu  $H_x$  durch  $u$ .

2. Es sei  $x \in \mathbb{R}^3$ ,  $x \neq 0$ , dann ist durch

$$H_x := \{y \in \mathbb{R}^3 : \langle x, y \rangle = 0\}$$

eine Ursprungsebene im Raum  $\mathbb{R}^3$  definiert. Diese Ursprungsebene teilt den  $\mathbb{R}^3$  in zwei disjunkte Teilmengen

$$H_x^+ := \{y \in \mathbb{R}^3 : \langle x, y \rangle > 0\}, \quad H_x^- := \{y \in \mathbb{R}^3 : \langle x, y \rangle < 0\}.$$

Weiter lässt sich eine solche Ebene in einfacher Weise im  $\mathbb{R}^3$  verschieben. Sei dazu  $u \in \mathbb{R}^2$  und  $b = \langle u, x \rangle$ , dann definiert

$$H_{x,b} := \{y \in \mathbb{R}^3 : \langle x, y \rangle = b\}$$

eine Ebene parallel zu  $H_x$  verschoben durch  $u \in \mathbb{R}^n$ .

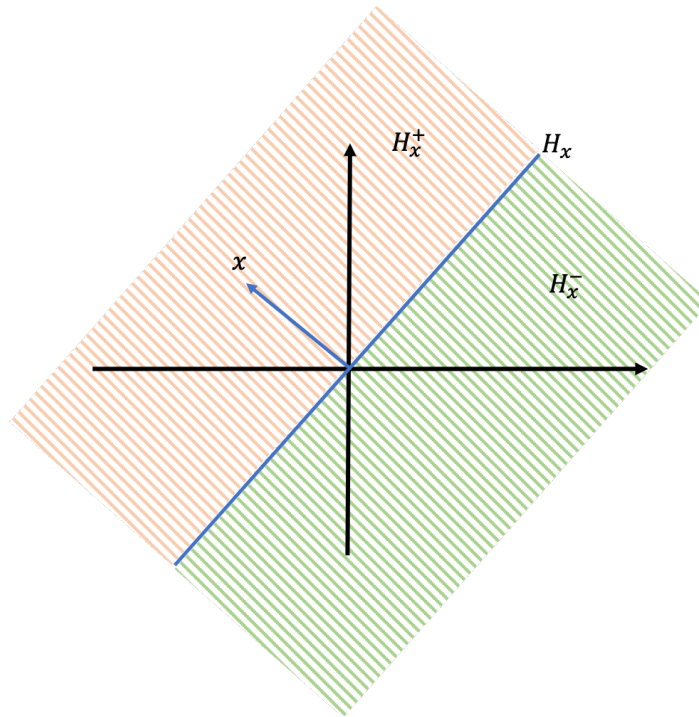


Figure 6: Hyperebene und disjunkte Unterteilung im  $\mathbb{R}^2$

In Bemerkung 3.9 werden  $H_x$  und  $H_{x,b}$  auch (die durch  $x$  definierte) Hyperebene genannt. Solche elementaren geometrischen Konstruktionen werden zum Beispiel sehr intensiv in

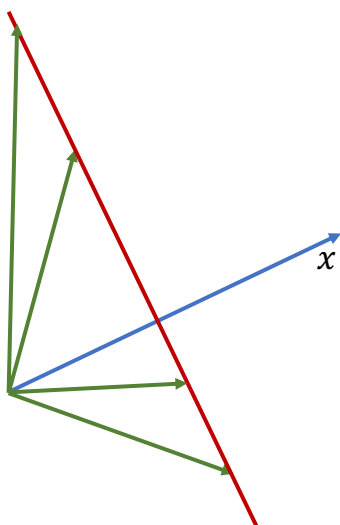


Figure 7: Das Skalarprodukt der grün dargestellten Vektoren mit  $x$  ist konstant (hier positiv) da die orthogonale Projektion auf  $x$  jeweils identisch ist. Darüber lässt sich die rote Gerade definieren.

der Computer-Grafik verwendet. Der Begriff der Hyperfläche ist sofort in den  $\mathbb{R}^n$  übertragbar.



**Definition 3.10** 1. Eine **Hyperebene** (im  $\mathbb{R}^n$ ) ist eine Menge der Form

$$H_{x,b} := \{y \in \mathbb{R}^n : \langle x, y \rangle = b\},$$

wobei  $b \in \mathbb{R}$  und  $x \in \mathbb{R}^n$  mit  $x \neq 0$  heißt **Normalenvektor** der Hyperebene.

2. Gilt in (1)  $n = 2$ , so ist eine Hyperebene eine Gerade im  $\mathbb{R}^2$ . Im Fall von  $n = 3$  ist eine Hyperebene eine Ebene im  $\mathbb{R}^3$ . Gilt dabei  $\|x\| = 1$  so heißt  $H_{x,b}$  **Hessesche Normalform** der Gerade bzw. der Ebene.

3. Eine Hyperebene teilt den Raum  $\mathbb{R}^n$  in zwei disjunkte **Halbräume**:

$$H_{x,b}^+ := \{y \in \mathbb{R}^n : \langle x, y \rangle > b\},$$

$$H_{x,b}^- := \{y \in \mathbb{R}^n : \langle x, y \rangle < b\},$$

4. Im Falle von  $b = 0$  schreiben wir kurz  $H_x, H_x^+, H_x^-$ .

Das Separieren von Halbräumen in Vektorräumen mittels Hyperebenen dient als Grundlage vieler Verfahren des maschinellen Lernens (lineare Klassifikatoren, support vector machine, kernel methods, ...) .

In der restringierten Optimierung, etwa der linearen Optimierung in oftmals hochdimensionalen Räumen, wird das Gebiet zulässiger Lösungen mathematisch typischerweise mittels Hyperebenen und Halbräumen modelliert (vgl. Modul lineare Optimierung (Master)).

## 4 Vektorräume und Lineare Abbildungen

In diesem Abschnitt betrachten wir Abbildungen der Form

$$x \mapsto Ax$$

wobei  $A \in \mathbb{R}^{m \times n}$ . Dazu führen wir zunächst die natürlichen Definitionsbereiche und Wertebereiche dieser Abbildungen, den **Vektorraum**, ein. Wir zeigen außerdem, dass jede sogenannte lineare Abbildung von  $\mathbb{R}^n$  nach  $\mathbb{R}^m$  in der oberen Form ausgedrückt werden kann.

**Definition 4.1** Es sei  $U \subset \mathbb{R}^n$ ,  $U \neq \emptyset$ , dann heißt  $U$  **Untervektorraum**, oder kurz **Vektorraum** von  $\mathbb{R}^n$ , wenn gilt:

1. für alle  $v, w \in U$  gilt  $v + w \in U$ ,
2. für alle  $\lambda \in \mathbb{R}$  und alle  $v \in U$  gilt  $\lambda v \in U$ .

**Bemerkung 4.2** Es sei ausdrücklich darauf hingewiesen, dass ein Unterraum  $U \subset \mathbb{R}^n$ , auch den Fall  $U = \mathbb{R}^n$  einschließt. Wir verwenden daher auch den Begriff **Vektorraum** dafür. Für eine kurze Einführung zu abstrakten Vektorräumen sei auf den Anhang verwiesen.

Die folgende Aussage liefert ein einfaches Kriterium mit dem sich häufig schon schnell ausschließen lässt, dass gewisse Mengen Unterräume sind.

**Bemerkung 4.3** 1. Aus Definition 4.1 (2) ergibt sich sofort, dass jeder Unterraum den Nullvektor enthält.

2. Die Menge  $\{0\}$ , die nur aus dem Nullvektor besteht ist immer ein Unterraum.

3.  $\mathbb{R}^n$  ist ein Vektorraum

**Satz 4.4** Es seien  $v_1, \dots, v_k \in \mathbb{R}^n$ , dann ist  $\text{Span}(v_1, \dots, v_k)$  ein Untervektorraum von  $\mathbb{R}^n$ .

**Beweis.** Offensichtlich gilt  $\text{Span}(v_1, \dots, v_k) \neq \emptyset$ . Es seien

$$v = \sum_{j=1}^k \lambda_j v_j, \quad u = \sum_{j=1}^k \alpha_j v_j$$

wobei  $\lambda_j, \alpha_j \in \mathbb{R}$ ,  $j = 1, \dots, k$ . Damit gilt

$$v + u = \sum_{j=1}^k (\lambda_j + \alpha_j) v_j \in \text{Span}(v_1, \dots, v_k)$$

woraus Punkt 1. in Definition 4.1 folgt. Punkt 2. kann auf die gleiche Weise gezeigt werden.  $\square$

Wir zeigen im Folgenden, dass umgekehrt zu Satz 4.4 jeder Vektorraum als Span geschrieben werden kann.

**Definition 4.5** Für den Vektorraum  $\mathbb{R}^n$  ist die sogenannte **Standardbasis** das  $n$ -Tupel  $(e_1, e_2, \dots, e_n)$  der Vektoren

$$e_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad e_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (4.6)$$

Wir schreiben auch  $e_j^n = e_j$ ,  $j = 1, \dots, n$ , wenn dies zu Zwecken der Eindeutigkeit des zugrundeliegenden Vektorraums nötig ist.

Ist  $x \in \mathbb{R}^n$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

als **eindeutige Linearkombination** mit Hilfe der Standardbasis darstellbar:

$$x = x_1 e_1 + x_2 e_2 + \dots + x_n e_n.$$

Diese Beobachtung lässt sich über die Standardbasis hinaus verallgemeinern und führt zu dem grundlegenden Begriff der Basis. Dazu definieren wir zunächst das Folgende:

**Definition 4.6** Vektoren  $v_1, \dots, v_m \in \mathbb{R}^n$  ( $m \in \mathbb{N}$ ) heißen **linear unabhängig**, falls

$$\sum_{j=1}^m \lambda_j v_j = 0$$

impliziert, dass  $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$  gilt. Andernfalls heißen die Vektoren  $v_1, \dots, v_m$  **linear abhängig**.

**Bemerkung 4.7** Man überzeugt sich leicht, dass gemäß Definition 4.6 eine Menge von Vektoren, die den Nullvektor enthält stets linear abhängig ist.

**Beispiel 4.8** 1.

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

sind linear unabhängig.

2.

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix},$$

sind linear abhängig.

**Satz 4.9** 1. Vektoren  $v_1, \dots, v_m \in \mathbb{R}^n$  sind genau dann linear abhängig, wenn mindestens einer der Vektoren als Linearkombination der anderen Vektoren dargestellt werden kann. Das heißt für ein  $k \in \{1, \dots, m\}$  gilt

$$v_k = \sum_{j=1, j \neq k}^m \lambda_j v_j$$

für gewisse  $\lambda_j \in \mathbb{R}$ ,  $j = 1, \dots, k-1, k+1, \dots, m$ . Dies ist weiter äquivalent zu der Aussage

$$\text{Span}(v_1, \dots, v_m) = \text{Span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m).$$

2. Sind  $v_1, \dots, v_m \in \mathbb{R}^n$  linear unabhängig und  $x \in \text{Span}(v_1, \dots, v_m)$ , dann gibt es eindeutige  $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$  mit

$$x = \sum_{j=1}^m \lambda_j v_j.$$

**Beweis.** Zu (1.): Es gilt

$$v_k = \sum_{j=1, j \neq k}^m \lambda_j v_j \Leftrightarrow 0 = -v_k + \sum_{j=1, j \neq k}^m \lambda_j v_j,$$

woraus die lineare Abhängigkeit der  $v_1, \dots, v_m$  folgt. Gilt umgekehrt mit  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  und  $\lambda_k \neq 0$  für ein  $k \in \{1, \dots, m\}$

$$0 = \sum_{j=1}^m \lambda_j v_j$$

so folgt sofort

$$v_k = \sum_{j=1, j \neq k}^m -\frac{\lambda_j}{\lambda_k} v_j.$$

Dass dies äquivalent ist zu

$$\text{Span}(v_1, \dots, v_m) = \text{Span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_m).$$

kann als Übung gezeigt werden.

Punkt 2. Übungsaufgabe.

□

Mit Hilfe des Begriffs der linearen Abhängigkeit können wir nun den zentralen Begriff der Basis definieren.

**Definition 4.10** Es sei  $V \subset \mathbb{R}^n$  ein Vektorraum. Dann heißt ein Tupel  $(v_1, \dots, v_m)$  bestehend aus Vektoren  $v_1, \dots, v_m \in V$ , **Basis** von  $V$ , falls gilt

1.  $v_1, v_2, \dots, v_m$  sind linear unabhängig, und
2.  $\text{Span}(v_1, \dots, v_m) = V$ .

Es gilt offensichtlich

$$\text{Span}(e_1^n, e_2^n, \dots, e_n^n) = \mathbb{R}^n.$$

Zusammen mit der leicht verifizierbaren linearen Unabhängigkeit der Vektoren erhalten wir.

**Satz 4.11**  $(e_1^n, \dots, e_n^n)$  ist eine Basis des  $\mathbb{R}^n$ . Somit existiert zu jedem dieser Räume eine Basis.

Die Eindeutigkeit der Darstellung aus Satz 4.9 (2) stellt sicher, dass Folgendes wohldefiniert ist.

**Definition 4.12** Es sei  $V \subset \mathbb{R}^n$  ein Vektorraum und  $\mathcal{B} := (v_1, \dots, v_m)$  eine Basis von  $V$ . Für  $x \in V$  heißen die eindeutig bestimmten Skalare  $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$  so, dass

$$x = \sum_{j=1}^m \lambda_j v_j$$

gilt, die **Koordinaten** von  $x$  bezüglich  $\mathcal{B}$ . Wir schreiben hierfür auch

$$x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}_{\mathcal{B}}.$$

Mit Hilfe des folgenden Resultats lassen sich weitere Basen zu erzeugen.

**Satz 4.13** (*Austauschlemma*) Es seien  $v_1, \dots, v_m \in \mathbb{R}^n$  linear unabhängig. Sei weiter  $x \in \mathbb{R}^n$ ,  $x$  mit

$$x = \sum_{j=1}^m \lambda_j v_j, \quad \lambda_j \in \mathbb{R}, \quad j = 1, \dots, m,$$

und gilt dabei  $\lambda_k \neq 0$  für mindestens ein  $k \in \{1, \dots, m\}$ , dann sind  $v_1, \dots, v_{k-1}, x, v_{k+1}, \dots, v_m$  linear unabhängig, und es gilt

$$\text{Span}(v_1, \dots, v_m) = \text{Span}(v_1, \dots, v_{k-1}, x, v_{k+1}, \dots, v_m).$$

**Beweis.** Ohne Einschränkung sei  $k = 1$ . Wegen der linearen Unabhängigkeit von  $v_1, \dots, v_m$  und der damit einhergehend Eindeutigkeit der Linearkombinationen (vgl. Satz 4.9 (2)) gilt für  $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m \in \mathbb{R}$

$$\sum_{j=1}^m \alpha_j v_j = \beta_1 x + \sum_{j=2}^m \beta_j v_j$$

genau dann, wenn

$$\beta_1 \lambda_1 = \alpha_1 \quad \text{und} \quad \beta_1 \lambda_j + \beta_j = \alpha_j, \quad j = 2, \dots, m,$$

was wiederum genau dann gilt, wenn

$$\beta_1 = \frac{\alpha_1}{\lambda_1} \quad \text{und} \quad \beta_j = \alpha_j - \frac{\alpha_1 \lambda_j}{\lambda_1}, \quad j = 2, \dots, m.$$

Dies zeigt  $\text{Span}(v_1, \dots, v_m) = \text{Span}(v_1, \dots, v_{k-1}, x, v_{k+1}, \dots, v_m)$ . Weiterhin folgt aus dem Oberen

$$\alpha_j = 0 \quad \forall j = 1, \dots, m \Leftrightarrow \beta_j = 0 \quad \forall j = 1, \dots, m,$$

womit, wiederum wegen der linearen Unabhängigkeit von  $v_1, \dots, v_m$ , die lineare Unabhängigkeit von  $x, v_2, \dots, v_m$  folgt.  $\square$

Damit können wir nun Folgendes herleiten.

**Satz 4.14** (*Basiseristenzsatz und Basisergänzungssatz*)

1. Es sei  $V \subset \mathbb{R}^n$ ,  $V \neq \{0\}$  ein Vektorraum, dann existiert eine Basis  $(v_1, \dots, v_m)$  von  $V$  und jede Basis hat die gleiche Länge (Anzahl Vektoren) und dabei gilt  $m \leq n$ .
2. Es sei  $V \subset \mathbb{R}^n$  ein Vektorraum und  $v_1, \dots, v_k \in V$  linear unabhängig. Dann gibt es Vektoren  $v_{k+1}, \dots, v_m$  so, dass  $(v_1, \dots, v_m)$  eine Basis von  $V$  bilden.

**Beweis.** Zu (1): Wir wählen sukzessive linear unabhängige Vektoren  $v_1, \dots, v_m \in V \setminus \{0\}$ . Dieser Prozess muss spätestens nach  $n$  Iterationen stoppen, da jeder Vektor als Linearkombination der  $e_1, \dots, e_n$  dargestellt werden kann und weil die  $e_1, \dots, e_n$  nicht mehr als  $n$  linear unabhängige Vektoren darstellen können (\*Übung). Sei also  $(v_1, \dots, v_m)$  mit  $m \leq n$  das Resultat dieser Prozedur. Da alle Vektoren als Linearkombination der  $e_1, \dots, e_n$  darstellbar sind, gibt es somit keinen weiteren Vektor in  $V$ , der zusammen mit  $v_1, \dots, v_m$  eine linear unabhängige Menge von Vektoren bildet. Gäbe es nun einen Vektor  $u \in V$  mit  $u \notin \text{Span}(v_1, \dots, v_m)$ , dann wären  $u, v_1, \dots, v_m$  nach Satz 4.9 linear unabhängig. Somit folgt  $V = \text{Span}(v_1, \dots, v_m)$ .

Zu (2): Es sei  $(u_1, \dots, u_m)$  eine Basis von  $V$ . Mit Hilfe des Austauschlemmas 4.13 lässt sich sukzessive eine Basis erzeugen in der jeweils ein Vektor der  $u_1, \dots, u_m$  durch einen Vektor der  $v_1, \dots, v_k$  ersetzt wurde. Die lineare Unabhängigkeit der  $v_1, \dots, v_k$  garantiert dabei, dass für die Darstellung eines jeden  $v_j$  mindestens ein Vektor der  $u_1, \dots, u_m$  mit einem Koeffizienten ungleich 0 eingehen muss. Es verbleiben am Ende  $m - k$  Vektoren der ursprünglichen Basis  $(u_1, \dots, u_m)$ . Durch Umbenennung dieser Vektoren folgt die Behauptung.

□

Wir haben das Ergebnis aus Satz 4.9 somit erweitert und wissen nun, dass jeder Vektorraum  $V$  wie folgt dargestellt werden kann

$$V = \text{Span}(v_1, \dots, v_m).$$

Satz 4.14 zeigt, dass zwei Basen eines Vektorraums stets die gleiche Anzahl von Vektoren enthalten, was die folgende Definition rechtfertigt.



**Definition 4.15** Es sei  $V \subset \mathbb{R}^n$  ein Vektorraum und  $(v_1, \dots, v_m)$  ( $m \in \mathbb{N}$ ) eine Basis, dann heißt  $m$  die **Dimension** von  $V$  und wir schreiben hierfür

$$\dim V = m.$$

Man sagt auch  $V$  ist ein  $m$ -dimensionaler Vektorraum. Weiter setzen wir für den Vektorraum, der nur aus dem Nullvektor besteht

$$\dim\{0\} = 0.$$

**Bemerkung 4.16** Ist  $V \subset \mathbb{R}^n$  ein Vektorraum mit  $\dim V = m$ . Dann ist jede Menge von  $m$  linear unabhängigen Vektoren in  $V$  eine Basis von  $V$ .

**Definition 4.17** Es seien  $V \subset \mathbb{R}^m$  und  $U \subset \mathbb{R}^n$  Vektorräume und  $f : U \rightarrow V$  eine Abbildung, dann heißt  $f$  **linear** falls gilt:

1. für alle  $v, w \in U$  ist  $f(v + w) = f(v) + f(w)$ ,
2. für alle  $\lambda \in \mathbb{R}$  und alle  $v \in U$  ist  $f(\lambda v) = \lambda f(v)$ .

Die obige Definition besagt, dass lineare Abbildungen die strukturerhaltende Abbildungen (also Homomorphismen) auf Vektorräumen sind.

Aus Satz 2.8 ergibt sich sofort:

**Satz 4.18** Es sei  $A \in \mathbb{R}^{m \times n}$ , dann ist die Abbildung  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  definiert durch

$$f(x) = Ax$$

linear.

Wir werden bald sehen, dass alle linearen Abbildungen zwischen  $\mathbb{R}^n$  und  $\mathbb{R}^m$  stets als  $x \mapsto Ax$  mit einer Matrix  $A$  geschrieben werden können.

**Bemerkung 4.19** Für jede lineare Abbildung  $f : U \rightarrow V$  gilt  $f(0) = 0$ .

Die obere Bemerkung liefert eine Art Schnelltest auf Linearität einer Abbildung

**Beispiel 4.20** 1. Die Abbildung  $(x_1, x_2)^T \mapsto 3x_1 + 4x_2$  ist linear

2. Die Abbildung  $x \mapsto \|x\|$  mit  $x \in \mathbb{R}^n$  ist nicht linear, denn es etwa  $e_1 \mapsto \|e_1\| = 1$  und  $-e_1 \mapsto \|-e_1\| = 1$ , jedoch  $e_1 - e_1 \mapsto \|e_1 - e_1\| = \|0\| = 0$ .

**Definition 4.21** Es seien  $V \subset \mathbb{R}^n$  und  $U \subset \mathbb{R}^m$  Vektorräume.

1. Eine bijektive lineare Abbildung  $f : U \rightarrow V$  heißt **Isomorphismus**.
2. Existiert zwischen  $U$  und  $V$  ein Isomorphismus, so sind  $U$  und  $V$  **isomorph** zueinander.

Man kann isomorphe Vektorräume im Bezug auf ihre Vektorraumeigenschaften als gleich betrachten.

**Satz 4.22** Es seien  $V \subset \mathbb{R}^n$  und  $U \subset \mathbb{R}^m$  Vektorräume und  $f : U \rightarrow V$  ein Isomorphismus, dann ist auch die inverse Abbildung  $f^{-1} : V \rightarrow U$  ein Isomorphismus.

**Beweis.** Wegen der Bijektivität ist klar, dass die inverse Abbildung  $f^{-1} : V \rightarrow U$  existiert und ebenfalls bijektiv ist (vgl. mathematische Grundlagen). Für beliebige  $v_1, v_2 \in V$  gibt es daher eindeutige  $u_1, u_2 \in U$  mit  $f(u_1) = v_1$  und  $f(u_2) = v_2$  und wegen der Linearität von  $f$  folgt  $f(u_1 + u_2) = f(u_1) + f(u_2) = v_1 + v_2$ . Damit ergibt sich

$$f^{-1}(v_1) + f^{-1}(v_2) = u_1 + u_2 = f^{-1}(v_1 + v_2).$$

Für  $\lambda \in \mathbb{R}$  gilt wegen der Linearität von  $f$  weiter  $f(\lambda u_1) = \lambda v_1$  und somit

$$f^{-1}(\lambda v_1) = \lambda u_1 = \lambda f^{-1}(v_1).$$

□

Abbildungen auf endlichdimensionalen Räumen, wie sie in dieser Vorlesung betrachtet werden, lassen sich stets mithilfe von Matrizen darstellen. Es ergibt sich daher folgende Charakterisierung:

**Satz 4.23** Es sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  eine Abbildung. Dann ist  $f$  genau dann eine lineare Abbildung, wenn es eine eindeutig definierte Matrix  $A \in \mathbb{R}^{m \times n}$  gibt, so dass

$$f(x) = Ax$$

für alle  $x \in \mathbb{R}^n$  gilt.

Beweis [vgl. Übung]

Somit können wir alle linearen Abbildungen auf den Gesamträumen (also  $\mathbb{R}^n$ ) in eindeutiger Weise mit Matrizen identifizieren, was die folgende Definition rechtfertigt.

**Definition 4.24** Es seien  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  eine lineare Abbildung und  $A$  die nach Satz 4.23 eindeutig bestimmte Matrix, dann nennen wir  $A$  die **darstellende Matrix** (von  $f$ ), und umkehrt heißt  $f$  die durch  $A$  definierte lineare Abbildung.

**Beispiel 4.25** Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  definiert durch

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 2x_1 - 3x_2 \\ x_2 \\ -x_1 + 7x_2 \end{pmatrix},$$

dann ergibt sich folgende darstellende Matrix  $A = (a_1 \ a_2) \in \mathbb{R}^{3 \times 2}$  mit  $a_1 = f(e_1) \in \mathbb{R}^3$ ,  $a_2 = f(e_2) \in \mathbb{R}^3$ , also

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

Die Verallgemeinerung dieser Prozedur liefert die Konstruktion der Matrix für den Beweis von Satz 4.23.

**Bemerkung 4.26** Sind  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  zwei lineare Abbildungen mit darstellenden Matrizen  $A \in \mathbb{R}^{n \times m}$  und  $B \in \mathbb{R}^{m \times k}$ , dann ist  $g \circ f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  ebenfalls linear und die zugehörige darstellende Matrix ist gegeben durch  $BA$ .

Da wir nun wissen, dass lineare Abbildungen von  $\mathbb{R}^n$  nach  $\mathbb{R}^m$  stets durch eine eindeutige Matrix beschrieben sind, beschränken wir uns in den weiteren Definitionen und Sätzen dieses Abschnitts darauf lineare Abbildungen als Matrizen zu betrachten.

Die fundamentalen Vektorräume im Zusammenhang mit Matrizen sind die folgenden:

**Definition 4.27** Es sei  $A \in \mathbb{R}^{m \times n}$  mit Spaltenvektoren  $a_1, \dots, a_n \in \mathbb{R}^m$ . Dann sind

1.  $\text{Kern } A := \{x \in \mathbb{R}^n : Ax = 0\}$  der **Kern** von  $A$ , und
2.  $\text{Im } A := \text{Span}(a_1, \dots, a_n)$  das **Bild** von  $A$ .

**Satz 4.28** Es sei  $A \in \mathbb{R}^{m \times n}$ .

*Die Abbildung  $x \mapsto Ax$  ist genau dann injektiv, wenn  $\text{Kern } A = \{0\}$  gilt.*

*Die Abbildung  $x \mapsto Ax$  bildet surjektiv nach  $\mathbb{R}^m$  ab, wenn  $\text{Im } A = \mathbb{R}^m$  gilt.*

*$\text{Kern } A$  ist ein Unterraum von  $\mathbb{R}^n$  und  $\text{Im } A$  ein Unterraum von  $\mathbb{R}^m$ .*

**Beweis.** 1. Ist die Abbildung injektiv dann gilt wegen  $A0 = 0$ , dass  $Ax \neq 0$  für alle  $x \in \mathbb{R}^n \setminus \{0\}$  und somit  $\text{Kern } A = \{0\}$ . Gilt umgekehrt  $\text{Kern } A = \{0\}$ , dann gilt für beliebige  $x, y \in \mathbb{R}^n$

$$Ax = Ay \Leftrightarrow A(x - y) = 0 \Leftrightarrow x - y \in \text{Kern } A \Leftrightarrow x = y.$$

2. Die Aussage folgt unmittelbar aus der Beobachtung, dass  $\text{Im } A$  die Bildmenge von  $\mathbb{R}^n$  unter der Abbildung  $x \mapsto Ax$  ist.

3. vgl. Übung □

**Definition 4.29** Der **Rang** einer Matrix  $A \in \mathbb{R}^{m \times n}$  ist die maximale Anzahl linear unabhängiger Spaltenvektoren. Für eine Matrix  $A$  mit Rang  $k$  schreiben wir

$$\text{rang } A = k.$$

**Bemerkung 4.30** 1. Der Rang gemäß Definition 4.29 wird, aus naheliegenden Gründen, auch **Spaltenrang** genannt, und entsprechend nennt man  $\text{rang } A^T$  den **Zeilenrang** von  $A$ . Wir werden später sehen, dass diese Unterscheidung nicht notwendig ist, da diese beiden Größen identisch sind.

2. Man kann, etwa mithilfe des Austauschlemmas leicht zeigen, dass

$$\text{rang } A = \dim(\text{Im } A)$$

gilt.

## 5 Inverse Matrix, lineare Gleichungssysteme

In diesem Abschnitt betrachten wir folgende Gleichung

$$\boxed{Ax = b} \quad (5.7)$$

mit  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$ .

Wir setzen diese Gleichung in Beziehung zu linearen Abbildungen und studieren damit die Lösbarkeit und die Struktur von Lösungsmengen dieser Gleichung.

Wie bereits aus Kapitel 1 bekannt, ist die Einheitsmatrix  $E$  das **neutrale** Element bezüglich des Matrix-Matrix-Produkts der quadratischen Matrizen ist, d.h. es gilt

$$AE = EA = A, \quad \forall A \in \mathbb{R}^{n \times n}.$$

Die Existenz eines neutralen Elements zieht gleich die Frage nach einem inversen Element nach sich, also die Frage ob es zu einer Matrix  $A \in \mathbb{R}^{n \times n}$  eine Matrix  $B \in \mathbb{R}^{n \times n}$  gibt so, dass  $BA = E$ .

**Definition 5.1** Es sei  $A \in \mathbb{R}^{n \times n}$ . Falls es eine Matrix  $B \in \mathbb{R}^{n \times n}$  gibt, so dass  $BA = E$  gilt, dann heißt  $A$  **invertierbar**, und  $A^{-1} := B$  heißt **inverse Matrix** von  $A$ .

**Beispiel 5.2** Es gilt

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

**Bemerkung 5.3** Ist  $A$  invertierbar, so gilt  $A^{-1}A = AA^{-1} = E$ . Weiter ist dann die inverse Matrix  $A^{-1}$  eindeutig bestimmt und selbst wieder invertierbar. Dabei gilt

$$(A^{-1})^{-1} = A.$$

Es sind jedoch nicht alle quadratischen Matrizen invertierbar. Die Invertierbarkeit hängt unmittelbar mit den Eigenschaften der durch die Matrix definierten Abbildung  $x \mapsto Ax$  zusammen, wie folgende Überlegung zeigt: Ist  $A$  invertierbar, dann ist die Abbildung  $x \mapsto A^{-1}Ax = Ex = x$  die Identität und somit ist  $x \mapsto A^{-1}x$  die inverse

Abbildung (Umkehrabbildung) von  $x \mapsto Ax$ , und umgekehrt ist  $x \mapsto A^{-1}x$  die inverse Abbildung von  $x \mapsto A^{-1}x$ . Es ergibt sich das folgende Ergebnis:

**Satz 5.4** *Es sei  $A \in \mathbb{R}^{n \times n}$  invertierbar. Dann gilt Folgendes:*

1. *Die Abbildung  $x \mapsto A^{-1}x$  ist die inverse Abbildung (Umkehrabbildung) von  $x \mapsto Ax$*
2. *Ist weiter  $B \in \mathbb{R}^{n \times n}$  ebenfalls invertierbar, dann ist auch  $AB$  invertierbar und es gilt*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

**Beweis.** (1.) ergibt sich unmittelbar aus den Überlegungen vor Satz 5.4.

Zu (2.): Da die inverse Matrix nach Bemerkung 5.3 eindeutig ist folgt die Behauptung aus

$$AB B^{-1}A^{-1} = AEA^{-1} = E.$$

□

Daher sind im Fall einer invertierbaren Matrix  $A \in \mathbb{R}^{n \times n}$  die Abbildungen  $x \mapsto Ax$  und  $x \mapsto A^{-1}x$  Isomorphismen. Das zeigt, dass diese Eigenschaft schon notwendig für die Invertierbarkeit ist. Die umgekehrt Implikation gilt auch.

**Satz 5.5** *Es sei  $A \in \mathbb{R}^{n \times n}$ . Dann sind die folgenden Aussagen äquivalent:*

1.  *$A$  ist invertierbar,*
2.  *$x \mapsto Ax$  ist ein Isomorphismus von  $\mathbb{R}^n$  nach  $\mathbb{R}^n$ .*

**Beweis.** Die Implikation (1.)  $\Rightarrow$  (2.) ist schon in Satz 5.4 formuliert.

Zur Implikation (1.)  $\Leftarrow$  (2.): Es gelte also  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $f(x) := Ax$  ist ein Isomorphismus. Dann ist nach Satz 4.22 die Umkehrfunktion  $f^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  ebenfalls linear und kann nach Satz 4.23 durch eine Matrix  $B \in \mathbb{R}^{n \times n}$  dargestellt werden, d.h.  $f^{-1}(x) = Bx$  für alle  $x \in \mathbb{R}^n$ . Es gilt dann

$$x = f^{-1}(f(x)) = B(Ax) = BA \quad \forall x \in \mathbb{R}^n.$$

Daraus ergibt sich  $BA = E$  (Übungsaufgabe) und somit muss  $B = A^{-1}$  gelten. □

**Beispiel 5.6** Die Matrix

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & -1 \\ -2 & -4 & 7 \end{pmatrix}$$

nicht invertierbar. Dies folgt zum Beispiel aus  $A2e_1 = Ae_2$  (wieso? Übung).

Mit Hilfe der inversen Matrix lässt sich ein lineares Gleichungssystem

$$Ax = b$$

mit  $A \in \mathbb{R}^{n \times n}$  invertierbar und  $b \in \mathbb{R}^n$  sofort lösen, denn Multiplikation beider Seiten mit  $A^{-1}$  von links liefert

$$x = Ax = A^{-1}b.$$

Es lassen sich aber auch allgemeine Matrix-Gleichungen auflösen, wie folgendes Beispiel zeigt.

**Beispiel** Es seien  $A, B \in \mathbb{R}^{n \times n}$  invertierbar und  $C, D, F \in \mathbb{R}^{m \times n}$ , dann lässt sich die Gleichung

$$(C + D)AB = F$$

beispielsweise nach  $D$  auflösen:

$$D = FB^{-1}A^{-1} - C.$$

Da das Matrix-Matrix-Produkt nicht kommutativ ist, muss in solchen Fällen zwischen Multiplikation von rechts und links zu unterscheiden.

Die oberen Rechnungen setzen jedoch die Kenntnis der inversen Matrix voraus. In der Praxis ist die Berechnung inverser Matrizen relativ aufwändig.

Im folgenden betrachten wir allgemeine **lineare Gleichungssysteme (LGS)**, also eines Gleichungssystems der Form



$$Ax = b \tag{5.8}$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , wobei  $x \in \mathbb{R}^n$  gesucht ist.

Neben theoretischer Grundlagen zu Lösbarkeit, wird das Gauß-Verfahren (auch Gaußsches Eliminationsverfahren oder Gauß-Algorithmus genannt) als Verfahren zum Lösen solcher LGS eingeführt.

Wir bezeichnen mit

$$\text{Lsg}(A, b) := \{x \in \mathbb{R}^n : Ax = b\}$$

die **Lösungsmenge** von (5.8).

Das Gauß-Verfahren verwendet die folgenden drei Operation.

**Definition 5.7** Die folgenden Operationen auf Matrizen heißen **elementare Zeilenumformungen**:

**Z1** Vertauschen zweier Zeilen.

**Z2** Addition des  $\lambda$ -fachen einer  $j$ -ten Zeile zu einer  $k$ -ten Zeile, wobei gilt  $j \neq k$  und  $\lambda \in \mathbb{R}$ .

**Z3** Multiplikation einer Zeile mit einem Skalar  $\lambda \neq 0$ .

Diese elementaren Umformungen, Z1, Z2, Z3, werden im Gauß-Verfahren zur Bestimmung von  $\text{Lsg}(A, b)$  angewendet. Diese Umformungen entsprechen der Multiplikation der Matrix  $A$  mit sogenannten Elementarmatrizen von links, die sich wie folgt ergeben:

Es sei für das Folgende  $A \in \mathbb{R}^{m \times n}$ .

1. Es sei  $E_{Z1}(j, k) \in \mathbb{R}^{m \times m}$  die Matrix, die durch Vertauschen der  $j$ -ten

und  $k$ -ten Zeile der Einheitsmatrix  $E_m$  entsteht. Zum Beispiel

$$E_{Z1}(3, 6) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{9 \times 9}$$

Dann ist das Produkt  $E_{Z1}(j, k) A$  identisch mit der Matrix, die durch Vertauschen der  $j$ -ten und  $k$ -ten Zeile von  $A$  entsteht, und entspricht somit einer Umformung  $Z1$ .

2. Es sei  $E_{Z2}(\lambda, j, k) \in \mathbb{R}^{m \times m}$  mit  $j \neq k$  die Matrix, die durch Addition des  $\lambda$ -fachen der  $j$ -ten Zeile zur  $k$ -ten Zeile der Einheitsmatrix  $E_m$  entsteht. Zum Beispiel:

$$E_{Z2}(5, 2, 6) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, E_{Z2}(9, 5, 3) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \in \mathbb{R}^{7 \times 7}$$

Dann ist das Produkt  $E_{Z2}(\lambda, j, k) A$  identisch zu der Matrix, die durch das Addieren des  $\lambda$ -fachen der  $j$ -ten Zeile zur  $k$ -ten Zeile von  $A$  entsteht, und entspricht somit einer Umformung  $Z2$ .

3. Es sei  $E_{Z3}(\lambda, j) \in \mathbb{R}^{m \times m}$  mit  $\lambda \neq 0$  die Matrix, die durch Multiplikation

der  $j$ -ten Zeile der Einheitsmatrix  $E_m$  mit  $\lambda$  entsteht.

$$E_{Z2}(6, 4) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{9 \times 9}$$

Dann ist das Produkt  $E_{Z3}(\lambda, j) A$  identisch mit der Matrix, die durch Multiplizieren der  $j$ -ten Zeile von  $A$  mit  $\lambda$  entsteht, und entspricht somit einer Umformung  $Z3$ .

Die hier eingeführten Matrizen heißen **Elementarmatrizen**.

Im Gauß-Verfahren wird, allgemein gesprochen, ein System (5.8) durch iteratives Anwenden elementarer Zeilenumformungen in ein Gleichungssystem umgeformt, das unmittelbar (durch eine einfache Prozedur) lösbar ist, vorausgesetzt  $\text{Lsg}(A, b) \neq \emptyset$ . Das heißt

$$Ax = b$$

wird umgeformt zu

$$B_s B_{s-1} \cdots B_2 B_1 Ax =: A'x = B_s B_{s-1} \cdots B_2 B_1 b =: b',$$

wobei  $B_1, \dots, B_s$  Elementarmatrizen sind. Dabei ist dann  $A'x = b'$  von einer Form, die unmittelbares Lösen ermöglicht, falls  $\text{Lsg}(A, b) \neq \emptyset$ .

Um dies tun zu dürfen, muss sichergestellt sein, dass

$$\text{Lsg}(A, b) = \text{Lsg}(A', b')$$

gilt. Weil alle Elementarmatrizen invertierbar sind, folgt das aus dem folgenden Ergebnis.

**Satz 5.8** *Es sei  $B \in \mathbb{R}^{m \times m}$  invertierbar und  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , dann gilt*

$$\text{Lsg}(A, b) = \text{Lsg}(BA, Bb).$$

**Beweis.** Die Behauptung folgt sofort aus der Beobachtung, dass durch Multiplikation von

$$Ax = b$$

mit  $B$  von links, sofort das Gleichungssystem

$$BAx = Bb$$

entsteht, und durch Multiplikation des letzten Gleichungssystems mit  $B^{-1}$  von links wiederum das ursprüngliche Gleichungssystem entsteht. Das Multiplizieren mit einer invertierbaren Matrix ist hier somit eine äquivalente Umformung.  $\square$

Im Anhang wird eine algorithmische Beschreibung des Gauß-Verfahrens geliefert.

An dieser Stelle wird das Verfahren ohne eine genaue Festlegung der Rechenschritte eingeführt. Es wird vielmehr ein Ziel definiert, und das grundsätzliche Vorgehen dazu dargestellt

## Gauß-Verfahren

Es seien

$$A = \begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & & \\ a_{m1} & a_{m2} \dots & a_{mn} \end{pmatrix}, \quad \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = b$$

gegeben. Dieses System wird im Gauß-Verfahren mittels elementarer Zeilenumformungen

so umgeformt, dass  $A$  in eine Matrix  $A'$  in **Zeilenstufenform** umgeformt wird:

$$A' = \begin{pmatrix} 0 & \dots & \boxed{a'_{1k_1}} & (*) & \dots & (*) & (*) & (*) & (*) & \dots & \dots & \dots & \dots & \dots & (*) \\ 0 & \dots & 0 & \dots & 0 & \boxed{a'_{2k_2}} & (*) & (*) & (*) & \dots & \dots & \dots & \dots & \dots & (*) \\ 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{a'_{3k_3}} & (*) & \dots & \dots & \dots & \dots & (*) \\ \vdots & & & & \vdots & & & & \vdots & \vdots & & & & \vdots & \\ 0 & \dots & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & 0 & \boxed{a'_{rk_r}} & (*) & \dots & (*) \\ 0 & \dots & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 \\ \vdots & & & & \vdots & & & & \vdots & & & & \vdots & & \\ 0 & \dots & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 \end{pmatrix}. \quad (5.9)$$

Die unteren Zeilenumformungen werden **stets simultan auf  $A$  und  $b$  angewendet**, sodass sich gemäß Satz 5.8 die Lösungsmenge nicht ändert.

Dazu geht man spaltenweise von der linken Spalte beginnend vor. Zunächst erzeugt man in der erste Spalte Nulleinträge, mit Ausnahme des obersten Eintrags, indem man zum Beispiel

$$E_{Z2} \left( -\frac{a_{k1}}{a_{11}}, 1, k \right)$$

für  $k = 2, \dots, m$  von links auf  $A$  und  $b$  anwendet.

Falls  $a_{11} = 0$  gilt ist das natürlich nicht möglich. In diesem Fall tauscht man mittels  $Z1$  Zeilen in geeigneter Weise.

Dadurch soll  $A$  zu einer Matrix  $A'$  der Form

$$\begin{pmatrix} a'_{11} & a'_{12} \dots & a'_{1n} \\ 0 & a'_{22} \dots & a'_{2n} \\ \vdots & & \\ 0 & a'_{m2} \dots & a'_{mn} \end{pmatrix}$$

umgeformt sein. Anschließend geht man zur zweiten Spalte von links und verfährt analog. Man erzeugt 0–Einträge, mit Ausnahme der obersten beiden Einträge. Wichtig ist dabei, dass man nur noch mit den Zeilen  $2, \dots, m$  arbeitet. Ansonsten könnten die Nulleinträge der ersten Spalte zerstört werden.

In dieser Weise verfährt man bis man entweder bei der  $n$ –ten Spalte angelangt ist.

Generell ist zu beachten, dass das Erzeugen von Nulleinträgen in einer  $k$ -ten Spalte simultan auch Nulleinträge in den gleichen Zeilen der  $k + 1$ -ten Spalte, oder weiteren rechts gelegenen Spalten, entstehen können. Solche Spalten bedürfen dann keiner weiteren Umformung und man kann direkt zu einer nächsten Spalte gehen. Daher können die Stufen in der Zeilenstufenform (5.9) aus mehr als einem Eintrag in der Breite bestehen. Wichtig ist, dass **jede Stufe aus einem Eintrag in der Höhe besteht**.

Mit der Matrix  $A'$  und entsprechend umgeformten  $b'$  lässt sich dann wie folgt die Lösung(en) bestimmen.

Seien dazu die Indizes  $r \in \{1, \dots, m\}$  und  $k_1 < k_2 < \dots < k_r$ ,  $k_j \in \{1, \dots, n\}$ ,  $j = 1, \dots, r$  ( $r \leq m$  und  $r \leq n$ ) und

$$a'_{1k_1}, a'_{1k_2}, \dots, a'_{rk_r}$$

die entsprechenden Einträge in der Zeilenstufenform (5.9) so dass

$$a'_{jk_j} \neq 0, \quad j = 1, \dots, r.$$

Damit entspricht  $r$  der **Anzahl der Stufen**. Sei weiter

$$b' = \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_m \end{pmatrix}$$

der Vektor, der durch die Zeilenumformungen aus  $b$  entsteht.

Damit kann

$$\text{Lsg}(A, b) := \{x \in \mathbb{R}^n : Ax = b\}$$

wie folgt bestimmt werden.

1. Fall: Es gibt ein  $l > r$  mit  $b'_l \neq 0$ . Dann gilt

$$\text{Lsg}(A, b) = \emptyset.$$

2. Fall: Für alle  $l > r$  gilt  $b'_l = 0$ . Dann  $\text{Lsg}(A, b) \neq \emptyset$  Eine Lösung

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

kann dann beispielsweise wie folgt konstruiert werden. Für  $j \notin \{k_1, \dots, k_r\}$  setze

$$x_j = 0.$$

(Die freien Variablen werden  $= 0$  gesetzt). Weiter setzt man, beginnend an der rechten unteren Seite der Matrix

$$x_{k_r} = \frac{b_r}{a'_{rk_r}},$$

und dann rekursiv

$$x_{k_j} = \frac{1}{a'_{jk_j}} \left( b_j - \sum_{l=j+1}^r x_{k_l} a'_{jk_l} \right), \quad j = r-1, \dots, 1.$$

Setzt man die freien Variablen allgemein für  $j \notin \{k_1, \dots, k_r\}$  durch

$$x_j = \xi_j$$

für irgendwelche  $\xi_j \in \mathbb{R}$ , so ergibt sich die zugehörige Lösung weiter durch

$$x_{k_r} = \frac{1}{a'_{rk_r}} \left( b_r - \sum_{l=k_r+1}^n \xi_l a'_{jl} \right),$$

und dann rekursiv

$$x_{k_j} = \frac{1}{a'_{jk_j}} \left( b_j - \sum_{l=j+1}^r x_{k_l} a'_{jk_l} - \sum_{\substack{l=k_j+1 \\ l \notin \{k_{j+1}, \dots, k_r\}}}^n \xi_l a'_{jl} \right), \quad j = r-1, \dots, 1.$$

Aus der Zeilenstufenform lassen sich allgemeine Eigenschaften der Matrix  $A$  ablesen:

**Satz 5.9** *Es sei  $A'$  in (5.9) eine Zeilenstufenform von  $A$ , dann gilt:*

1. *Ist die Anzahl der Stufen gleich  $m$ , dann ist*

$$Ax = b$$

*für jedes  $b \in \mathbb{R}^m$  lösbar. Die Anzahl der Stufen ist gleich  $m$  genau dann, wenn die Abbildung  $x \mapsto Ax$  surjektiv (nach  $\mathbb{R}^m$ ) ist, und dies ist äquivalenten dazu, dass  $\text{Im}A = \mathbb{R}^m$  gilt.*

2. *Besteht jede Stufe in der Breite nur aus einem Eintrag und gibt es keine Spalte, die vollständig aus Nulleinträgen besteht, dann ist*

$$Ax = b$$

*entweder eindeutig lösbar oder nicht lösbar. Diese Bedingung ist äquivalent dazu, dass die Abbildung  $x \mapsto Ax$  injektiv ist, und diese ist äquivalenten dazu, dass  $\text{Kern}A = \{0\}$  gilt.*

3. *Ist  $A$  quadratisch (also  $A \in \mathbb{R}^{n \times n}$ ), dann ist  $A$  genau dann invertierbar, wenn die Einträge  $a'_{jj}$ ,  $j = 1, \dots, n$  ungleich Null sind.*

**Folgerung 5.10** Sei ein lineares Gleichungssystem  $Ax = b$  gegeben mit  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Sei weiter  $A'$  wie (5.9) eine Zeilenstufenform zu  $A$ , und  $b' \in \mathbb{R}^m$  der Vektor, der sich durch simultane Anwendung der Zeilenumformungen aus  $b$  ergibt. Dann folgt aus Satz 5.9 das folgende:

1.  $\text{Lsg}(A, b) = \emptyset$  genau dann, wenn  $A'$  eine Zeile enthält, die vollständig aus Nulleinträgen besteht und dabei der entsprechende Eintrag in  $b'$  ungleich Null ist.
2.  $Ax = b$  ist genau dann eindeutig lösbar, wenn 1. nicht zutrifft und jede Stufe  $A'$  in der Breite nur aus einem Eintrag besteht und es keine Spalte gibt, die vollständig aus Nulleinträgen besteht.

Es gilt weiterhin:



**Satz 5.11** Für  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  und  $v \in \mathbb{R}^n$  gelte  $Av = b$ , dann folgt

$$\text{Lsg}(A, b) = \{v + u : u \in \text{Kern} A\}$$

**Beweis.** Übung. □

Mit Satz 5.9 (3) folgt, dass sich invertierbare Matrizen aus der Zeilenstufenform durch elementare Zeilenumformungen so weiter umformen lassen, dass die Einträge oberhalb der Diagonalen zu Nulleinträgen werden. Danach ist somit die Matrix so umgeformt, dass  $a'_{jk} = 0$  für  $j \neq k$  gilt. Multipliziert man anschließend jede Zeile mit dem Kehrwert des entsprechenden Diagonaleintrags (elementare Zeilenumformung Z3) so entsteht die Einheitsmatrix. Dieses beschreibt im Prinzip schon das aus dem Gauß-Verfahren abgeleitete Verfahren zur Berechnung der inversen Matrix. Dieses Verfahren wird auch **Gauß-Jordan-Verfahren** genannt. Mathematischer kann man dies wie folgt formulieren:

Zur Bestimmung der inversen Matrix zu  $A \in \mathbb{R}^{n \times n}$  kann man das folgende Gleichungssystem aufstellen

$$AX = E, \tag{5.10}$$

wobei  $E$  wie gewohnt die Einheitsmatrix ist und  $X \in \mathbb{R}^{n \times n}$  die gesuchte inverse Matrix. Gelingt es nun, wie vorher textuell beschrieben, Elementarmatrizen  $B_1, \dots, B_s \in \mathbb{R}^{n \times n}$  so zu finden, dass  $B_s B_{s-1} \cdots B_1 A = E$  gilt, so folgt mit (5.10)

$$B_s B_{s-1} \cdots B_1 AX = EX = X = B_s B_{s-1} \cdots B_1 E,$$

und damit gilt

$$A^{-1} = X.$$

Eine algorithmische Beschreibung des Gauß-Jordan-Verfahrens wird im Anhang gegeben. Für  $A \in \mathbb{R}^{n \times n}$  kann das oben beschriebene Gauß-Verfahren zur Gewinnung einer **Matrixfaktorisierung** dienlich sein. In einigen Fällen (Genaueres folgt) ist es möglich die Elementarmatrizen  $B_1, \dots, B_s$ , die zu einer Zeilenstufenform führen so zu wählen, dass

$$(B_s \cdot \dots \cdot B_2 B_1)^{-1} =: L$$

eine untere Dreiecksmatrix mit  $[L]_{jj} = 1$  für alle  $j = 1, \dots, n$  ist. Dabei ist  $L$  eine solche untere Dreiecksmatrix genau dann, wenn auch

$$B_s \cdot \dots \cdot B_2 B_1 = L^{-1}$$

eine solche ist. Bezeichnen wir mit  $R$  die Zeilenstufenform, die in diesem Fall (quadratische Matrix) eine obere Dreiecksmatrix ist, so ergibt sich durch das Gauß-Verfahren dann

$$L^{-1}A = R \Leftrightarrow$$

$$\boxed{A = LR \text{ LR-Zerlegung}} \quad (5.11)$$

Engl. **LU-decomposition** (lower-upper triangular)).

Dabei gilt folgendes:

**Satz 5.12** *Es sei  $A \in \mathbb{R}^{n \times n}$  invertierbar.*

1. *Besitzt  $A$  eine LR-Zerlung, so ist diese eindeutig.*
2.  *$A$  besitzt eine LR-Zerlung genau dann, wenn alle (Untermatrizen)  $A^{(l)} \in \mathbb{R}^{l \times l}$   $k = 1, \dots, n$ , definiert durch*

$$[A^{(l)}]_{jk} = [A]_{jk} \quad j, k = 1, \dots, l,$$

*invertierbar sind.*

## 6 Zerlegung von Vektorräumen und Orthogonalität

Es seien  $x, y, w \in \mathbb{R}^n$  und

$$x = y + w.$$

Beispielsweise könnte  $y$  ein (Nutz-)Signal,  $w$  das Störsignal oder Rauschen, und  $x$  das gemessene Signal sein. In einem solchen Fall stellt sich die Frage ob man aus der Messung  $x$  das Signal  $y$  rekonstruieren kann. Es stellt sich dabei heraus, dass ein Vorwissen der Form

$$y \in V, \quad w \in U,$$

wobei  $V, U$  zwei Unterräume sind, zur Zerlegung von  $x$  in  $y, w$  genutzt werden kann.

Da Folgende liefert das mathematische Fundament um solche und einige andere Anwendungsprobleme zu bearbeiten.

**Satz und Definition 6.1** Es seien  $V$  und  $U$  zwei Vektorräume in  $\mathbb{R}^n$ .

1. Dann ist die **Summe** von  $U$  und  $V$  definiert als

$$U + V = \{x \in \mathbb{R}^n : x = u + v, \text{ mit } v \in V, u \in U\}.$$

2. Falls  $V \cap U = \{0\}$  gilt, dann heißt die Summe dieser Räume auch **direkte Summe** und wir schreiben in diesem Fall anstelle von  $V + U$  auch

$$V \oplus U.$$

3. Sind  $U, V$  wie folgt gegeben  $V = \text{Span}(v_1, \dots, v_m) \subset \mathbb{R}^n$  und  $U = \text{Span}(u_1, \dots, u_k) \subset \mathbb{R}^n$ , so gilt

$$U + V = \text{Span}(v_1, \dots, v_m, u_1, \dots, u_k)$$

**Bemerkung 6.2** Man sieht sofort, dass die Summe von Unterräumen wieder einen Unterraum bildet.

Im Falle mehrerer Unterräume  $V_1, \dots, V_m \subset \mathbb{R}^n$  ergibt sich die Definition der Summe dieser Räume durch iteratives Anwenden der obigen Definition:

$$\sum_{k=1}^m V_k = \{v_1 + v_2 + \dots + v_m : v_k \in V_k, k = 1, \dots, m\}.$$

Falls weiter  $V_j \cap V_k = \{0\}$  für all  $j \neq k$  gilt, dann heißt diese Summe wieder direkte Summe und wir schreiben dafür

$$\bigoplus_{k=1}^m V_k.$$

**Beispiel** 1. Es seien  $V = \text{Span}((1, 0)^T)$  und  $U = \text{Span}((0, 1)^T)$ , dann gilt

$$V + U = V \oplus U = \mathbb{R}^2.$$

2. Es seien  $V, U$  Unterräume des  $\mathbb{R}^8$  gegeben durch  $V = \text{Span}(e_1, e_2, e_3)$ ,  $U = \text{Span}(e_3, e_8)$ . Dann gilt  $V \cap U \neq \{0\}$  und

$$V + U = \text{Span}(e_1, e_2, e_3, e_8).$$

Das folgende Beispiel gibt eine geometrische Interpretation der Summe von Vektorräumen.

**Beispiel** Es seien

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

Dann beschreiben  $V = \text{Span}(v_1)$  und  $U = \text{Span}(v_2)$  geometrisch Ursprungsgeraden im  $\mathbb{R}^3$ . Die (direkte) Summe  $V \oplus U$  beschreibt geometrisch die Ursprungsebene, die durch die Vektoren  $v_1, v_2$  aufgespannt wird.

**Satz 6.3** *Es seien  $V, U \subset \mathbb{R}^n$  Unterräume. Dann ist für jedes  $x \in U + V$  die Darstellung  $x = y + w$  mit  $y \in V$  und  $w \in U$  genau dann eindeutig, wenn  $V \cap U = \{0\}$  gilt, also  $V + U = V \oplus U$ .*

In der Situation zu Beginn dieses Abschnitts gibt dieser Satz somit Auskunft wann ein Signal eindeutig in Nutz- und Störsignal zerlegt werden kann.

**Beweis.** Es gelte zuerst  $V \cap U = \{0\}$ . Es sei  $x = y_1 + w_1 = y_2 + w_2$  mit  $y_1, y_2 \in V$  und  $w_1, w_2 \in U$ . Durch Umformung dieser Gleichung, und weil  $V, U$  Unterräume sind gilt

dann  $y_1 - y_2 = w_2 - w_1 \in V \cap U$ , und somit nach Voraussetzung  $y_1 - y_2 = w_2 - w_1 = 0$ . Also ist  $y_1 = y_2$  und  $w_1 = w_2$ .

Sei nun die Eindeutigkeit der Darstellung  $x = y + w$  für all  $x \in V + U$  gegeben. Für  $x \in V \cap U$  sind dann  $x = 0 + x = 1/2x + 1/2x$  mögliche Darstellungen aus zwei Vektoren aus  $V$  und  $U$ . Wegen der Eindeutigkeit der Darstellung gilt dann  $1/2x = x$  und  $1/2x = 0$ . Jede dieser Gleichungen impliziert für sich schon  $x = 0$ .  $\square$

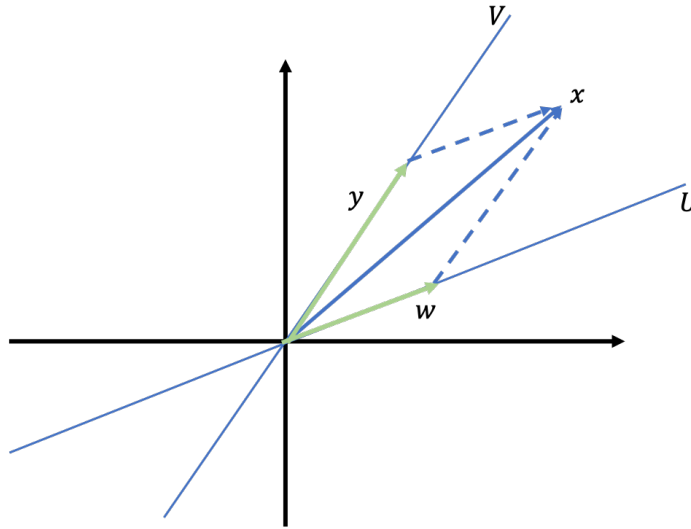


Figure 8: Eindeutige Linearkombination von  $x$  in  $V \oplus U = \mathbb{R}^2$  durch  $y \in V$  und  $w \in U$ .

In vielen Fällen können Vektorräume in zueinander orthogonal Räume zerlegt werden. In diesen Fällen ergeben sich einige schöne Vereinfachungen.

**Satz und Definition 6.4** Es sei  $V \subset \mathbb{R}^n$  ein Unterraum. Dann ist

$$V^\perp := \{u \in \mathbb{R}^n : \langle u, v \rangle = 0, \text{ für alle } v \in V\}$$

ein Untervektorraum und es gilt  $V \oplus V^\perp = \mathbb{R}^n$ . Der Raum  $V^\perp$  heißt **orthogonales Komplement** oder auch **Orthogonalraum** von  $V$ .

Nach Satz 6.3 gibt es zu jedem  $x \in \mathbb{R}^n$  eine eindeutige Darstellung  $x = v + u$  mit

$v \in V$  und  $u \in V^\perp$ . Weiterhin ist sofort ersichtlich, dass

$$(V^\perp)^\perp = V. \quad (6.12)$$

Der Beweis zu Satz 6.4 erfolgt relativ leicht mithilfe des unten eingeführten Orthogonalisierungs-Verfahren von Gram-Schmidt.

Wir bringen nun die oberen Begriffe mit den Unterräumen  $\text{Kern}(A)$ ,  $\text{Im}(A)$ , vgl. Def. 4.27 einer linearen Abbildung

$$x \mapsto Ax$$

in einen Zusammenhang. Dieser Satz kann als **Fundamentalsatz der linearen Abbildungen** betrachtet werden. Es lassen sich wesentliche Eigenschaften linearer Abbildungen daraus ableiten (vgl. Übung)

**Satz 6.5** *Es sei  $A \in \mathbb{R}^{m \times n}$ , dann gilt:*

1.  $(\text{Kern } A)^\perp = \text{Im } A^T$  (also auch  $(\text{Kern } A^T)^\perp = \text{Im } A$ ), und  $\dim(\text{Im } A) = \dim(\text{Im } A^T)$ .
2.  $n = \dim(\text{Kern } A) + \dim(\text{Im } A)$  (Dimensionsformel)
3.  $\text{rang } A = \text{rang } A^T$

**Beweis.** Nach Definition des Kerns gilt  $\text{Kern } A \subset (\text{Im } A^T)^\perp$  (\* Übungsaufgabe). Wir zeigen nun  $\text{Kern } A \supset (\text{Im } A^T)^\perp$ . Sei dazu  $v \in (\text{Im } A^T)^\perp$  gegeben. Dann gilt nach Definition von  $\text{Im } A^T$ , dass

$$0 = \langle v, A^T x \rangle = \langle Av, x \rangle \quad \forall x \in \mathbb{R}^m,$$

woraus  $Av = 0$  (\*\* Übungsaufgabe) folgt. Somit ist  $v \in \text{Kern } A$ . Es gilt also  $\text{Kern } A = (\text{Im } A^T)^\perp$ , was nach (6.12) äquivalent ist zu

$$(\text{Kern } A)^\perp = \text{Im } A^T.$$

Es sei nun  $(v_1, \dots, v_k)$  eine Basis von  $(\text{Im } A^T)^\perp$ . Wegen dem oben Gezeigten liegen die  $v_1, \dots, v_k$  damit nicht in  $\text{Kern } A$ , woraus die lineare Unabhängigkeit von  $Av_1, \dots, Av_k \in \text{Im } A$  folgt (\* \* \* Übungsaufgabe). Daher muss  $\dim(\text{Im } A) \geq \dim(\text{Im } A^T)^\perp$  gelten. Ausgehend von einer Basis von  $\text{Im } A$  zeigt man genauso  $\dim(\text{Im } A) \leq \dim(\text{Im } A^T)^\perp$  und erhält daher  $\dim(\text{Im } A) = \dim(\text{Im } A^T)^\perp$ . Damit ist (1) gezeigt.

Zu 2. Sei  $(v_1, \dots, v_k)$  eine Basis von  $\text{Im } A^T$  und  $(u_1, \dots, u_l)$  eine Basis von  $\text{Kern } A$ . Mit (1.) und (6.4) gilt  $\text{Kern } A \oplus \text{Im } A^T = \mathbb{R}^n$  und darum bildet  $(v_1, \dots, v_k, u_1, \dots, u_l)$  eine Basis des  $\mathbb{R}^n$  (\*\* \* Übungsaufgabe). Wiederum mit 1. folgt dann

$$n = l + k = \dim(\text{Kern } A) + \dim(\text{Im } A^T) = \dim(\text{Kern } A) + \dim(\text{Im } A).$$

3. folgt direkt aus  $\dim(\text{Im } A) = \dim(\text{Im } A^T)$  und Bemerkung 4.30 (3).

□

**Folgerung 6.6** Es sei  $A \in \mathbb{R}^{m \times n}$ , dann gilt:

1. Ist  $n > m$  so kann  $x \rightarrow Ax$  von  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  nicht injektiv sein. Ist  $n < m$  so kann  $x \rightarrow Ax$  von  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  nicht surjektiv sein.
2. Ist  $B \in \mathbb{R}^{m \times m}$  invertierbar, dann gilt  $\text{rang}(A) = \text{rang}(BA)$ . Insbesondere verändern elementare Zeilenumformungen (Gauß-Verfahren) den Rang nicht.
3. Die Anzahl der Zeilen in der Zeilenstufenform (vgl. 5.9) entspricht dem Rang der Matrix.

**Beispiel 6.7** Für

$$A = \begin{pmatrix} 1 & 4 \\ -3 & 4 \\ 0 & 1 \end{pmatrix}$$

ist die Abbildung  $x \rightarrow Ax$  nicht surjektiv. Mit anderen Worten

$$Ax = b$$

ist nicht für jedes  $b \in \mathbb{R}^3$  lösbar. Da die Spalten von  $A$  offensichtlich linear unabhängig sind gilt  $\dim(\text{Im } A) = 2$  und nach der Dimensionsformel (Satz 6.5 (2.)) gilt

$$2 = \dim(\text{Im } A) + \dim(\text{Kern } A),$$

woraus  $\dim(\text{Kern } A) = 0$  und damit  $\text{Kern } A = \{0\}$ , folgt. Die Abbildung  $x \mapsto A^T x$  kann umgekehrt nicht injektiv sein und es gilt mit Satz 6.5 (1.)(2.)

$$3 = \dim(\text{Im } A^T) + \dim(\text{Kern } A^T) = \dim(\text{Im } A) + \dim(\text{Kern } A^T) = 2 + \dim(\text{Kern } A^T),$$

und somit  $\dim(\text{Kern} A^T) = 1$ .

Im Folgenden werden konkrete Hilfsmittel zum Arbeiten mit orthogonalen Räumen und zur Konstruktion von orthogonalen Basen geliefert.

- Definition 6.8**
1. Ein Vektor  $x \in \mathbb{R}^n$  heißt **normiert**, wenn  $\|x\| = 1$  gilt.
  2. Eine Basis  $(b_1, \dots, b_m)$  eines (Unter-) Vektorraums  $V \subset \mathbb{R}^n$  heißt **orthogonale Basis** oder auch **Orthogonalbasis**, wenn  $\langle b_j, b_k \rangle = 0$  für alle  $j \neq k$  gilt.
  3. Eine **Orthogonalbasis**  $(b_1, \dots, b_m)$  eines Vektorraums  $V \subset \mathbb{R}^n$  heißt **Orthonormalbasis** (ONB), wenn  $\|b_j\| = 1$  für alle  $j = 1, \dots, m$  gilt.

**Satz 6.9** *Es seien  $V \subset \mathbb{R}^n$  und  $v_1, \dots, v_m \in V \setminus \{0\}$  orthogonal, dann sind diese Vektoren linear unabhängig.*

**Beweis.** Es seien  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$  und

$$0 = \sum_{k=1}^m \lambda_k v_k.$$

Dann folgt mit den Eigenschaften des Skalarprodukts 3.1 (1), (2)

$$0 = \left\langle \sum_{j=1}^m \lambda_j v_j, \sum_{k=1}^m \lambda_k v_k \right\rangle = \sum_{j=1}^m \sum_{k=1}^m \lambda_k \lambda_j \langle v_j, v_k \rangle.$$

Wegen der Orthogonalität der  $v_1, \dots, v_m$  sind die Skalarprodukte innerhalb der rechten Summe genau dann ungleich 0, wenn  $j = k$  gilt, und somit gilt dann

$$\sum_{j=1}^m \lambda_j^2 \langle v_j, v_j \rangle = 0.$$

Da nach Voraussetzung alle  $v_j \neq 0$  und somit  $\langle v_j, v_j \rangle > 0$ ,  $j = 1, \dots, m$ , folgt aus der letzten Gleichung  $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$ . Nach Definition 4.6 sind daher die  $v_1, \dots, v_m$  linear unabhängig.  $\square$



Für ONBs des  $\mathbb{R}^n$  gilt das folgende nützliche Resultat:

**Satz und Definition 6.10** Es sei  $A \in \mathbb{R}^{n \times n}$ . Dann sind folgende Aussagen äquivalent:

a)  $A$  ist invertierbar und

$$A^{-1} = A^T.$$

b) Die Spaltenvektoren von  $A$  bilden eine ONB des  $\mathbb{R}^n$ .

c) Die Zeilenvektoren von  $A$  bilden eine ONB des  $\mathbb{R}^n$ .

Wenn diese Bedingungen gelten, heißt  $A$  **Orthonormalmatrix**.

**Beweis.** Der Beweis folgt leicht durch Betrachtung des Matrix-Matrix-Produkts und kann als Übung nachvollzogen werden.  $\square$

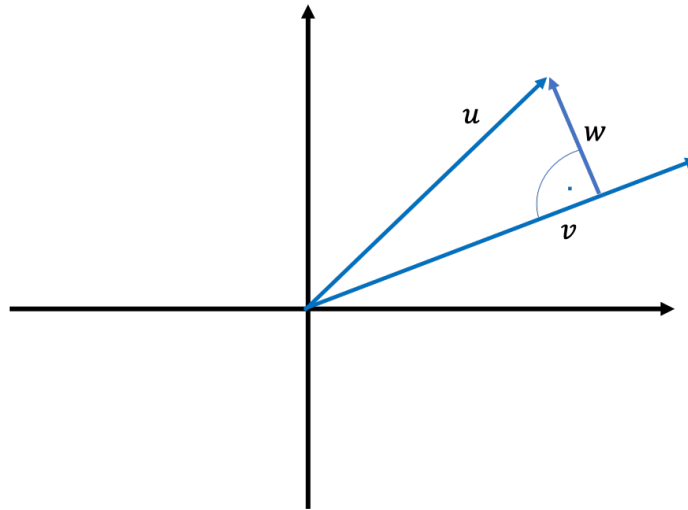


Figure 9: Orthogonale Projektion auf eine Gerade

Abbildung 6 illustriert am Beispiel des  $\mathbb{R}^2$  das Folgende: Seien  $v, u \in \mathbb{R}^n$ , dann kann  $u$  dargestellt werden als

$$u = w + \lambda v$$

mit  $w$  orthogonal zu  $v$  und  $\lambda \in \mathbb{R}$  ein Skalar. Wie in der oberen Graphik dargestellt,

wird dazu  $u$  auf die von  $v$  definierte Ursprungsgerade projiziert.

**Lemma 6.11** *Es seien  $v, u \in \mathbb{R}^n$ ,  $v \neq 0$  zwei Vektoren, dann ist*

$$w := u - \frac{\langle v, u \rangle}{\langle v, v \rangle} v$$

*orthogonal zu  $v$ . Durch Umformung ergibt sich:*

$$u = w + \frac{\langle v, u \rangle}{\langle v, v \rangle} v.$$

**Beweis.** Es gilt

$$\left\langle v, u - \frac{\langle v, u \rangle}{\langle v, v \rangle} v \right\rangle = \langle v, u \rangle - \langle v, v \rangle \frac{\langle v, u \rangle}{\langle v, v \rangle} = 0,$$

und damit folgt die Behauptung.  $\square$

Diese Idee der orthogonalen Zerlegung wird nun auf den Fall mehrerer Vektoren, bzw. allgemeiner Unterräume erweitert.

Falls in Lemma 6.11  $u$  und  $v$  linear unabhängig sind, ist  $w \neq 0$  und damit  $w, v$  orthogonal und linear unabhängig. Dieses Prinzip lässt sich iterativ auf eine Familie linear unabhängiger Vektoren anwenden:

### Gram-Schmidtsches Orthogonalisierungsverfahren

Es sei  $(w_1, \dots, w_k)$  eine Basis eines Vektorraums  $V \subset \mathbb{R}^n$ . Das Gram-Schmidt-Verfahren generiert daraus wie folgt eine Orthogonalbasis (von  $V$ ):

$$\begin{aligned} v_1 &:= w_1 \\ v_2 &:= w_2 - \frac{\langle v_1, w_2 \rangle}{\langle v_1, v_1 \rangle} v_1 \\ v_3 &:= w_3 - \frac{\langle v_1, w_3 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle v_2, w_3 \rangle}{\langle v_2, v_2 \rangle} v_2 \\ &\vdots \\ v_k &:= w_k - \sum_{j=1}^{k-1} \frac{\langle v_j, w_k \rangle}{\langle v_j, v_j \rangle} v_j. \end{aligned} \tag{6.13}$$

Durch iterative Anwendung von Lemma 6.11 sieht man, dass die Vektoren

$v_1, \dots, v_k$  paarweise orthogonal sind. Durch Normierung

$$u_j := \frac{v_j}{\|v_j\|} \quad j = 1, \dots, k$$

ergibt sich die ONB  $(u_1, \dots, u_k)$  von  $V$ .

Aus dem Gram-Schmidt-Verfahren ergibt sich insbesondere, dass es zu jedem Vektorraum  $V \subset \mathbb{R}^n$  eine ONB gibt.

Sei weiter  $(v_1, \dots, v_m)$  eine Basis des Unterraums  $V \subset \mathbb{R}^n$ , dann gibt es Vektoren  $(w_1, \dots, w_{n-m})$  so, dass  $(v_1, \dots, v_m, w_1, \dots, w_{n-m})$  eine Basis des  $\mathbb{R}^n$  ist. Wenden wir nun das Gram-Schmidt-Verfahren auf diese Basis an, so erhalten wir eine ONB  $(b_1, \dots, b_n)$  des  $\mathbb{R}^n$ . Vorausgesetzt, dass Orthogonalisierung in der vorgesehenen Reihenfolge durchgeführt wurde, gilt dann:

1.  $(b_1, \dots, b_m)$  ist eine ONB von  $V$ .
2. Für  $U = \text{Span}(b_{m+1}, \dots, b_n)$  gilt  $V \cap U = \{0\}$  und für alle  $v \in V$  und  $u \in U$  gilt  $\langle v, u \rangle = 0$ .
3. Sei  $w \in \mathbb{R}^n$  so, dass  $\langle w, v \rangle = 0$  für alle  $v \in V$ , dann folgt  $w \in U$ .

Damit ergibt sich folgendes:

**Definition 6.12** Es sei  $V \subset \mathbb{R}^n$  ein Unterraum. Für  $x \in \mathbb{R}^n$  mit  $x = y + w$  mit  $y \in V$  und  $w \in V^\perp$  ( $y, w$  sind eindeutig Satz 6.4 und Satz 6.3) setzen wir

$$P_V(x) = y.$$

Die dadurch definierte Abbildung  $P_V : \mathbb{R}^n \rightarrow V$  ist linear und heißt **orthogonale Projektion** auf  $V$ .

Orthogonale Projektionen kommen in der Praxis an vielen Stellen vor. Betrachten wir dazu zum Beispiel die Situation zu Beginn des Abschnitts noch einmal. Hier war ein Signal  $x = y + w$  mit Nutzsignal  $y$  und Störsignal  $w$ . Ist nun bekannt, dass  $y \in V$  und  $w \in V^\perp$  für einen Unterraum  $V$  gilt, dann kann mittels  $y = P_V(x)$  das Signal entrauscht werden.

Im Fall, dass eine ONB des Raumes  $V$  bekannt ist, lässt sich  $P_V$  sehr einfach berechnen:

**Satz 6.13** 1. Es sei  $\mathcal{B} = (v_1, v_2, \dots, v_k)$  eine ONB eines Vektorraums  $V \subset \mathbb{R}^n$  und  $U := (v_1 v_2 \dots v_k) \in \mathbb{R}^{n \times k}$  (spaltenweise Konkatenieren). Dann gilt für jedes  $x \in \mathbb{R}^n$

$$P_V(x) = \sum_{j=1}^k \langle v_j, x \rangle v_j = UU^T x.$$

2. Weiter gilt für alle  $v \in V$ ,  $v \neq P_V(x)$ :

$$\|x - P_V(x)\| < \|x - v\|.$$

Es gilt somit in Satz 6.13(1.) für alle  $x \in V$ :

$$x = \begin{pmatrix} \langle v_1, x \rangle \\ \langle v_2, x \rangle \\ \vdots \\ \langle v_k, x \rangle \end{pmatrix}_{\mathcal{B}}.$$

Und Satz 6.13 (2.) impliziert

$$\|x - P_V(x)\| = \min_{v \in V} \|x - v\|.$$

**Beweis.** Zu (1.) Es sei  $(u_1, \dots, u_{n-k})$  eine ONB von  $V^\perp$ . Da  $(v_1, \dots, v_k)$  und  $(u_1, \dots, u_{n-k})$  ONBs von  $V$  und  $V^\perp$  ist, ist  $(v_1, \dots, v_k, u_1, \dots, u_{n-k})$  ein ONB des  $\mathbb{R}^n$ . Für

$$A = (v_1, \dots, v_k, u_1, \dots, u_{n-k}) \in \mathbb{R}^{n \times n}$$

gilt dann nach Satz 6.10  $A^{-1} = A^T$ . Für die Berechnung der Koordinaten ist folgendes zu lösen

$$x = \lambda_1 v_1 + \dots + \lambda_k v_k + \alpha_1 u_1 + \dots + \alpha_{n-k} u_{n-k}$$

Dies ist äquivalent ist zu

$$x = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \alpha_1 \\ \vdots \\ \alpha_{n-k} \end{pmatrix} \Leftrightarrow A^T x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \alpha_1 \\ \vdots \\ \alpha_{n-k} \end{pmatrix}.$$

Hieraus lässt sich sofort  $\lambda_j = \langle v_j, x \rangle$  für  $j = 1, \dots, k$  ablesen.

Zu (2.): Seien  $(v_1, \dots, v_k)$  und  $(u_1, \dots, u_{n-k})$  wie oben. Es gilt aufgrund der Eindeutigkeit der Zerlegung von  $x$ , dass

$$x - P_V(x) = w \in V^\perp.$$

Da  $V$  ein Unterraum ist gilt weiter für jedes  $v \in V$  dass  $v - P_V(x) \in V$ , und somit folgt  $\langle v - P_V(x), x - P_V(x) \rangle = 0$ . Damit ergibt sich

$$\begin{aligned} \|x - v\|^2 &= \|x - v + P_V(x) - P_V(x)\|^2 = \langle x - v + P_V(x) - P_V(x), x - v + P_V(x) - P_V(x) \rangle \\ &= \|x - P_V(x)\|^2 + \|v - P_V(x)\|^2 + 2\langle v - P_V(x), x - P_V(x) \rangle \\ &\geq \|x - P_V(x)\|^2, \end{aligned}$$

wobei im letzten Schritt ' $>$ ' gilt falls  $v \neq P_V(x)$ . □

Das Gram-Schmidt-Verfahren liefert eine weitere Matrixfaktorisierung, die sogenannte **QR-Zerlegung**. Sei im Gram-Schmidt-Verfahren dazu  $k = n$ , die Basen beziehen sich also auf den gesamten  $\mathbb{R}^n$ . In (6.13) ist,

$$v_l := w_l - \sum_{j=1}^{l-1} \frac{\langle v_j, w_l \rangle}{\langle v_j, v_j \rangle} v_j, \quad l = 1, \dots, n.$$

und man beobachtet hier, dass die orthogonalen Vektoren  $v_l$  jeweils als Linearkombination von  $w_l$  und  $v_1, \dots, v_{l-1}$  berechnet werden. Sukzessives Einsetzen von  $v_1$  ausgehend liefert daher, dass  $v_l$  als Linearkombination der Form

$$v_l = \sum_{j=1}^l r_{lj}^- w_j$$

geschrieben werden kann, wobei entsprechend (6.13)  $r_{jj}^- = 1$ ,  $j = 1, \dots, n$  gilt. Setzen wir weiter  $r_{kj}^- = 0$  für  $k > j$ , dann ergibt sich

$$(v_1 \dots v_n) = (w_1 \dots w_n) R^- \tag{6.14}$$

mit

$$R^- := \begin{pmatrix} r_{11}^- & r_{12}^- & r_{13}^- & \dots & r_{1n}^- \\ 0 & r_{22}^- & r_{23}^- & \ddots & r_{2n}^- \\ 0 & 0 & r_{33}^- & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_{nn}^- \end{pmatrix}. \tag{6.15}$$

Dabei ist  $R^-$  invertierbar. Weiterhin kann die Normalisierung im Gram-Schmidt-Verfahren durch eine invertierbare Diagonalmatrix  $D$  implementiert werden sodass sich aus 6.14 die ONB  $(u_1, \dots, u_n)$  ergibt als

$$(u_1 \dots u_n) = (w_1 \dots w_n) R^- D \quad (6.16)$$

Sei nun  $A \in \mathbb{R}^{n \times n}$  invertierbar, dann bilden die Spalten eine Basis des  $\mathbb{R}^n$ . Identifizieren wir nun  $A = (w_1 \dots w_n)$ ,  $Q = (u_1 \dots u_n)$  (**Orthonormalmatrix**) und  $R = (R^- D)^{-1}$  (**obere Dreiecksmatrix**) so folgt aus 6.16

$$\boxed{A = QR \text{ QR-Zerlegung}} \quad (6.17)$$

Durch leichte Modifikation des Oberen lässt sich zeigen, dass eine QR-Zerlegung auch für  $A \in \mathbb{R}^{m \times n}$  möglich ist. Dabei ist dann  $Q$  allgemein nicht mehr quadratisch besteht aber weiterhin aus orthonormalen Spalten.

## 7 Kleinste-Quadrate-Problem und Lineare Regression

Sei  $A \in \mathbb{R}^{m \times n}$  eine Matrix und  $b \in \mathbb{R}^m$ . Falls das LGS

$$Ax = b$$

keine Lösung besitzt, sucht man ein  $x \in \mathbb{R}^n$ , sodass  $Ax$  möglichst nahe an  $b$  liegt. Dies führt auf das folgende **Optimierungsproblem**:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|. \quad (7.18)$$

Dieses **lineare kleinste-Quadrate-Problem (linear least squares)** tritt insbesondere in der linearen Regression auf, wo eine Funktion  $b \approx Ax$  durch Wahl von  $x$  optimiert wird. Geometrisch entspricht dies der Suche nach  $y \in \text{Im}A$ , sodass  $\|y - b\|$  minimal ist. Die Lösung ergibt sich aus der orthogonalen Projektion (siehe Abbildung 6).

**Satz 7.1** Sei  $A \in \mathbb{R}^{m \times n}$  mit  $m \geq n$ ,  $b \in \mathbb{R}^m$  und  $\text{rang}(A) = n$ . Dann ist  $A^T A$  invertierbar, und die Lösung von (7.18) ist

$$x^* = (A^T A)^{-1} A^T b.$$

Weiter gilt für die Projektion von  $b$  auf  $\text{Im}A$ :

$$P_{\text{Im}A}(b) = A(A^T A)^{-1} A^T b.$$

**Beweis.** Aus  $\text{rang}(A) = n$  folgt die Invertierbarkeit von  $A^T A$  (Übung). Zudem gilt nach Satz 6.13 (2.):

$$\min_{y \in \text{Im}A} \|b - y\| = \min_{x \in \mathbb{R}^n} \|b - Ax\|.$$

Zu zeigen bleibt noch  $P_{\text{Im}A}(b) = A(A^T A)^{-1} A^T b$ . Sei  $v = Au \in \text{Im}A$  mit  $u \in \mathbb{R}^n$ . Dann folgt

$$\begin{aligned} \langle v, b - A(A^T A)^{-1} A^T b \rangle &= \langle Au, b - A(A^T A)^{-1} A^T b \rangle \\ &= \langle u, A^T b - A^T A(A^T A)^{-1} A^T b \rangle \\ &= \langle u, A^T b - A^T b \rangle \\ &= 0. \end{aligned}$$

Daraus folgt  $b - A(A^T A)^{-1} A^T b \in (\text{Im}A)^\perp$  und somit  $P_{\text{Im}A}(b) = A(A^T A)^{-1} A^T b$ .  $\square$

### Anwendung: Lineare Regression

Die **lineare Regression** beschreibt den Zusammenhang zwischen einer abhängigen Variablen  $y$  und mehreren unabhängigen Variablen. Sei  $y \in \mathbb{R}^m$  ein Vektor mit gemessenen Werten und  $X \in \mathbb{R}^{m \times n}$  eine Matrix mit erklärenden Variablen (Features). Das Modell lautet:

$$\mathbf{y} \approx X\beta$$

mit unbekanntem Parametervektor  $\beta \in \mathbb{R}^n$ . Die Schätzung von  $\beta$  ergibt sich aus der Lösung des kleinsten-Quadrate-Problems:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}.$$

**Beispiel 7.2 (Einfache lineare Regression)** Ein Forscher untersucht den Einfluss der Temperatur  $T$  (in °C) auf den Eisverkauf  $S$  (in tausend Stück) anhand folgender Daten:

$$X = \begin{pmatrix} 1 & 15 \\ 1 & 20 \\ 1 & 25 \\ 1 & 30 \\ 1 & 35 \end{pmatrix}, \quad y = \begin{pmatrix} 40 \\ 50 \\ 65 \\ 80 \\ 95 \end{pmatrix}.$$

Hierbei ist die erste Spalte von  $X$  eine Eins-Spalte für den Achsenabschnitt. Die Schätzung liefert die Regressionsgerade:

$$S(T) \approx 5T - 35.$$

Das Modell sagt voraus, dass ein Temperaturanstieg um 1°C den Eisverkauf um 5.000 Stück erhöht.

**Beispiel 7.3 (Multiple lineare Regression)** Betrachtet man neben der Temperatur  $T$  auch den Werbeaufwand  $W$  (in 1.000€), ergibt sich das Modell:

$$S(T, W) = \beta_0 + \beta_1 T + \beta_2 W.$$



Mit den Daten

$$X = \begin{pmatrix} 1 & 15 & 2 \\ 1 & 20 & 3 \\ 1 & 25 & 5 \\ 1 & 30 & 7 \\ 1 & 35 & 9 \end{pmatrix}, \quad y = \begin{pmatrix} 40 \\ 50 \\ 65 \\ 80 \\ 95 \end{pmatrix}$$

erhält man durch  $\hat{\beta} = (X^T X)^{-1} X^T y$  die Schätzung:

$$S(T, W) \approx 3T + 4W - 20.$$

Dies bedeutet, dass eine Erhöhung der Temperatur um 1°C den Eisverkauf um 3.000 Stück steigert, während 1.000€ mehr Werbeausgaben zusätzlich 4.000 Stück mehr bringen.

## 8 Erweiterung auf komplexe Vektorräume

Die bisherigen Definitionen und Resultate lassen sich (fast) vollständig auf Vektoren und Matrizen mit Einträgen in  $\mathbb{C}$  übertragen.

In vielen Fällen werden in der Praxis komplexe Zahlen im Zusammenhang mit den Verfahren der linearen Algebra benötigt. Wir führen daher das Folgende ein:

Wir meinen mit  $\mathbb{K}$  entweder wie bisher die reellen Zahlen oder die komplexen Zahlen, also

$$\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}.$$

Die Addition von Vektoren, das Matrix-Vektor-Produkt, und das Matrix-Matrix-Produkt sind auf  $\mathbb{K}^n$  weiterhin genauso definiert wie im  $\mathbb{R}^n$  vgl. Definitionen 1.1, 2.4, 2.5, 2.12.

Weiterhin gelten sämtliche Rechenregeln und nachfolgenden Definitionen aus den ersten beiden Kapiteln. Das Folgende führt das Rechnen im  $\mathbb{C}^2$  beispielhaft vor.

**Beispiel 8.1** 1.

$$(-1 + i2) \begin{pmatrix} 3 \\ i \end{pmatrix} = \begin{pmatrix} -3 + i6 \\ -2 - i \end{pmatrix}.$$

2.

$$\begin{pmatrix} 1 \\ 1 + i \end{pmatrix} + \begin{pmatrix} 4 + i3 \\ i2 \end{pmatrix} = \begin{pmatrix} 5 + i3 \\ 1 + i3 \end{pmatrix}.$$

3.

$$\begin{pmatrix} 1 & 2 + i \\ 1 + i & 4 - i \end{pmatrix} + \begin{pmatrix} i & -2 \\ 4 & i7 \end{pmatrix} = \begin{pmatrix} 1 + i & i \\ 5 + i & 4 + i6 \end{pmatrix}.$$

4.

$$\begin{pmatrix} i2 & 0 \\ 0 & 3 - i \end{pmatrix} \cdot \begin{pmatrix} i & -2 \\ 4 & i7 \end{pmatrix} = \begin{pmatrix} -2 & -i4 \\ 12 - 4i & 7 - i21 \end{pmatrix}.$$

Aus Gründen, die nachfolgend erläutert werden wird der Begriff der transponierten Matrix und der symmetrischen Matrix erweitert

**Definition 8.2** 1. Sei  $A \in \mathbb{K}^{m \times n}$ , dann ist die **komplex konjugierte Matrix**  $\bar{A} \in \mathbb{K}^{m \times n}$  definiert durch

$$[\bar{A}]_{jk} = \overline{[A]_{jk}}, \quad j = 1, \dots, m, \quad k = 1, \dots, n.$$

(für  $A \in \mathbb{R}^{m \times n}$  ist  $\bar{A} = A$ )

2. Zu  $A \in \mathbb{K}^{m \times n}$  heißt  $A^H := \bar{A}^T = \overline{A^T}$  die **adjungierte Matrix** von  $A$  (für  $A \in \mathbb{R}^{m \times n}$  ist  $A^H = A^T$ ).

3.  $A \in \mathbb{K}^{n \times n}$  heißt **hermitesch**, falls

$$A = A^H$$

gilt (für  $A \in \mathbb{R}^{n \times n}$  ist hermitesch äquivalent mit symmetrisch).

4. Eine invertierbare Matrix  $A \in \mathbb{K}^{n \times n}$  heißt **unitär**, falls

$$A^{-1} = A^H$$

gilt (für  $A \in \mathbb{R}^{n \times n}$  ist unitär äquivalent mit orthonormal).

**Beispiel 8.3** Die Matrix zur sogenannten diskreten Fourier-Transformation,  $W \in \mathbb{C}^{N \times N}$  ist gegeben durch

$$W = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & e^{-2\pi i 1/N} & e^{-2\pi i 2/N} & \dots & e^{-2\pi i (N-1)/N} \\ 1 & e^{-2\pi i 2/N} & e^{-2\pi i 4/N} & \dots & e^{-2\pi i 2(N-1)/N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-2\pi i (N-1)/N} & e^{-2\pi i (N-1)2/N} & \dots & e^{-2\pi i (N-1)^2/N} \end{pmatrix}, \quad (8.19)$$

oder, äquivalent,

$$[W]_{j,k} = \frac{1}{\sqrt{N}} e^{-2\pi i (j-1)(k-1)/N}, \quad j, k = 1, \dots, N.$$

Diese Matrix ist unitär, es gilt also  $W^{-1} = W^H$ . Dabei liefert  $W^H$  die inverse diskrete Fourier-Transformation und es gilt

$$[W^H]_{j,k} = \frac{1}{\sqrt{N}} e^{2\pi i (j-1)(k-1)/N}, \quad j, k = 1, \dots, N.$$

An anderer Stelle definiert man  $W$  teilweise auch als  $D = \sqrt{N}W$ , wodurch dann  $D^{-1} = 1/ND^H$  gilt. In dieser Konvention ist die diskrete Fourier-Transformation nicht unitär.

Es soll weiterhin für  $v \in \mathbb{K}^n$  durch

$$\sqrt{\langle v, v \rangle}$$

eine *Länge* von  $v$  beschrieben sein. Dazu muss die Definition des Skalarprodukts wie folgt erweitert werden. Für reelle Vektoren ergibt sich dabei keine Änderung.

**Definition 8.4** Eine Abbildung  $\langle \cdot, \cdot \rangle : \mathbb{K}^n \rightarrow \mathbb{K}$  heißt **Skalarprodukt**, wenn für  $x, y, z \in \mathbb{K}^n$  und  $\lambda \in \mathbb{K}$  folgendes gilt:

1.  $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ ,
2.  $\langle x, y \rangle = \overline{\langle y, x \rangle}$ ,
3.  $\langle x, \lambda y \rangle = \lambda \langle x, y \rangle$  und  $\langle \lambda x, y \rangle = \bar{\lambda} \langle x, y \rangle$ ,
4.  $\langle x, x \rangle \geq 0$ , wobei  $\langle x, x \rangle = 0$  genau dann gilt, wenn  $x = 0$ .

**Satz und Definition 8.5** Für  $x, y \in \mathbb{K}^n$  setzen wir

$$\langle x, y \rangle := \sum_{k=1}^n \overline{x_k} y_k.$$

Diese Abbildung  $\langle \cdot, \cdot \rangle : \mathbb{K}^n \times \mathbb{K}^n \rightarrow \mathbb{K}$  erfüllt die Eigenschaften aus Definition 8.4 und heißt **kanonisches Skalarprodukt** oder auch kurz Skalarprodukt auf  $\mathbb{K}^n$ . Zwei Vektoren  $x, y \in \mathbb{K}^n$  heißen **orthogonal**, wenn gilt

$$\langle x, y \rangle = 0.$$

Es gilt

Mit dieser Erweiterung der Definition des Skalarprodukts ergibt sich folgende Änderung in Satz 3.4 für den allgemeinen Fall in  $\mathbb{K}^n$ .

**Satz 8.6** *Es seien  $A \in \mathbb{K}^{m \times n}$ ,  $x \in \mathbb{K}^n$  und  $y \in \mathbb{K}^m$ , dann gilt*

$$\langle y, Ax \rangle = \langle A^H y, x \rangle.$$

Die Definition der Norm in (3.5) mit ihren Eigenschaften in Satz 3.6 und die Cauchy-Schwarze Ungleichung in Satz 3.7 bleiben unverändert erhalten.

Im Allgemeinen gilt für  $x, y \in \mathbb{C}^n$  jedoch  $\langle x, y \rangle \notin \mathbb{R}$ . Somit ist in dem allgemeinen Fall kein Winkel zwischen  $x, y$  wie in Definition 3.8 definiert. Daher lassen sich Hyperebenen und Halbräume, die auf der Größenrelation der Ergebnisse von Skalarprodukten beruhen, nicht ohne Weiteres auf den Fall  $\mathbb{C}^n$  verallgemeinern.

Die Definition und Resultate aus Kapitel 4, Kapitel 5, Kapitel 6 bleiben für  $\mathbb{K}^n$  erhalten.

## 9 Determinante

Ein wichtiges Hilfsmittel, vor allem zur theoretischen Analyse, innerhalb der linearen Algebra ist die Determinante. Die Definition der Determinante wird zu Anfang etwas undurchsichtig und erklärungsbedürftig erscheinen. Die geometrische Interpretation im späteren Teil liefert jedoch zumindest für  $2 \times 2$  und  $3 \times 3$  Matrizen einen einleuchtenden Zugang zu diesem Thema.

**Definition 9.1** Eine bijektive Abbildung

$$\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}, \quad (n \in \mathbb{N})$$

heißt **Permutation** (von  $\{1, \dots, n\}$ ). Weiter setzen wir

$$S_n := \{\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}, \sigma \text{ ist bijektiv}\},$$

die Menge aller Permutationen (von  $\{1, \dots, n\}$ ).

Für  $\sigma \in S_n$  hat sich die folgende definierende Schreibweise etabliert:

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}.$$

**Definition 9.2** Ein **Fehlstand** einer Permutation  $\sigma \in S_n$  ist ein Paar  $(j, k)$  mit  $1 \leq j < k \leq n$  so, dass  $\sigma(j) > \sigma(k)$  gilt. Das **Vorzeichen** oder **Signum** einer Permutation  $\sigma \in S_n$  ist definiert durch

$$\text{sign}(\sigma) := \begin{cases} 1, & \text{Anzahl der Fehlstände von } \sigma \text{ ist gerade} \\ -1, & \text{Anzahl der Fehlstände von } \sigma \text{ ist ungerade} \end{cases}.$$

**Beispiel 9.3** 1. Für

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

ist  $\text{sign}(\sigma) = -1$

2. Für

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix}$$

ist  $\text{sign}(\sigma) = 1$

Hiermit lässt sich nun die Determinante wie folgt definieren:

**Definition 9.4** Es sei  $A \in \mathbb{K}^{n \times n}$  mit Einträgen  $a_{jk} := [A]_{jk}$ ,  $j, k = 1, \dots, n$ , dann ist die **Determinante** von  $A$  definiert durch

$$\det(A) := \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \dots \cdot a_{n\sigma(n)}.$$

Die obere Formel heißt **Leibniz-Formel**.

Für die Fälle  $n = 2$  und  $n = 3$  ergeben sich folgende Formeln zur Berechnung der Determinante.

**Bemerkung 9.5** Für  $n = 2$  gilt

$$\begin{aligned} \det A &= \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned}$$

Weiter gilt für  $n = 3$ :

$$\begin{aligned} \det(A) &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}. \end{aligned}$$

Für nicht-quadratische Matrizen ist die Determinante nicht definiert. Das folgende Resultat liefert wichtige Eigenschaften der Abbildung  $\det : \mathbb{K}^{n \times n} \rightarrow \mathbb{K}$ .

**Satz 9.6** Für Determinante gilt:

1)  $\det$  ist linear in jeder Zeile, d. h. für  $A = \begin{pmatrix} A_{1Z} \\ \vdots \\ A_{jZ} \\ \vdots \\ A_{nZ} \end{pmatrix} \in \mathbb{K}^{n \times n}$ , mit  $A_{jZ} \in \mathbb{K}^{1 \times n}$ ,  $j = 1, \dots, n$ , die Zeilenvektoren, und  $j \in \{1, \dots, n\}$  gilt für jedes  $b \in \mathbb{K}^{1 \times n}$

$$\det \begin{pmatrix} A_{1Z} \\ \vdots \\ A_{jZ} + b \\ \vdots \\ A_{nZ} \end{pmatrix} = \det(A) + \det \begin{pmatrix} A_{1Z} \\ \vdots \\ b \\ \vdots \\ A_{nZ} \end{pmatrix},$$

und für jedes  $\lambda \in \mathbb{K}$ :

$$\det \begin{pmatrix} A_{1Z} \\ \vdots \\ \lambda A_{jZ} \\ \vdots \\ A_{nZ} \end{pmatrix} = \lambda \det(A).$$

2)  $\det(A) = 0$ , falls zwei Zeilen von  $A$  übereinstimmen,

3)  $\det(E) = 1$ .

Eine sehr anschauliche Interpretation der Aussagen aus Satz 9.6 liefert die folgende Bemerkung

Für zwei Vektoren  $v, u \in \mathbb{R}^2$  definiert

$$P = \{x = \lambda_1 v + \lambda_2 u \in \mathbb{R}^2 : \lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 \leq 1\}$$

das von  $v, u$  aufgespannte Parallelogramm. Für drei Vektoren  $v, u, w \in \mathbb{R}^3$  definiert

$$S = \{x = \lambda_1 v + \lambda_2 u + \lambda_3 w \in \mathbb{R}^3 : \lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 + \lambda_3 \leq 1\}$$

den von  $v, u, w$  aufgespannten Spat.



**Bemerkung 9.7** 1. Es seien  $v, u \in \mathbb{R}^2$ , dann gilt für  $A \in \mathbb{R}^{2 \times 2}$ , bestehend aus Zeilenvektoren  $u, v$ , und das von  $u, v$  aufgespannte Parallelogramm  $P$

$$|\det(A)| = \text{Flächeninhalt von } P.$$

2. Seien  $v, u, w \in \mathbb{R}^3$ , dann gilt für  $A \in \mathbb{R}^{3 \times 3}$ , bestehend aus Zeilenvektoren  $u, v, w$ , und den von  $u, v, w$  aufgespannten Spat  $S$

$$|\det(A)| = \text{Volumen von } S.$$

3. Analoges gilt für Volumina in höheren Dimensionen.  
Dies ist in der folgenden Graphik illustriert.

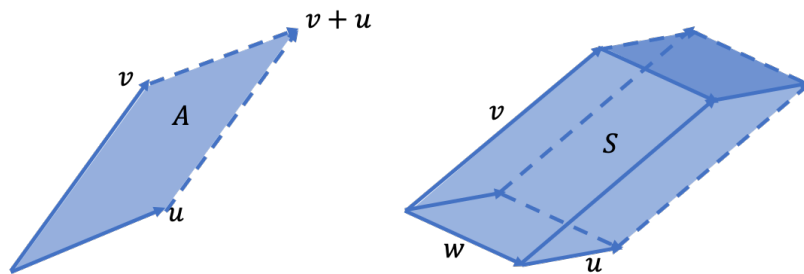


Figure 10: Parallelogramm im  $\mathbb{R}^2$  und Spat im  $\mathbb{R}^3$ .

Die Determinante wird auch zur Berechnung höherdimensionaler Volumina verwendet.

Die Eigenschaft aus Satz 9.6 (2) ergibt sich graphisch wie folgt:

Die Eigenschaft aus Satz 9.6 (1) ergibt sich graphisch wie folgt:

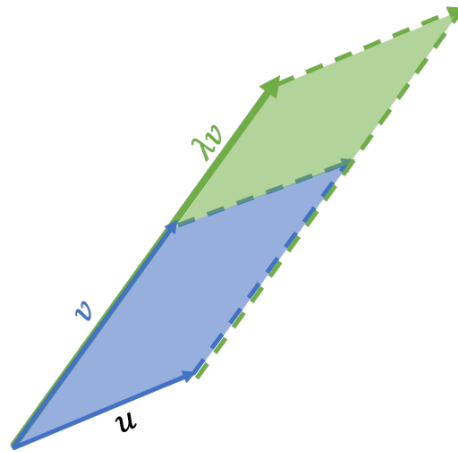
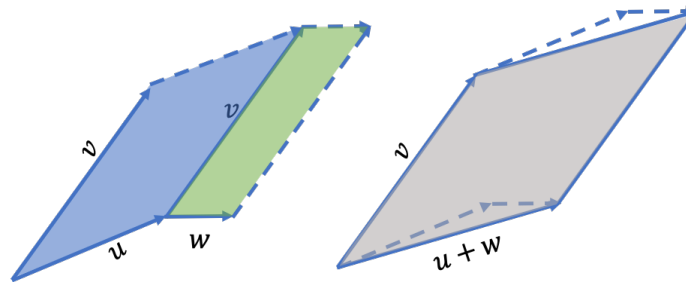


Figure 11: Zeilenweise Homogenität der Determinante

Figure 12: Zeilenweise Additivität der Determinante: Der Flächeninhalt der linken und rechten Figur im  $\mathbb{R}^2$  sind identisch.

Eine leichter eingängige Methode zur Berechnung der Determinante liefert der Laplace'sche Entwicklungssatz. Zur Formulierung dieses Satzes definieren wir Folgendes.

**Definition 9.8** Es seien  $A \in \mathbb{K}^{n \times n}$ , wobei  $n > 1$ , und  $j, k \in \{1, \dots, n\}$ . Die Matrix  $S_{jk}(A) \in \mathbb{K}^{(n-1) \times (n-1)}$ , die aus  $A$  durch Streichen der  $j$ -ten Zeile und der  $k$ -ten Spalte entsteht heißt *Streichungsmatrix* (bzgl.  $(j, k)$ ) von  $A$ .

Ist etwa

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

so ist

$$S_{23}(A) = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

**Satz 9.9 (Laplace'scher Entwicklungssatz)** Es sei  $A \in \mathbb{K}^{n \times n}$  mit Einträgen  $(a_{jk})$ ,  $j, k \in \{1, \dots, n\}$ . Dann gilt für  $k = 1, \dots, n$

$$\det A = \sum_{j=1}^n a_{jk} (-1)^{j+k} \det(S_{jk}(A)) \quad (9.20)$$

(“Entwicklung nach der  $k$ -ten Spalte”) und für  $j = 1, \dots, n$

$$\det A = \sum_{k=1}^n a_{jk} (-1)^{j+k} \det(S_{jk}(A)) \quad (9.21)$$

(“Entwicklung nach der  $j$ -ten Zeile”).

Der folgende Satz fasst wesentliche Eigenschaften der Determinante zusammen

**Satz 9.10** Es sei  $A \in \mathbb{K}^{n \times n}$  mit Einträgen  $(a_{jk})$ ,  $j, k = 1, \dots, n$ , dann gilt:

1.  $A$  ist invertierbar genau dann, wenn  $\det(A) \neq 0$ ,
2.  $\det(A) = \det(A^T)$ ,
3.  $\det(AB) = \det(A) \det(B)$  (mit  $B \in \mathbb{K}^{n \times n}$ ).

Mit diesen Eigenschaften lassen sich relativ leicht die Determinanten der Elementarmatrizen herleiten.

Für alle  $j \neq k$  ist

$$\det(E_{Z1}(j, k)) = -1.$$

Für alle  $\lambda \in \mathbb{R}$  und  $j \neq k$  ist

$$\det(E_{Z2}(\lambda, j, k)) = 1.$$

Für alle  $\lambda \in \mathbb{R} \setminus \{0\}$  und  $j$  ist

$$\det(E_{Z3}(\lambda, j)) = \lambda.$$

Im Allgemeinen ist die Berechnung der Determinante aufwändig. In manchen Fällen lassen sich die Eigenwerte jedoch sofort ablesen :

Ähnlich zu den Bezeichnungen in 2.1, 2.2, 2.3, heißt eine Matrix  $B \in \mathbb{K}^{n \times n}$  der Form

$$C = \begin{pmatrix} & & & 0 & \dots & & \\ & B_1 & & \vdots & \ddots & \vdots & \dots \\ & & & 0 & \dots & 0 & \\ 0 & \dots & 0 & & B_2 & & 0 & \dots \\ \vdots & \ddots & \vdots & & & \vdots & \ddots & \dots \\ & \dots & 0 & & & 0 & \dots & \\ & & & 0 & \dots & 0 & & \\ \vdots & & & \vdots & & \vdots & \ddots & \vdots \\ & & & & & & 0 & \dots & 0 \\ & & & & & \dots & 0 & & \\ & & & & & \dots & \ddots & \vdots & \\ & & & & & & \dots & 0 & B_k \\ & & & & & & & & \end{pmatrix} \quad (9.22)$$

mit  $B_j \in \mathbb{K}^{n_j \times n_j}$ ,  $n = n_1 + n_2 + \dots + n_k$  heißt **Blockmatrix** (mit Blöcken  $B_1, \dots, B_k$ )

**Satz 9.11** 1. Es sei  $A \in \mathbb{K}^{n \times n}$  eine untere- oder obere Dreiecksmatrix, dann gilt

$$\det(A) = \prod_{j=1}^n [A]_{jj}.$$

2. Es sei  $B \in \mathbb{K}^{n \times n}$  eine Blockmatrix mit Blöcken  $B_j \in \mathbb{C}^{n_j \times n_j}$ ,  $j = 1, \dots, k$ , dann gilt

$$\det(B) = \prod_{j=1}^k \det(B_j).$$

## 10 Eigenwerte

Zum Einstieg in diesen Abschnitt betrachten wir eine Diagonalmatrix, zum Beispiel

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

und für  $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$  das Produkt der Vektoren

$$Ax.$$

Wegen der speziellen Struktur dieser Matrix lässt sich dieses Produkt in sehr einfacher Weise berechnen:

$$Ax = \begin{pmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \lambda_3 x_3 \end{pmatrix}.$$

Weiter kann man hier beispielsweise das dynamische Verhalten der Folge von Vektoren

$$Ax, A Ax = A^2 x, A A Ax = A^3 x, \dots, A^k x, \dots$$

( $k \in \mathbb{N}_+$ ), leicht beschreiben, denn es gilt:

$$A^k x = \begin{pmatrix} \lambda_1^k x_1 \\ \lambda_2^k x_2 \\ \lambda_3^k x_3 \end{pmatrix}.$$

Mit Hilfe der sogenannten Eigenwerte und Eigenvektoren, die wir im Folgenden einführen, lassen sich obige Zusammenhänge für allgemeine Matrizen erweitern. Dabei ist

$$\boxed{Av = \lambda v} \tag{10.23}$$

die zentrale Gleichung, wobei gilt  $\lambda \in \mathbb{K}$ .

**Definition 10.1** Es sei  $A \in \mathbb{K}^{n \times n}$ , dann heißt ein  $\lambda \in \mathbb{K}$  **Eigenwert** von  $A$ , falls ein  $v \in \mathbb{K}^n \setminus \{0\}$  so existiert, dass

$$Av = \lambda v.$$

Ein Vektor  $v \in \mathbb{K}^n \setminus \{0\}$ , der diese Gleichung erfüllt heißt **Eigenvektor** zum Eigenwert  $\lambda$ .

**Satz 10.2** Die in 10.1 gegebene Definition für Eigenwerte/Eigenvektoren zu  $A \in \mathbb{K}^{n \times n}$  lässt sich auf mehrere Arten in äquivalenter Weise beschreiben:

1.  $\lambda$  ist Eigenwert von  $A$  mit Eigenvektor  $v \in \mathbb{K}^n \setminus \{0\}$  genau dann, wenn

$$(A - \lambda E)v = 0.$$

2.  $\lambda$  ist Eigenwert von  $A$  genau dann, wenn  $\text{Kern}(A - \lambda E) \neq \{0\}$ .
3.  $\lambda$  ist Eigenwert von  $A$  genau dann, wenn  $\text{rang}(A - \lambda E) < n$ .
4.  $\lambda$  ist Eigenwert von  $A$  genau dann, wenn  $\det(A - \lambda E) = 0$ .
5.  $v \in \mathbb{K}^n \setminus \{0\}$  ist Eigenvektor zum Eigenwert  $\lambda$  genau dann, wenn  $v \in \text{Kern}(A - \lambda E)$ .

**Beispiel 10.3** Es sei

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

Dann sind  $\pm 1$  Eigenwerte von  $A$  (denn

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{und} \quad A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -1 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Aus Satz 10.2 (4) ergibt sich, dass die Menge der Eigenvektoren zu einem festen Eigenwert sich in die bekannte Struktur der (Unter-)Vektorräume einfügen, was im Folgenden festgehalten wird.

**Satz und Definition 10.4** Es sei  $A \in \mathbb{K}^{n \times n}$  und  $\lambda \in \mathbb{K}$  ein Eigenwert von  $A$ . Dann setzen wir

$$\text{Eig}(\lambda) := \text{Kern}(A - \lambda E) \quad (= \{v \in \mathbb{K}^n : v \text{ ist Eigenvektor zu } \lambda\} \cup \{0\}).$$

Damit ist  $\text{Eig}(\lambda)$  ein Untervektorraum des  $\mathbb{K}^n$  und wir nennen  $\text{Eig}(\lambda)$  den **Eigenraum** zum Eigenvektor  $\lambda$ . Weiter heißt  $\dim(\text{Eig}(\lambda))$  die **geometrische Vielfachheit** von  $\lambda$ .

**Beweis.** Seien  $v, w$  zwei Eigenvektoren zu einem festen Eigenwert  $\lambda$  und sei  $\alpha \in \mathbb{K}$  ein Skalar. Dann gelten

$$\begin{aligned} A(v + w) &= Av + Aw = \lambda v + \lambda w = \lambda(v + w), \\ A\alpha v &= \alpha Av = \alpha \lambda v = \lambda(\alpha v), \end{aligned}$$

so, dass gemäß Definition  $\text{Eig}(\lambda)$  ein Unterraum ist. (Die Aussage folgt auch direkt aus der Tatsache, dass der Kern einer Matrix ein Unterraum ist und  $\text{Eig}(\lambda) = \text{Kern}(A - \lambda E)$  gilt.  $\square$ )

Das folgende Resultat fasst wesentliche Eigenschaften von Eigenräumen und Eigenwerten zusammen.

**Satz 10.5** *Es seien  $A \in \mathbb{K}^{n \times n}$  und  $\lambda_1, \dots, \lambda_m$  (mit  $\lambda_j \neq \lambda_k$  für  $j \neq k$ ) die zugehörigen Eigenwerte und seien  $v_1, \dots, v_m \in \mathbb{K}^n \setminus \{0\}$  mit  $v_j \in \text{Eig}(\lambda_j)$  für  $j = 1, \dots, m$ . Dann gilt Folgendes:*

1.  $m \leq n$  (es gibt höchstens  $n$  Eigenwerte),
2.  $v_1, \dots, v_m$  sind linear unabhängig,
- 3.

$$\sum_{j=1}^m \dim(\text{Eig}(\lambda_j)) \leq n,$$

4.  $\text{Eig}(\lambda_j) \cap \text{Eig}(\lambda_k) = \{0\}$  für alle  $j \neq k$ .

**Beweis.** Zu 1.: folgt sofort aus 3., wenn man bedenkt, dass für jeden Eigenwert  $\dim(\text{Eig}(\lambda)) \geq 1$  gelten muss.

Zu 2.: Wir zeigen die Behauptung durch Induktion. Für den Induktionsanfang  $m = 1$  gilt 2. sofort. Im Induktionsschritt nehmen wir also an, dass jeweils  $m-1$  Eigenvektoren linear unabhängig sind. Wir zeigen, nun dass kein  $v_j$  sich als Linearkombination der übrigen  $(m-1)$  Vektoren darstellen lässt. Zu diesem Zweck nehmen wir an, dass dies doch geht. Es gelte also ohne Einschränkung

$$v_m = \sum_{j=1}^{m-1} \alpha_j v_j.$$

Da  $v_m \neq 0$  gilt können dabei nicht alle  $\alpha_j$  verschwinden. Durch Multiplikation beider Seiten mit  $A$  ergibt sich damit

$$Av_m = \lambda_m v_m = \sum_{j=1}^{m-1} \alpha_j Av_j = \sum_{j=1}^{m-1} \alpha_j \lambda_j v_j,$$

und durch Ersetzen von  $v_m$  folgt dann

$$\lambda_m \sum_{j=1}^{m-1} \alpha_j v_j = \sum_{j=1}^{m-1} \alpha_j \lambda_j v_j$$

und damit

$$0 = \sum_{j=1}^{m-1} (\lambda_j - \lambda_m) \alpha_j v_j.$$

Da nicht alle  $(\lambda_j - \lambda_m) \alpha_j$  verschwinden können, ist dies ein Widerspruch zur Induktionsannahme.

3. Genau wie in 2. zeigt man, dass Basen sämtlicher Eigenräume zusammen eine linear unabhängige Menge von Vektoren bilden. Die Anzahl dieser Vektoren kann damit die Dimension von  $\mathbb{K}^n$  nicht überschreiten, woraus die Behauptung folgt.

4. vgl. Übung. □

und das kannn man auch Das obige Beispiel 10.3 zeigt einen Fall, in dem die Anzahl der Eigenwerte mit der Anzahl der Zeilen oder Spalten der Matrix übereinstimmt. Das folgende Beispiel zeigt, dass aber auch der andere Fall, nämlich, dass es gar keinen reellen Eigenwert gibt vorkommen kann.

**Beispiel 10.6** Es sei

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

Dann hat  $A$  keinen Eigenwert in  $\mathbb{R}$ .

Wir werden im Folgenden jedoch sehen, dass jede Matrix mindestens einen (möglicherweise komplexen) Eigenwert hat.



**Satz und Definition 10.7** Es sei  $A \in \mathbb{K}^{n \times n}$ . Dann ist  $P : \mathbb{K} \rightarrow \mathbb{K}$ , definiert durch

$$P(\lambda) := P_A(\lambda) := \det(A - \lambda E) \quad (\lambda \in \mathbb{K}),$$

ein Polynom vom Grad  $n$  (mit höchstem Koeffizient  $(-1)^n$ ). Das Polynom  $P_A$  heißt **charakteristisches Polynom** von  $A$

Im Fall von Matrizen in  $\mathbb{K}^{2 \times 2}$ ,  $\mathbb{K}^{3 \times 3}$  ergeben sich aus Satz 10.7 die folgenden charakteristischen Polynome:

**Bemerkung 10.8** 1. Für

$$A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{K}^{2 \times 2}$$

ist das **charakteristische Polynom** gegeben durch

$$P_A(\lambda) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21}.$$

2. Für

$$A := \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in \mathbb{K}^{3 \times 3}$$

ist das **charakteristische Polynom** gegeben durch

$$\begin{aligned} P_A(\lambda) &= (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{31}(a_{22} - \lambda)a_{13} - a_{32}a_{23}(a_{11} - \lambda) - (a_{33} - \lambda)a_{21}a_{12}. \end{aligned}$$

Zusammen mit Satz 10.2 (4) und dem Fundamentalsatz der Algebra (siehe Anhang) ergibt sich das folgende wichtige Resultat.

**Satz und Definition 10.9** Es sei  $A \in \mathbb{K}^{n \times n}$ .

1. Die Eigenwerte von  $A$  sind genau die Nullstellen des Polynoms  $P_A$ . Gilt weiter  $\mathbb{K} = \mathbb{C}$ , dann hat  $A$  mindestens einen Eigenwert.
2. Ist das charakteristische Polynom zu  $P_A$  gegeben durch

$$P_A(z) = (\lambda_1 - z)^{\alpha_1} (\lambda_2 - z)^{\alpha_2} \cdot \dots \cdot (\lambda_k - z)^{\alpha_k}.$$

Für  $A \neq 0$ , gilt dann nach dem Fundamentalsatz der Algebra  $\alpha_1 + \dots + \alpha_k = n$ , und  $\alpha_j \in \mathbb{N}_+$  heißt die **algebraische Vielfachheit** des Eigenwerts  $\lambda_j$ ,  $j = 1, \dots, k$ .

3. Die algebraische Vielfachheit eines Eigenwerts ist größer oder gleich seiner geometrischen Vielfachheit.

Für Matrizen mit speziellen Strukturen lassen sich die Eigenwerte häufig unmittelbar ablesen:

**Satz 10.10** 1. Es sei  $A \in \mathbb{K}^{n \times n}$  eine untere- oder obere Dreiecksmatrix, dann sind die Diagonaleinträge  $[A]_{jj}$ ,  $j = 1, \dots, n$  die Eigenwerte von  $A$ .

2. Es sei  $B \in \mathbb{K}^{n \times n}$  eine Blockmatrix mit Blöcken  $B_j \in \mathbb{K}^{n_j \times n_j}$ ,  $j = 1, \dots, k$ , dann sind die Eigenwerte von  $B$  gegeben durch die Vereinigung der Eigenwerte der  $B_j$ ,  $j = 1, \dots, k$ .

Wir widmen wir uns der Frage unter welchen Bedingungen die Eigenvektoren den gesamten Raum aufspannen, und welchen Bedingungen Eigenvektoren sogar eine ONB bilden können. Dazu zunächst folgende Definition.

**Definition 10.11** Es seien  $A, B \in \mathbb{K}^{n \times n}$ . Dann heißen  $A$  und  $B$  *ähnlich*, falls eine invertierbare Matrix  $C \in \mathbb{K}^{n \times n}$  so existiert, dass

$$B = C^{-1}AC.$$

Für solche Matrizen gilt das Folgende:

**Satz 10.12** Sind  $A, B \in \mathbb{K}^{n \times n}$  ähnlich, so ist

$$P_A = P_B.$$

Somit haben ähnliche Matrizen die selben Eigenwerte.

**Beweis.** Es gilt mit Satz 9.10

$$\begin{aligned} P_B(z) &= \det(B - zE) = \det(C^{-1}AC - zC^{-1}C) \\ &= \det(C^{-1}(A - zE)C) = \det(C^{-1}) \det(A - zE) \det(C) \\ &= \det(A - zE) = P_A(z). \end{aligned}$$

□

Damit kann man die Untersuchung der Potenzen  $B^k$  auf die Potenzen  $A^k$  zurückführen (sofern man  $C$  kennt) denn es gilt

$$B^k = C^{-1}AC C^{-1}AC \cdot \dots \cdot CC^{-1}AC = C^{-1}A^kC. \quad (10.24)$$

Dies ist insbesondere dann nützlich, wenn  $A$  eine Diagonalmatrix ist, wie zu Beginn dieses Abschnitts und im nachfolgenden Beispiel erläutert.

**Beispiel 10.13** Es seien

$$v_0 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad B := \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

und

$$v_k = Bv_{k-1} \quad (k \in \mathbb{N}_+).$$

Durch diese Iteration erhält man die Fibonacci Zahlen  $F_0 = 0$ ,  $F_1 = F_2 = 1$  und  $F_{k+2} := F_{k+1} + F_k$  für  $k > 2$ , denn es gilt

$$\begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix} = v_k.$$

Das charakteristische Polynom zu  $B$  ist gegeben durch

$$P_B(\lambda) = \lambda^2 - \lambda - 1,$$

und man zeigt leicht, dass die Nullstellen und somit die Eigenwerte gegeben sind durch

$$\lambda_1 = \frac{1 + \sqrt{5}}{2}, \quad \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

Die zugehörigen Eigenvektoren sind  $u_1 = (\lambda_1, 1)^T$  und  $u_2 = (\lambda_2, 1)^T$ . Mit

$$v_0 = \frac{1}{\lambda_1 - \lambda_2}(u_1 - u_2)$$

kann man nun  $v_k$ , somit auch die Fibonacci Zahlen, sehr zügig berechnen, denn es gilt:

$$\begin{aligned} v_k &= B^k v_0 = B^k \left( \frac{u_1 - u_2}{\lambda_1 - \lambda_2} \right) \\ &= \frac{1}{\lambda_1 - \lambda_2} (B^k u_1 - B^k u_2) = \frac{1}{\lambda_1 - \lambda_2} (\lambda_1^k u_1 - \lambda_2^k u_2). \end{aligned} \quad (10.25)$$

Mit der Matrix  $U = (u_1 \ u_2)$  kommt man zur gleichen Vereinfachung, denn es gilt

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = U^{-1} B U$$

und somit lässt sich  $B^k$  gemäß (10.24) leicht berechnen durch

$$B^k = U \begin{pmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{pmatrix} U^{-1}.$$

Das vorangehende Beispiel steht stellvertretend für eine ganze Reihe von Anwendungen, die durch **Diagonalisierung** stark vereinfacht werden. Ist  $A \in \mathbb{K}^{n \times n}$  und ist  $C \in \mathbb{K}^{n \times n}$  ein Matrix, die Spaltenweise aus linear unabhängigen Eigenvektoren besteht, dann ist

$$D = C^{-1} A C$$

eine Diagonalmatrix mit den Eigenwerten als Einträge auf der Hauptdiagonalen. Ein solches  $C$  existiert jedoch nicht in allen Fällen, weshalb man folgendes definiert:

**Definition 10.14** Ist  $A \in \mathbb{K}^{n \times n}$ , so heißt  $A$  **diagonalisierbar**, falls  $A$  ähnlich zu einer Diagonalmatrix (vgl. (2.1)) ist. D.h. es gibt  $C \in \mathbb{K}^{n \times n}$  so, dass

$$D = C^{-1}AC.$$

Die Existenz einer Matrix  $C \in \mathbb{K}^{n \times n}$  bestehend aus linear unabhängigen Eigenvektoren ist notwendig und hinreichend für die Diagonalisierbarkeit, was im Wesentlichen der Inhalt des folgenden Theorems ist.

**Satz 10.15** Es sei  $A \in \mathbb{K}^{n \times n}$ . Dann sind äquivalent:

- a)  $A$  ist diagonalisierbar.
- b) Sind  $\lambda_1, \dots, \lambda_k$  alle Eigenwerte von  $A$ , dann ist

$$\bigoplus_{j=1}^k \text{Eig}(\lambda_j) = \mathbb{K}^n.$$

- c) Es existiert eine Basis  $(v_1, \dots, v_n)$  bestehend aus Eigenvektoren von  $A$ .

Das folgende wichtige Resultat zeigt, dass die wichtige Klasse der symmetrischen Matrizen stets (in besonderer Weise) diagonalisierbar sind. Wir beschränken uns hierbei auf reelle Matrizen.

**Satz 10.16 (Spektralsatz)** Es sei  $A \in \mathbb{R}^{n \times n}$  symmetrisch, dann sind alle Eigenwerte reell und es gibt eine ONB  $(v_1, \dots, v_n)$  bestehend aus Eigenvektoren von  $A$ .

Im oberen Satz gilt damit für  $Q = (v_1 \dots v_n) \in \mathbb{R}^{n \times n}$

$$D = Q^{-1}AQ = Q^T AQ,$$

wobei  $D \in \mathbb{R}^{n \times n}$  eine Diagonalmatrix ist, deren Diagonaleinträge die Eigenwerte von  $A$  (gemäß der Ordnung der Eigenvektoren in  $C$ ) sind.

**Beweis.** Wir zeigen zunächst, dass Eigenvektoren zu unterschiedlichen Eigenwerten orthogonal zueinander stehen. Seien dazu  $\lambda_1 \neq \lambda_2$  reell und  $v_1, v_2$  zugehörige Eigenvektoren. Da  $A$  symmetrisch ist, ist auch

$$B := A - \lambda_1 E$$

symmetrisch und daher ist  $\text{Im } B = \text{Im } B^T$  und damit  $Bv_2 = (\lambda_2 - \lambda_1)v_2 \in \text{Im } B^T$ . Weiter gilt  $Bv_1 = 0$ , sodass  $v_1 \in \text{Kern } B$ . Nach Satz 6.5 gilt sogar  $(\text{Kern } B)^\perp = \text{Im } B^T$ , und daher  $\langle v_1, v_2 \rangle = 0$ .

Wir zeigen nun, dass jeder Eigenwert reell sein muss. Dazu betrachten wir  $A$  als Matrix in den komplexen Zahlen und nehmen an, dass  $\lambda \in \mathbb{C}$  ein Eigenwert, und  $v \in \mathbb{C}^n$  ein zugehöriger Eigenvektor sind. Dann ist

$$\bar{v}^T A v = \lambda \bar{v}^T v = \lambda \langle v, v \rangle.$$

Da  $A$  symmetrisch ist, ist die linke Seite reell (\*Übung) und weiter ist  $\langle v, v \rangle$  reell, sodass  $\lambda$  ebenfalls reell sein muss.

Nun müssen wir noch zeigen, dass es eine Basis von Eigenvektoren gibt. Wir nehmen dazu an, dass bereits  $l$  Eigenwerte  $\lambda_1, \dots, \lambda_l$  gefunden seien und es seien  $v_1, \dots, v_l$  eine ONB aus Eigenvektoren zu

$$V = \text{Span}(v_1, \dots, v_l) = \bigoplus_{j=1}^l \text{Eig}(\lambda_j).$$

(Man beachte, dass aufgrund von Vielfachheit von Eigenwerten  $k > l$  gelten kann) Wir zeigen, dass  $V^\perp$  einen weiteren Eigenvektor (zu einem weiteren Eigenwert), enthält vorausgesetzt  $V^\perp \neq \{0\}$ , d.h.  $k < n$ . Unter dieser Voraussetzung gibt es eine ONB  $w_{k+1}, \dots, w_n$  von  $V^\perp$ , und wir definieren  $U = (w_{k+1} \dots w_n) \in \mathbb{R}^{n \times n-k}$ . Gemäß Satz 6.13, implementiert

$$P := UU^T$$

die orthogonale Projektion auf  $V^\perp$ . Gilt dabei  $AP = 0$  (=Nullmatrix), dann ist jeder Vektor  $\neq 0$  ein Eigenvektor zum Eigenwert 0 und wir wählen  $v$  als einen beliebigen Vektor aus  $V^\perp \setminus \{0\}$ ,  $\|v\| = 1$ , so dass für  $\lambda = 0$  gilt

$$APv = \lambda v.$$

Gilt hingegen  $AP \neq 0$  so hat  $AP$  mindestens einen Eigenwert  $\lambda \neq 0$ . Da  $P$  und  $A$  beide symmetrisch sind, ist  $AP$  symmetrisch (\*\*Übung) und somit ist  $\lambda$  nach dem oben Gezeigten reell. Sei in diesem Fall  $v$  mit  $\|v\| = 1$  ein Eigenvektor zu  $\lambda$ . Wegen  $Pv_j = 0$  für alle  $j = 1, \dots, k$  sind alle  $v_j$  Eigenvektoren von 0 (bzgl.  $AP$ ) und können daher keine Eigenvektoren zu  $\lambda$  sein. In den beiden Fällen gilt  $v \in V^\perp$  um somit  $Pv = v$ , woraus

$$\lambda v = APv = Av$$

folgt. Damit ist  $v$  ein Eigenvektor von  $\lambda$  bezüglich  $A$ . Falls nun  $AP = 0$  gilt, bildet eine ONB von  $V^\perp$  zusammen mit  $v_1, \dots, v_k$  eine ONB des  $\mathbb{R}^n$ . Gilt  $AP \neq 0$  gilt  $\text{Eig}(\lambda) \subset V^\perp$  und eine ONB zu  $\text{Eig}(\lambda)$  ergibt zusammen mit  $v_1, \dots, v_k$  eine ONB zu

$$\text{Eig}(\lambda) \oplus \left( \bigoplus_{j=1}^l \text{Eig}(\lambda_j) \right).$$

Die iterative Anwendung dieser Argumentation liefert dann schließlich eine ONB von Eigenvektoren des  $\mathbb{R}^n$ .  $\square$

Dieses Resultat führt uns nun zu den sogenannten **Singulärwerten**, mit Hilfe derer die Idee der Eigenwerte auf allgemeinere (nicht quadratische) Matrizen erweitert werden. Zuvor definieren wir noch folgendes:

**Definition 10.17** Es sei  $A \in \mathbb{R}^{n \times n}$ .

1) Dann heißt  $A$  **positiv (semi-)definit**, falls für alle  $v \in \mathbb{R}^n$ ,  $v \neq 0$  gilt

$$v^T A v > 0, (v^T A v \geq 0)$$

2)  $A$  **negativ (semi-)definit**, falls für alle  $v \in \mathbb{R}^n$ ,  $v \neq 0$  gilt

$$v^T A v < 0, (v^T A v \leq 0)$$

**Bemerkung 10.18**  $A \in \mathbb{R}^{n \times n}$  mit Eigenwert  $\lambda$  und Eigenvektor  $v$ . Aus

$$v^T A v = \lambda \langle v, v \rangle$$

folgt dann unmittelbar.

- Ist  $A$  positiv (semi-) definit, dann sind alle Eigenwerte größer (größer oder gleich) null.
- Ist  $A$  negativ (semi-) definit, dann sind alle Eigenwerte kleiner (kleiner oder gleich) null.

Für  $A \in \mathbb{R}^{m \times n}$  sieht man leicht, dass

$$A^T A \in \mathbb{R}^{n \times n}$$

positiv semi-definit und symmetrisch ist. Damit sind alle Eigenwerte von  $A^T A$  nicht negative reelle Zahlen. Nach dem Spektralsatz gibt es weiter eine ONB bestehend aus Eigenvektoren  $v_1, \dots, v_n$ . Zählen wir nun die Eigenwerte entsprechend der Dimension der zugehörigen Eigenräume, ordnen diese der Größe nach, so erhalten wir eine endliche Folge

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0 \quad (10.26)$$

und

$$\lambda_{r+1} = \dots = \lambda_n = 0$$

für ein  $r \leq n$ .

**Definition 10.19** Aus 10.26 ergeben sich die **Singulärwerte**  $\sigma_1, \dots, \sigma_r$  von  $A$  durch

$$\sigma_1 := \sqrt{\lambda_1} \geq \sigma_2 := \sqrt{\lambda_2} \geq \dots \geq \sigma_r := \sqrt{\lambda_r} > 0.$$

Setzen wir nun  $V = (v_1 \dots v_n) \in \mathbb{R}^{n \times n}$  (und setzen voraus die  $v_1, \dots, v_n$  sind gemäß der  $\sigma_j$  geordnet (mit den etwaigen Eigenvektoren zu  $\lambda = 0$  zuletzt), dann gilt

$$A^T A = V \begin{pmatrix} D & 0_{r, n-r} \\ 0_{n-r, r} & 0_{n-r, n-r} \end{pmatrix} V^T, \quad (10.27)$$

wobei

$$D = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & \dots & & \sigma_r^2 \end{pmatrix} \in \mathbb{R}^{r \times r},$$



und  $0_{k,j} \in \mathbb{R}^{k \times j}$  nur aus Nulleinträgen besteht. Es gilt also  $A^T A v_j = \sigma_j^2 v_j$  für alle  $j = 1, \dots, r$ . Setzen wir nun

$$u_j := \frac{A v_j}{\sigma_j}, \quad j = 1, \dots, r, \quad (10.28)$$

dann gilt

$$A A^T u_j = A \frac{1}{\sigma_j} (A^T A v_j) = \sigma_j^2 u_j,$$

und somit sind  $u_1, \dots, u_r$  normierte ( $\|u_j\| = 1$ ) Eigenvektoren von  $A A^T$  zu den Eigenwerten  $\sigma_1^2, \dots, \sigma_r^2 > 0$ . Da die  $v_j$  orthogonal sind, sind auch  $u_1, \dots, u_r$  orthogonal (\*Übungsaufgabe). Mit Satz 6.5 gilt

$$\text{rang}(A^T A) = \text{rang}(A A^T),$$

woraus folgt, dass  $A A^T$  keine weiteren Eigenwerte ungleich Null haben kann. Erweitern wir  $u_1, \dots, u_r$  zu einer ONB  $u_1, \dots, u_r, \dots, u_m$  des  $\mathbb{R}^m$ , dann bilden  $(u_{r+1}, \dots, u_m)$  somit eine ONB zu  $\text{Kern}(A A^T)$ . Setzen wir  $U = (u_1 \dots u_m) \in \mathbb{R}^{m \times m}$ , dann erhalten wir, analog zu (10.27),

$$A A^T = U \begin{pmatrix} D & 0_{r, m-r} \\ 0_{m-r, r} & 0_{m-r, m-r} \end{pmatrix} U^T. \quad (10.29)$$

mit  $D$  und  $0_{k,j} \in \mathbb{R}^{k \times j}$  wie in (10.27). Setzen wir nun

$$\Sigma_r = \begin{pmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & \dots & & \sigma_r \end{pmatrix} \in \mathbb{R}^{r \times r},$$

und

$$\Sigma = \begin{pmatrix} \Sigma_r & 0_{r, n-r} \\ 0_{m-r, r} & 0_{m-r, n-r} \end{pmatrix}$$

so gilt nach (10.28)

$$A V = U \Sigma,$$

und da  $V$  orthonormal ist gilt damit

$$\boxed{A = U \Sigma V^T}. \quad (10.30)$$

## Prinzipien der Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (PCA) ist eine Methode zur Dimensionsreduktion, die auf den Konzepten der Eigenwerte und Eigenvektoren aufbaut. Sie dient dazu, hochdimensionale Daten in eine niedrigdimensionale Struktur zu überführen, wobei der maximale Informationsgehalt erhalten bleibt.

Ein typisches Problem in der Datenanalyse ist die Verarbeitung hochdimensionaler Daten. Beispielsweise haben Bild- oder Textdaten oft viele Merkmale, von denen jedoch nicht alle gleichermaßen zur Analyse beitragen. PCA ermöglicht:

- Datenkompression durch Reduktion der Dimension mit minimalem Informationsverlust.
- Rauschunterdrückung, indem nur die wichtigsten Komponenten der Daten berücksichtigt werden.
- Visualisierung hochdimensionaler Daten in niedrigeren Dimensionen (z. B. 2D oder 3D).

Gegeben sei eine zentrierte Datenmatrix  $X \in \mathbb{R}^{m \times n}$  mit  $m$  Beobachtungen und  $n$  Merkmalen. Die Kovarianzmatrix  $C$  von  $X$  ist definiert als:

$$C = \frac{1}{m} X^T X.$$

Die Eigenvektoren dieser Kovarianzmatrix geben die Richtungen maximaler Varianz in den Daten an, während die Eigenwerte die Varianz entlang dieser Richtungen bestimmen.

Eine effiziente Methode zur Berechnung der Hauptkomponenten ist die Singulärwertzerlegung (SVD) der Datenmatrix  $X$ :

$$X = U \Sigma V^T.$$

- Die Spalten von  $V$  sind die Hauptkomponenten, also die Eigenvektoren von  $X^T X$ .
- Die Diagonalelemente von  $\Sigma$  (die Singulärwerte) sind mit den Wurzeln der Eigenwerte von  $C$  verknüpft.
- Durch die Auswahl der ersten  $k$  Spalten von  $V$  kann man eine dimensionsreduzierte Darstellung der Daten berechnen:

$$X_k = X V_k.$$

Hierbei repräsentiert  $X_k$  die Projektion der Daten auf den neuen, niedrigdimensionalen Raum.

**Beispiel:** Ein Graustufenbild mit  $100 \times 100$  Pixeln hat eine Dimension von 10.000. Durch PCA kann dieses auf wenige Hauptkomponenten reduziert werden, ohne dass visuelle Informationen verloren gehen.

## 10.1 Graphische Darstellung mit TikZ

```
[<img alt="TikZ code for a PCA plot showing principal components x1 and x2, and data points." data-bbox="134 297 847 359"/>] (-2,0) -- (2,0) node[right] x1; [<img alt="TikZ code for a PCA plot showing principal components x1 and x2, and data points." data-bbox="134 297 847 359"/>] (0,-2) -- (0,2) node[above] x2; [<img alt="TikZ code for a PCA plot showing principal components x1 and x2, and data points." data-bbox="134 297 847 359"/>] (<img alt="TikZ code for a PCA plot showing principal components x1 and x2, and data points." data-bbox="134 297 847 359"/>0,0)
-- (1.5,1) node[right] PC1; [<img alt="TikZ code for a PCA plot showing principal components x1 and x2, and data points." data-bbox="134 297 847 359"/>] (0,0) -- (-1,1.5) node[above] PC2; in
-1.5,-1,-0.5,0,0.5,1,1.5 in -1.5,-1,-0.5,0,0.5,1,1.5 [black] (,) circle (1pt);
```

## A Zusammenfassung Matrix Faktorisierung

### A.1 LR-Zerlegung (LU-Decomposition)

Gegeben  $A \in \mathbb{R}^{n \times n}$  invertierbar.

Dann können im Gauß-Verfahren die Elementarmatrizen  $B_1, \dots, B_s$ , die zu einer Zeilenstufenform führen so zu wählen, dass

$$(B_s \cdot \dots \cdot B_2 B_1)^{-1} =: L$$

eine untere Dreiecksmatrix mit  $[L]_{jj} = 1$  für alle  $j = 1, \dots, n$  ist. Dabei ist  $L$  eine solche untere Dreiecksmatrix genau dann, wenn auch

$$B_s \cdot \dots \cdot B_2 B_1 = L^{-1}$$

eine solche ist. Die Matrix in Zeilenstufenform entspricht hier einer oberen Dreiecksmatrix  $R$  und es gilt

$$L^{-1}A = R \Leftrightarrow$$

$$\boxed{A = LR \text{ LR-Zerlegung}} \quad (\text{A.31})$$

### A.2 QR-Zerlegung

Gegeben  $A \in \mathbb{R}^{m \times n}$  mit  $m \geq n$  und  $\text{rang}(A) = n$ , was insbesondere invertierbare Matrizen einschließt. Seien  $a_1, \dots, a_n \in \mathbb{R}^m$  die linear unabhängigen Spalten von  $A$ .

Im Gram-Schmidt-Verfahren (zur Orthonormalisierung der  $a_j$ ) gilt

$$v_l := a_l - \sum_{j=1}^{l-1} \frac{\langle v_j, a_l \rangle}{\langle v_j, v_j \rangle} v_j, \quad l = 1, \dots, n.$$

und somit wird  $v_l$  jeweils als Linearkombination von  $a_l$  und  $v_1, \dots, v_{l-1}$  berechnet. Sukzessives Einsetzen von  $v_1$  ausgehend liefert daher, dass  $v_l$  als Linearkombination der Form

$$v_l = \sum_{j=1}^l r_{lj}^- a_j$$

geschrieben werden kann, wobei entsprechend  $r_{jj}^- = 1$ ,  $j = 1, \dots, n$  gilt. Setzen wir weiter  $r_{kj}^- = 0$  für  $k > j$ , dann ergibt sich

$$(v_1 \dots v_n) = (a_1 \dots a_n) R^- \quad (\text{A.32})$$

mit

$$R^- := \begin{pmatrix} r_{11}^- & r_{12}^- & r_{13}^- & \dots & a_{1n} \\ 0 & r_{22}^- & r_{23}^- & \ddots & r_{2n}^- \\ 0 & 0 & r_{33}^- & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_{nn}^- \end{pmatrix}. \quad (\text{A.33})$$

Dabei ist  $R^-$  invertierbar. Weiterhin kann die Normalisierung im Gram-Schmidt-Verfahren durch eine invertierbare Diagonalmatrix  $D$  implementiert werden. Durch spaltenweises Konkatenieren der resultierenden orthonormalen Vektoren ergibt sich  $(u_1 \dots u_n) =: Q \in \mathbb{R}^{m \times n}$  mit

$$Q = AR^-D \quad (\text{A.34})$$

Mit  $R := (R^-D)^{-1}$  gilt schließlich

$$\boxed{A = QR \text{ QR-Zerlegung}} \quad (\text{A.35})$$

Engl. **LU-decomposition** (lower-upper triangular)).

### A.3 Singulärwertzerlegung

Gegeben:  $A \in \mathbb{R}^{m \times n}$ .

$\Rightarrow A^T A \in \mathbb{R}^{n \times n}$  ist symmetrisch und positive semi-definit.  $\Rightarrow A^T A$  hat nicht negative Eigenwerte  $\lambda_1 > \lambda_2 > \dots \lambda_r > \lambda_{r+1} = \dots = \lambda_j = 0$ .

**Singulärwerte**  $\sigma_1, \dots, \sigma_r$  von  $A$  sind definiert durch

$$\sigma_1 := \sqrt{\lambda_1} \geq \sigma_2 := \sqrt{\lambda_2} \geq \dots \geq \sigma_r := \sqrt{\lambda_r} > 0.$$

Seien  $V = (v_1 \dots v_n) \in \mathbb{R}^{n \times n}$  (und setzen voraus die  $v_1, \dots, v_n$  sind gemäß der  $\sigma_j$  geordnet (mit den etwaigen Eigenvektoren zu  $\lambda = 0$  zuletzt), dann gilt

$$A^T A = V \begin{pmatrix} D & 0_{r, n-r} \\ 0_{n-r, r} & 0_{n-r, n-r} \end{pmatrix} V^T, \quad (\text{A.36})$$

wobei

$$D = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_r^2 \end{pmatrix} \in \mathbb{R}^{r \times r},$$

und  $0_{k,j} \in \mathbb{R}^{k \times j}$  nur aus Nulleinträgen besteht. Setze

$$u_j := \frac{Av_j}{\sigma_j}, \quad j = 1, \dots, r, \quad (\text{A.37})$$

Erweitern  $u_1, \dots, u_r$  zu einer ONB  $u_1, \dots, u_r, \dots, u_m$  des  $\mathbb{R}^m$ , und setze  $U = (u_1 \dots u_m) \in \mathbb{R}^{m \times m}$ , dann folgt

$$AA^T = U \begin{pmatrix} D & 0_{r, m-r} \\ 0_{m-r, r} & 0_{m-r, m-r} \end{pmatrix} U^T. \quad (\text{A.38})$$

mit  $D$  und  $0_{k,j} \in \mathbb{R}^{k \times j}$  Mit

$$\Sigma_r = \begin{pmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & \dots & & \sigma_r \end{pmatrix} \in \mathbb{R}^{r \times r},$$

und

$$\Sigma = \begin{pmatrix} \Sigma_r & 0_{r, n-r} \\ 0_{m-r, r} & 0_{m-r, n-r} \end{pmatrix}$$

$$AV = U\Sigma,$$

und da  $V$  orthonormal ist, folgt die Singulärwertzerlegung

$$\boxed{A = U\Sigma V^T}. \quad (\text{A.39})$$

## B Vektorräume

In diesem Abschnitt werden wir die abstrakte algebraische Struktur, die dem  $\mathbb{R}^n$  zugrunde liegt einführen.

**Definition B.1** Es sei  $K$  ein Körper und  $V$  eine Menge (Vektoren) mit zwei Abbildungen (Addition, skalare Multiplikation)

$$\begin{aligned} + : V \times V &\rightarrow V, (x, y) \mapsto x + y, \\ \cdot : K \times V &\rightarrow V, (\lambda, x) \mapsto \lambda \cdot x =: \lambda x \end{aligned}$$

so, dass folgende (Vektorraumaxiome) gelten:

1. Für alle  $v, w, z \in V$  gilt  $(v + w) + z = v + (w + z)$  (Assoziativgesetz).
2. Für alle  $v, w \in V$  gilt  $v + w = w + v$  (Kommutativgesetz).
3. Es gibt ein Element  $\vec{0} = 0 \in V$  mit  $v + 0 = v$  für alle  $v \in V$  (neutrales Element bzgl. Addition).
4. Für alle  $v \in V$  existiert ein  $-v \in V$ , so dass  $v + (-v) = 0$  (inverses Element bzgl. Addition).
5. Für alle  $\lambda, \mu \in K$  und  $v \in V$  gilt  $\lambda(\mu v) = (\lambda\mu)v$ .
6. Für alle  $v \in V$  gilt  $1v = v$  (mit 1 neutrales Element der Multiplikation auf  $K$ ).
7. Für alle  $\lambda \in K$  und  $v, w \in V$  gilt  $\lambda(v + w) = \lambda v + \lambda w$ .
8. Für alle  $\lambda, \mu \in K$  und  $v \in V$  gilt  $(\lambda + \mu)v = \lambda v + \mu v$ .

Dann heißt  $V$  ein  $K$ -Vektorraum, oder Vektorraum über  $K$ , oder auch kurz **Vektorraum**.

Im Fall des  $\mathbb{R}^n$  folgen die Vektorraumaxiome unmittelbar aus den Eigenschaften des Körpers der reellen Zahlen  $\mathbb{R}$  und Satz [1.2](#)

- Bemerkung B.2** 1. Die Axiome 1-4 aus Definition ?? sind äquivalent zu der Aussage, dass  $(V, +)$  eine kommutative (abelsche) Gruppe ist.
2. Wir betrachten vorwiegend den Spezialfall der reellen Vektorräume. Das heißt in Definition ?? ist dann  $K = \mathbb{R}$ .

**Satz B.3** *Es sei  $V$  ein  $K$ -Vektorraum, dann gilt:*

1.  $\lambda \vec{0} = \vec{0}$  für alle  $\lambda \in K$ ,
2.  $0v = \vec{0}$  für alle  $v \in V$ ,
3.  $(-1)v = -v$  für alle  $v \in V$ .

Der Raum  $\mathbb{R}^n$  ist insbesondere für Anwendungen der wichtigste Vektorraum. Jedoch gibt es in einigen Bereichen andere Vektorräume, wie das nachfolgende Beispiel zeigt. Es hat einige Jahrhunderte gedauert, bis die heutigen Vektorraumaxiome als passende algebraische Struktur herausgearbeitet waren. Eine der ersten Quellen, in der die Axiome in der heutigen Form erschienen sind ist das Buch *Moderne Algebra* aus dem Jahre 1930 von van der Waerden. In den vorangehenden Jahrhunderten hatten Wissenschaftler wie Leibniz, Gauß, Caley und einige andere sich schon intensiv mit Methoden der linearen Algebra, wie dem Lösen linearer Gleichungssysteme, beschäftigt. Der Vorteil allgemeiner algebraischer Strukturen wie Gruppen, Körpern, Vektorräumen usw. ist eine breite Anwendbarkeit ihrer Ergebnisse auf viele Anwendungsfälle. Außerdem helfen sie dabei in neuen Anwendungen zielsicher gewisse Eigenschaften zu betrachten, die anschließend unmittelbar eine Vielzahl mathematischer Werkzeuge und ein theoretisches Fundament liefern können.

- Beispiel B.4** 1. Die Menge  $\mathbb{R}^n$  mit den eingeführten Rechenregeln ist ein Vektorraum.
2. Es sei  $K = \{0, 1\}$  mit der Addition und der Multiplikation gemäß der Restklassen  $\mathbb{Z}_2$  (vgl. Grundlagen der Mathematik):

+	0	1	·	0	1
0	0	1	0	0	0
1	1	0	1	0	1



Damit ist  $K$  ein Körper. Sei weiter  $V = K^n$  die Menge aller  $n$ -Tupel von Elementen aus  $K$ . Für  $x, y \in K^n$  und  $\lambda \in K$  sei analog zum  $\mathbb{R}^n$  die Addition definiert durch

$$x + y := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

und die skalare Multiplikation definiert durch

$$\lambda x := \lambda \cdot x := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Damit ist  $V$  ein Vektorraum über  $K$ . Solche Vektorräume über endlichen Körpern finden zum Beispiel in der Codierungstheorie ihre Anwendung.

3. Sei  $T$  die Menge aller Funktionen  $f : \mathbb{R} \rightarrow \mathbb{R}$  der Form

$$f(x) = a_0 + \sum_{k=1}^n a_k \cos(kx) + \sum_{k=1}^n b_k \sin(kx)$$

$n \in \mathbb{N}$ ,  $a_0, \dots, a_n \in \mathbb{R}$ ,  $b_1, \dots, b_n \in \mathbb{R}$ . Mit der punktweisen Addition

$$(f + g)(x) := f(x) + g(x),$$

und der skalaren Multiplikation

$$(\lambda f)(x) := \lambda f(x),$$

$f, g \in T$  und  $\lambda \in \mathbb{R}$ , ist  $T$  ein Vektorraum über  $\mathbb{R}$ . Dieser Raum wird Raum der trigonometrischen Polynome genannt. Er spielt in der physikalischen Beschreibung von Wellen und damit auch im Bereich der Signalverarbeitung eine fundamentale Rolle.

Wir werden nun den wichtigen Begriff des Unterraums (Untervektorraums) einführen. Dieser Begriff wird später im Raum  $\mathbb{R}^n$  eine zentrale Rolle spielen.

**Definition B.5** Es seien  $V$  ein  $K$ -Vektorraum und  $U \subset V$ . Ist  $U$  mit der Addition und Skalarmultiplikation von  $V$  selbst wieder ein Vektorraum, so heißt  $U$  **Untervektorraum** (oder kurz Unterraum) von  $V$ .

In der Praxis wird die folgende Charakterisierung verwendet.

**Satz B.6** *Es sei  $V$  ein  $K$ -Vektorraum und  $U \subset V$ ,  $U \neq \emptyset$ , dann ist  $U$  genau dann ein Untervektorraum von  $V$ , wenn gilt:*

1. für alle  $v, w \in U$  gilt  $v + w \in U$ ,
2. für alle  $\lambda \in K$  und alle  $v \in U$  gilt  $\lambda v \in U$ .

Die folgende Aussage liefert ein einfaches Kriterium mit dem sich häufig schon schnell ausschließen lässt, dass gewisse Mengen Unterräume sind.

**Bemerkung B.7** 1. Mit Hilfe von Satz ?? (2) und Satz 4.1 (2) folgt sofort, dass jeder Unterraum den Nullvektor enthält.

2. Die Menge  $\{0\}$ , die nur aus dem Nullvektor besteht ist immer ein Unterraum.

Analog zu Definition 1.3 und Definition 1.5 lassen sich die Begriffe der Linearkombination und des aufgespannten Raums in dem abstrakten Rahmen definieren

**Definition B.8** Es sei  $V$  ein  $K$ -Vektorraum und seien  $v_1, \dots, v_n \in V$  und  $\lambda_1, \dots, \lambda_n \in K$  ( $n \in \mathbb{N}$ ). Dann heißt die folgende Summe

$$\sum_{j=1}^n \lambda_j v_j = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$$

**Linear Kombination** von  $v_1, \dots, v_n$ .

**Definition B.9** Es sei  $V$  ein  $K$ -Vektorraum und seien  $v_1, \dots, v_n \in V$  ( $n \in \mathbb{N}$ ). Dann definieren wir

$$\text{Span}(v_1, \dots, v_n) := \left\{ v \in V : v = \sum_{j=1}^n \lambda_j v_j \text{ mit } \lambda_1, \dots, \lambda_n \in K \right\},$$

und nennen  $\text{Span}(v_1, \dots, v_n)$  den von  $v_1, \dots, v_n$  **aufgespannten Raum**.

**Satz B.10** *Es sei  $V$  ein  $K$ -Vektorraum und seien  $v_1, \dots, v_n \in V$  ( $n \in \mathbb{N}$ ). Dann ist  $\text{Span}(v_1, \dots, v_n)$  ein Untervektorraum von  $V$ .*

## C Mengen und Abbildungen

### Definition von Mengen und bekannte Zahlenmengen

Wir starten mit der grundlegenden Definition einer Menge.

**Definition C.1** Eine **Menge**  $M$  ist eine Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen.

Ein solches Objekt  $x$  heißt **Element** der Menge  $M$ , Schreibweise:

$$x \in M.$$

Ist  $x$  nicht Element von  $M$ , so schreiben wir

$$x \notin M.$$

Die Menge ohne Elemente heißt die **leere Menge**, Schreibweise:

$$\emptyset \text{ oder } \{\}.$$

Es gibt verschiedene Möglichkeiten der Darstellung von Mengen, etwa die aufzählende Schreibweise

$$M := \{x_1, x_2, x_3, \dots, x_n\}$$

wobei  $x_1, x_2, x_3, \dots, x_n$  die Elemente von  $M$  sind. Oder die beschreibende Schreibweise, also eine Charakterisierung der Elemente. Die beschreibende Schreibweise hat allgemein die Form

$$M := \{x : x \text{ hat die Eigenschaft } E\},$$

wobei  $E$  irgendeine „Eigenschaft“ ist. Alternativ schreibt man statt  $x$  : auch  $x|$ .

#### Beispiel C.2

$$M = \{1, 2, 3, 4\}$$

$$M = \{x \in \mathbb{N} : 1 \leq x \leq 4\}.$$

Wir setzen die Kenntnis der folgenden üblichen Zahlenmengen voraus.

$$\begin{aligned}
\mathbb{N} &:= \{x : x \text{ natürliche Zahl}\} \\
\mathbb{N}_0 &:= \{x : x \text{ natürliche Zahl oder } x = 0\} \\
\mathbb{Z} &:= \{x : x \text{ ganze Zahl}\} \\
\mathbb{Q} &:= \{x : x \text{ rationale Zahl}\} \\
\mathbb{R} &:= \{x : x \text{ reelle Zahl}\}.
\end{aligned}
\tag{C.40}$$

## Teilmengen, Schnitt- und Vereinigung von Mengen

**Definition C.3** Es seien  $A, B$  Mengen.

1.  $A$  heißt **Teilmenge** von  $B$ , falls aus  $x \in A$  auch  $x \in B$  folgt. Schreibweise:

$$A \subset B \text{ oder auch } A \subseteq B$$

2.  $A$  und  $B$  heißen **gleich**, falls  $A \subset B$  und  $B \subset A$  gilt. Schreibweise

$$A = B$$

3. Die Menge

$$B \setminus A := \{x : x \in B \text{ und } x \notin A\}$$

heißt **Differenz** von  $B$  und  $A$ . Ist  $A \subset B$ , so heißt

$$A^c := A^{c(B)} := B \setminus A$$

**Komplement** von  $A$  bezüglich  $B$ . Die Bezeichnung  $A^{c(B)}$  wird jedoch nur verwendet, wenn Uneindeutigkeit bezüglich der Grundmenge  $B$  besteht.

4. Die Menge

$$B \cap A := \{x : x \in A \text{ und } x \in B\}$$

heißt **Schnitt** oder **Schnittmenge** von  $B$  und  $A$ .

5. Die Menge

$$B \cup A := \{x : x \in A \text{ oder } x \in B\}$$

heißt **Vereinigung** von  $B$  und  $A$ .

**Bemerkung C.4** Für Mengen  $M_1, M_2, \dots, M_n$  verwendet man die folgende Notation

$$\bigcap_{j=1}^n M_j = M_1 \cap M_2 \cap \dots \cap M_n$$

und

$$\bigcup_{j=1}^n M_j = M_1 \cup M_2 \cup \dots \cup M_n$$

für Schnitt und Vereinigung. Diese Notation kann auch auf unendlich viele Mengen erweitert werden.

Das folgende Ergebnis fasst grundlegende Rechenregeln für die oben eingeführten Mengenoperationen zusammen.

**Satz C.5** *Es seien  $A, B, C$  Mengen, dann gilt:*

1.  $A \cap B = B \cap A$  und  $A \cup B = B \cup A$  (Kommutativgesetze)
2.  $A \cap (B \cap C) = (A \cap B) \cap C$  und  $A \cup (B \cup C) = (A \cup B) \cup C$  (Assoziativgesetze)
3.  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  und  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$  (Distributivgesetze)

**Satz C.6** (De Morgansche Gesetze) *Es seien  $A, B$  Mengen, dann gilt*

1.  $(A \cap B)^c = A^c \cup B^c$  und
2.  $(A \cup B)^c = A^c \cap B^c$ .

## Kartesisches Produkt

Für die üblichen Zahlenmengen aus (C.40) sind höherdimensionale Konstrukte der Form  $(x_1, x_2)$ ,  $(x_1, x_2, x_3)$ , beispielsweise mit  $x_1, x_2, x_3 \in \mathbb{R}$ , möglicherweise schon bekannt. In der nächsten Definition führen wir diese sogenannten geordneten Tupel für allgemeine Mengen ein.

**Definition C.7** Es seien  $M_1, M_2$  zwei nichtleere Mengen, dann heißt

$$M_1 \times M_2 := \{(x_1, x_2) : x_1 \in M_1, x_2 \in M_2\}$$

(Menge der geordneten 2-Tupel) das **kartesische Produkt** von  $M_1$  und  $M_2$ .  
Seien allgemeiner mehrere nichtleere Mengen  $M_1, M_2, M_3, \dots, M_n$  gegeben, dann heißt

$$M_1 \times M_2 \times \dots \times M_n := \{(x_1, x_2, \dots, x_n) : x_j \in M_j, j = 1, \dots, n\}$$

das **kartesische Produkt** der Mengen  $M_1, M_2, M_3, \dots, M_n$ .

Gilt weiter  $M := M_1 = M_2 = \dots = M_n$ , so schreibt man auch

$$M^n = \underbrace{M \times M \times \dots \times M}_{n\text{-mal}}.$$

**Beispiel C.8** 1. Für eine natürliche Zahl  $n \in \mathbb{N}$  ist

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_j \in \mathbb{R}, j = 1, \dots, n\}.$$

2. Für  $M = \mathbb{Z} \setminus \mathbb{N}_0$  ist  $M^2 = M \times M$  die Menge der geordneten zwei-Tupel negativer ganzer Zahlen.

## Abbildungen (Funktionen)

Ähnlich wie bei der obigen Einführung von Mengen wollen wir auf eine eher intuitive Definition des zweiten grundlegenden Begriffes der Mathematik zurückgreifen, nämlich den einer Abbildung (oder Funktion).

(Im Zusammenhang mit Abbildungen sind hier die Variablennamen  $X, Y, Z$  für Mengen bevorzugt)

**Definition C.9** Es seien  $X$  und  $Y$  nichtleere Mengen. Eine **Abbildung** (oder **Funktion**)  $f$  von  $X$  nach  $Y$  ist eine „Vorschrift“, die jedem  $x \in X$  *genau ein* Element  $f(x) \in Y$  zuordnet. Dabei heißen

$X$  der **Definitionsbereich**

und

$Y$  der **Zielbereich**

von  $f$ .

Man schreibt

$$f : X \rightarrow Y$$

und formuliert die Vorschrift  $f$  gewöhnlich durch mathematische Formeln, an manchen Stellen aber auch durch eine wörtliche Beschreibung. Alternativ schreibt man zur Formulierung der Vorschrift auch

$$x \mapsto f(x).$$

**Bemerkung C.10** An anderer Stelle (z.B. theoretische Informatik) betrachtet man Abbildungen

$$f : X \rightarrow Y$$

die nicht auf der gesamten Menge  $X$  ausgewertet werden können. In diesem Fall ist der Definitionsbereich eine Teilmenge von  $X$ , nämlich der Bereich wo die Funktion ausgewertet werden kann. Dies kann zum Beispiel dann sinnvoll sein, wenn  $f$  einen Algorithmus beschreibt, der für manche Eingabedaten nicht zum Ende kommt und daher hierfür keine definierte Ausgabe erzeugt. Eine solche Eingabe wäre dann ein Element außerhalb des Definitionsbereichs.

**Definition C.11** Für eine Abbildung  $f : X \rightarrow Y$  heißt die Menge

$$\{(x, f(x)) : x \in X\} (\subset X \times Y)$$

der **Graph** der Abbildung  $f$ .



**Definition C.12** Sind  $f, g : X \rightarrow Y$  zwei Abbildungen, so heißen  $f$  und  $g$  **gleich**, falls  $f(x) = g(x)$  für alle  $x \in X$  gilt.

Im Weiteren verwenden wir oft die Schreibweise  $(x \in X)$  anstelle von „für alle  $x \in X$ “.

## Urbild, Bildmenge, Wertebereich

**Definition C.13** Sind  $X, Y$  Mengen und ist  $f : X \rightarrow Y$ , so heißt für  $B \subset Y$

$$f^{-1}(B) := \{x \in X : f(x) \in B\} (\subset X)$$

**Urbildmenge** von  $B$  unter  $f$  und für  $A \subset X$

$$f(A) := \{f(x) : x \in A\} = \{y \in Y : y = f(x) \text{ für ein } x \in A\} (\subset Y)$$

**Bildmenge** von  $A$  unter  $f$ . Speziell heißt

$$W(f) := f(X)$$

**Wertebereich** von  $f$ . Ist  $W(f)$  einpunktig, so heißt  $f$  **konstant**.

**Beispiel C.14** Es seien  $X := Y := \mathbb{N}$ , und es sei  $f : \mathbb{N} \rightarrow \mathbb{N}$  definiert durch

$$f(x) := \begin{cases} x, & \text{falls } x \text{ gerade} \\ 2x, & \text{falls } x \text{ ungerade} \end{cases}.$$

Dann gilt

$$f^{-1}(\{2, 4, 6\}) = f^{-1}(\{1, 2, 3, 4, 5, 6\}) = \{1, 2, 3, 4, 6\}$$

und

$$f(\{1, 2, 3\}) = \{2, 6\}.$$

Außerdem ist  $W(f) = \{y \in \mathbb{N} : y \text{ gerade}\}$ .

## Surjektiv, Injektiv, Bijektiv, Verkettung und Umkehrabbildung

**Definition C.15** Es seien  $X, Y$  Mengen. Eine Abbildung  $f : X \rightarrow Y$  heißt

1. **surjektiv**, falls für alle  $y \in Y$  ein  $x \in X$  so existiert, dass  $f(x) = y$  (d.h.  $W(f) = Y$ )
2. **injektiv**, falls gilt: sind  $x_1, x_2 \in X$  mit  $x_1 \neq x_2$ , so ist  $f(x_1) \neq f(x_2)$ ,
3. **bijektiv**, falls  $f$  injektiv und surjektiv ist.

**Definition C.16** Es seien  $X, Y, Z$  Mengen und  $f : X \rightarrow Y$  sowie  $g : Y \rightarrow Z$  Abbildungen. Dann heißt  $g \circ f : X \rightarrow Z$ , definiert durch

$$(g \circ f)(x) := g(f(x)) \quad (x \in X)$$

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

**Beispiel C.17** Sind  $X = Y = Z = \mathbb{N}$  und  $f, g : \mathbb{N} \rightarrow \mathbb{N}$  definiert durch

$$f(x) := x^2, \quad g(x) := x + 1 \quad (x \in \mathbb{N}),$$

so ist  $g \circ f : \mathbb{N} \rightarrow \mathbb{N}$  gegeben durch

$$(g \circ f)(x) = x^2 + 1 \quad (x \in \mathbb{N}).$$

Man beachte: Hier ist auch  $f \circ g : \mathbb{N} \rightarrow \mathbb{N}$  definiert und es gilt

$$(f \circ g)(x) = (x + 1)^2 \quad (x \in \mathbb{N}).$$

Dabei ist  $g \circ f \neq f \circ g$  (da etwa  $(g \circ f)(1) = 2 \neq 4 = (f \circ g)(1)$ ).

**Satz C.18** Es seien  $X, Y, Z$  Mengen und  $f : X \rightarrow Y$  und  $g : Y \rightarrow Z$  Abbildungen.

1. Sind  $f, g$  injektiv, dann ist auch die Komposition  $g \circ f : X \rightarrow Z$  injektiv.
2. Sind  $f, g$  surjektiv, dann ist auch die Komposition  $g \circ f : X \rightarrow Z$  surjektiv.

**Definition C.19** Für eine nichtleere Menge  $X$  heißt  $\text{id}_X : X \rightarrow X$ , definiert durch  $\text{id}(x) := \text{id}_X(x) := x$  ( $x \in X$ ), die **identische Abbildung** auf  $X$ .

**Definition C.20** Es seien  $X, Y$  Mengen und es sei  $f : X \rightarrow Y$  eine Abbildung. Die Abbildung  $f^{-1} : Y \rightarrow X$  heißt **Umkehrabbildung** von  $f$  oder **inverse Abbildung** (bezüglich  $\circ$ ) von  $f$ , falls

$$f^{-1} \circ f = \text{id}_X, \text{ und } f \circ f^{-1} = \text{id}_Y$$

gilt.

**Satz C.21** Es seien  $X, Y$  Mengen und es sei  $f : X \rightarrow Y$  bijektiv. Dann existiert zu jedem  $y \in Y$  genau ein  $x \in X$  mit  $f(x) = y$ . Damit ergibt sich die inverse Abbildung durch

$$f^{-1}(y) := x \quad (y \in Y),$$

wobei  $y = f(x)$ .

## Mächtigkeit von Mengen, Potenzmenge

Für endliche Mengen ist es offensichtlich, dass die Anzahl der Elemente Auskunft über die „Größe“ der Menge gibt.

Für eine Menge  $M = \{x_1, \dots, x_n\}$  ( $n \in \mathbb{N}$ ) setzen wir

$$|M| := n.$$

Haben jedoch zwei Mengen unendlich viele Elemente, so ist nicht klar wie man allgemein die „Größe“ solcher Mengen vergleichen kann. In der Mathematik gibt es hierzu einige Zugänge, die jeweils in ihrer Hinsicht etwas über die „Größe“ aussagen. Wir können hier jedoch nur den elementaren Begriff der Mächtigkeit einführen.

**Definition C.22** Es seien  $M$  und  $N$  zwei Mengen, dann heißen diese **gleichmächtig**, falls eine bijektive Abbildung  $f : M \rightarrow N$  (oder äquivalent  $f : N \rightarrow M$ ) existiert.

1. Wenn  $M$  und  $\{1, \dots, n\}$  für ein  $n \in \mathbb{N}$  gleichmächtig sind, dann heißt  $M$  **endliche Menge** und man sagt, dass  $M$  die **Kardinalität**  $n$  hat (gleichbedeutend  $|M| = n$ ).
2. Wenn  $M$  und  $\mathbb{N}$  gleichmächtig sind, dann heißt  $M$  **abzählbar unendlich** oder kurz **abzählbar**.
3. Wenn  $M$  endlich oder abzählbar unendlich ist, dann sagt man auch  $M$  ist **höchstens abzählbar**.
4. Wenn  $M$  nicht höchstens abzählbar ist, dann heißt  $M$  **überabzählbar**.

**Bemerkung C.23** Mit Hilfe von Satz C.18 folgt leicht das Folgende. Sind  $M_1, M_2, M_3$  Mengen mit so, dass  $M_1, M_2$  gleichmächtig sind und  $M_2, M_3$  gleichmächtig sind, dann sind auch  $M_1, M_3$  gleichmächtig.

**Definition C.24** Es sei  $M$  eine Menge, dann heißt

$$\mathcal{P}(M) := 2^M := \{N : N \subset M\}$$

die **Potenzmenge** von  $M$ .

Betrachten wir  $M = \{1, \dots, n\}$  und  $N \subset M$ , dann definiert

$$\begin{aligned} \varphi_N : \{1, \dots, n\} &\rightarrow \{0, 1\} \\ \varphi_N(x) &:= \begin{cases} 1, & \text{falls } x \in N \\ 0, & \text{falls } x \notin N \end{cases} \end{aligned}$$

eine eindeutige Zuordnung  $N \mapsto (\varphi_N(1), \varphi_N(2), \dots, \varphi_N(n)) \in \{0, 1\}^n$ . Diese Zuordnung definiert eine bijektive Abbildung

$$f : \{N : N \subset M\} \rightarrow \{0, 1\}^n, \quad N \mapsto (\varphi_N(1), \dots, \varphi_N(n)),$$

und somit gilt

$$|\{N : N \subset M\}| = |\{0, 1\}^n| = 2^n.$$

Dieser Sachverhalt für endliche Mengen ist der Grund für die Namensgebung in der oberen Definition.

**Satz C.25** *Es sei  $M$  eine Menge mit  $|M| = n$  mit  $n \in \mathbb{N}_0$ , dann gilt*

$$|\mathcal{P}(M)| = 2^n.$$

## D Komplexe Zahlen

Wir haben oben gesehen, dass das Lösen der Gleichung

$$4 \cdot x = 1$$

einer Erweiterung der Gruppe  $(\mathbb{Z}, +)$  zum Körper  $(\mathbb{Q}, +, \cdot)$  bedurfte. Weiter bedurfte das Lösen der Gleichung

$$x^2 = 2$$

einer Erweiterung von  $(\mathbb{Q}, +, \cdot)$  nach  $(\mathbb{R}, +, \cdot)$ .

In diesem Abschnitt wollen wir  $(\mathbb{R}, +, \cdot)$  in einer Weise erweitern, dass die Gleichung

$$x^2 = -1 \tag{D.41}$$

lösbar ist. Dies führt zum Körper der **komplexen Zahlen**. Dazu definieren wir das Folgende:

**Definition D.1** Die **imaginäre Einheit**  $i$  ist definiert als die Lösung der Gleichung (D.41), d.h. es gilt

$$i^2 = -1.$$

Eine **komplexe Zahl** ist ein Element  $(x, iy)$  mit  $x, y \in \mathbb{R}$ , und wir setzen dafür  $(x, iy) =: x + iy$ . Für  $z := x + iy$  heißen:

1.  $\operatorname{Re}(z) := x$  der **Realteil** von  $z$ ,
2.  $\operatorname{Im}(z) := y$  der **Imaginärteil** von  $z$ .

Wir setzten

$$\mathbb{C} := \{z = x + iy : x \in \mathbb{R}, y \in \mathbb{R}\},$$

die Menge der komplexen Zahlen.

Wie zu Beginn des Abschnitts bemerkt, wollen wir  $\mathbb{C}$  zu einem Körper machen, und benötigen daher entsprechende algebraische Operationen  $+$  und  $\cdot$ .

**Definition D.2** Für  $z = x + iy$  und  $w = u + iv$  aus  $\mathbb{C}$  definieren wir

1.  $z + w := x + u + i(y + v)$  **Addition,**
2.  $zw := z \cdot w := xu + i(xv + uy) - vy$  **Multiplikation.**

Damit gilt:

**Satz D.3**  $(\mathbb{C}, +, \cdot)$  ist ein Ring mit neutralem Element  $0 + i0$  bezüglich der Addition  $+$ , und  $1 + i0$  als neutralem Element bezüglich der Multiplikation  $\cdot$ .

Der Kürze halber schreiben wir auch

$$x + i0 = x, \quad 0 + iy = iy, \quad 0 + i0 = 0.$$

Die Erweiterung zum Körper ergibt sich nun dadurch, dass man für alle  $z \in \mathbb{C} \setminus \{0\}$  ein inverses Element bezüglich der Multiplikation finden kann:

**Satz und Definition D.4** Es sei  $z = x + iy \in \mathbb{C} \setminus \{0\}$ , und

$$z^{-1} := \frac{1}{z} = \frac{x}{x^2 + y^2} + i \frac{-y}{x^2 + y^2},$$

dann gilt

$$z \cdot z^{-1} = 1.$$

Somit ist  $z^{-1}$  das inverse Element von  $z$  bezüglich  $\cdot$ . Damit ist  $(\mathbb{C}, +, \cdot)$  ein Körper.

Für  $z, w \in \mathbb{C}$ ,  $w \neq 0$  schreiben wir auch

$$zw^{-1} = z \frac{1}{w} = \frac{z}{w}.$$

Damit können wir mit den komplexen Zahlen rechnen wie in  $\mathbb{R}$  gewohnt:

$$\begin{aligned} \frac{1}{z} + \frac{1}{w} &= \frac{z+w}{zw} & (z, w \neq 0), \\ \frac{1}{z} \cdot \frac{1}{w} &= \frac{1}{zw} & (z, w \neq 0), \\ \frac{z+w}{w} &= \frac{z}{w} + 1 & (w \neq 0). \end{aligned}$$

**Definition D.5** Es sei  $z = x + iy$  eine komplexe Zahl.

1. Die komplexe Zahl  $\bar{z} := x - iy$  heißt zu  $z$  **konjugiert komplex**.
2. Die Zahl  $|z| := \sqrt{x^2 + y^2} \in [0, \infty)$  heißt **Betrag** von  $z$ .

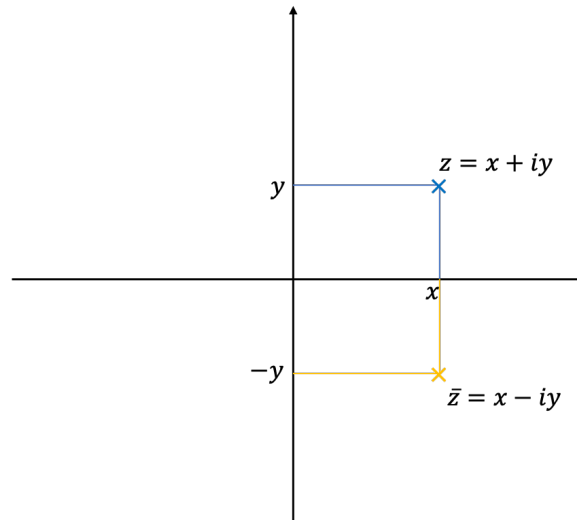


Figure 13: Gaußsche Zahlenebene (oder komplexe Ebene)

**Bemerkung D.6** Für  $z, w \in \mathbb{C}$  ergibt sich leicht

$$\overline{z + w} = \bar{z} + \bar{w}, \quad \overline{zw} = \bar{z} \cdot \bar{w}, \quad \overline{(\bar{z})} = z,$$

sowie

$$\operatorname{Re}(z) = \frac{1}{2}(z + \bar{z}) \quad \text{und} \quad \operatorname{Im}(z) = \frac{1}{2i}(z - \bar{z}).$$

Für den Betrag ergeben sich folgenden Rechenregeln:



**Satz D.7** Für  $z, w \in \mathbb{C}$  gilt

1.  $|z| \geq 0$ , und dabei ist  $|z| = 0$  genau dann, wenn  $z = 0$  ist,
2.  $|z| = |\bar{z}|$ ,  $|z| = |-z|$ ,  $|\operatorname{Re} z| \leq |z|$ ,  $|\operatorname{Im} z| \leq |z|$ ,
3.  $|z|^2 = z\bar{z}$  und  $1/z = \bar{z}/|z|^2$ , falls  $z \neq 0$ ,
4.  $|zw| = |z||w|$  und  $|z + w|^2 = |z|^2 + 2\operatorname{Re}(z\bar{w}) + |w|^2$ ,
5. (Dreiecksungleichung)  $|z \pm w| \leq |z| + |w|$ .

**Beispiel D.8** Für  $z = 3 - i$  gilt

$$|z| = \sqrt{9 + 1} = \sqrt{10}, \quad \bar{z} = 3 - i(-1) = 3 + i$$

und

$$z\bar{z} = (3 - i)(3 + i) = 9 + 1 (= |z|^2).$$

Wir kommen nun zu einem zentralen Ergebnis der Mathematik. Wie zu Beginn des Abschnitts erwähnt, haben die bisherigen Erweiterungen von  $\mathbb{Z}$  über  $\mathbb{Q}$  nach  $\mathbb{R}$  und schließlich nach  $\mathbb{C}$  uns ermöglicht Gleichungen der Form

$$4 \cdot x = 1, \quad x^2 = 2, \quad x^2 = -1$$

zu lösen. Wir werden nun sehen, dass die komplexen Zahlen ein natürliches Ende dieser Erweiterungen darstellt, denn hierin sind sämtliche Kombinationen solcher Gleichungen stets lösbar. Dazu zunächst folgende Definition

**Definition D.9** Eine **Polynomfunktion** (oder kurz **Polynom**) ist eine Funktion  $P : \mathbb{C} \rightarrow \mathbb{C}$  der Form

$$P(z) = \sum_{\nu=0}^d a_{\nu} z^{\nu}$$

mit  $a_0, \dots, a_d \in \mathbb{C}$ . Ist  $a_d \neq 0$ , so heißt  $\deg(P) := d$  der **Grad** von  $P$  und  $a_0, \dots, a_d$  die **Koeffizienten** von  $P$ . Ein  $z \in \mathbb{C}$  heißt **Nullstelle** von  $P$ , falls  $P(z) = 0$  gilt.

**Satz D.10** (*Fundamentalsatz der Algebra*) Es sei

$$P(z) = \sum_{\nu=0}^d a_{\nu} z^{\nu}$$

ein Polynom vom Grad  $d > 0$ . Dann existieren  $z_1, \dots, z_k \in \mathbb{C}$  und zugehörige  $\alpha_1, \dots, \alpha_k \in \mathbb{N}$ , so dass

$$P(z) = a_d(z - z_1)^{\alpha_1} (z - z_2)^{\alpha_2} \cdot \dots \cdot (z - z_k)^{\alpha_k}$$

gilt, dabei ist  $\alpha_1 + \alpha_2 + \dots + \alpha_k = d$ .

**Definition D.11** Die  $\alpha_1, \alpha_2, \dots, \alpha_k$  in Satz D.10 heißen **Vielfachheiten** der Nullstellen  $z_1, z_2, \dots, z_k$ .

**Beispiel D.12** 1. Das Polynom  $P(z) = z^2 + 1 = (z + i)(z - i)$  hat einfache Nullstellen an  $i$  und  $-i$

2. Das Polynom  $P(z) = z^3 + z^2 + 4z + 4$  hat nur eine einfache reelle Nullstelle an  $z = -1$ . Nach dem dem Fundamentalsatz der Algebra (Satz D.10) muss  $P$  noch zwei weitere Nullstellen (inklusive Vielfachheit) haben. Es gilt

$$P(z) = (z + 1)(z + 2i)(z - 2i),$$

woran man erkennt, dass  $P$  die beiden komplexen Nullstellen an  $z = 2i$  und  $z = -2i$  hat.

3. Das Polynom  $P(z) = z^3 + iz^2 + 4z + i4 = (z + i)(z - 2i)(z + 2i)$  hat ausschließlich komplexe Nullstellen an  $z = -i$ ,  $z = 2i$  und  $z = -2i$ .

In dem oberen Beispiel beobachtet man, dass die Nullstellen von Polynomen mit reellen Koeffizient stets symmetrisch bezüglich der reellen Achse in  $\mathbb{C}$  liegen. Dies gilt auch im Allgemeinen:

**Satz D.13** *Es sei*

$$P(z) = \sum_{\nu=0}^d a_{\nu} z^{\nu}$$

*ein Polynom mit  $a_{\nu} \in \mathbb{R}$  für  $\nu = 0, \dots, d$ . Dann gilt  $P(z) = 0$  genau dann, wenn  $P(\bar{z}) = 0$  gilt.*

# Index

Abbildung, 98  
abzählbar, 102  
abzählbar unendlich, 102  
Addition, 4, 105  
Anzahl der Stufen, 44  
aufgespannten Raum, 6, 92  
  
Basis, 30  
Betrag, 106  
bijektiv, 100  
Bild, 36  
Bildmenge, 99  
Blockmatrix, 72  
  
charakteristische Polynom, 77  
  
darstellende Matrix, 35  
Definitionsbereich, 98  
Determinante, 67  
Diagonalmatrix, 8  
Differenz, 95  
Dimension, 33  
direkte Summe, 49  
  
Eigenraum, 74  
Eigenvektor, 73  
Eigenwert, 73  
Einheitsmatrix, 15  
Element, 94  
elementare Zeilenumformungen, 41  
Elementarmatrizen, 42  
endliche Menge, 102  
euklidische Norm, 20  
euklidisches Skalarprodukt, 18  
  
Fehlstand, 66  
Funktion, 98  
  
Gauß-Jordan-Verfahren, 47  
gleich, 95, 99  
gleichmächtig, 102  
Grad, 107  
Graph, 98  
  
Halbräume, 25  
Hessesche Normalform, 25  
Hintereinanderausführung, 100  
Hyperebene, 25  
höchstens abzählbar, 102  
  
identische Abbildung, 101  
imaginäre Einheit, 104  
Imaginärteil, 104  
injektiv, 100  
inverse Abbildung, 101  
inverse Matrix, 38  
invertierbar, 38  
isomorph, 34  
Isomorphismus, 34  
  
kanonisches Skalarprodukt, 18, 64  
Kardinalität, 102  
kartesische Produkt, 97  
Kern, 36  
Koeffizienten, 107  
Komplement, 95  
komplexe, 104  
komplexen Zahlen, 104  
Komponenten, 3  
Komponentenschreibweise, 3

- Komposition, 100
- konjugiert komplex, 106
- konstant, 99
- Koordinaten, 30
  
- leere Menge, 94
- Leibniz-Formel, 67
- linear, 33
- linear abhängig, 28
- Linear Kombination, 92
- linear least squares, 60
- linear unabhängig, 28
- lineare Hülle, 6
- linearen Gleichungssystemen (LGS), 40
- lineares kleinste Quadrate Problem, 60
- Linearkombination, 5
- Lösungsmenge, 41
  
- Matrix, 7
- Matrix-Matrix-Produkt, 14
- Matrix-Vektor-Produkt, 10
- Matrizen, 7
- Menge, 94
- Multiplikation, 105
  
- neutrale, 38
- Normalenvektor, 25
- normiert, 54
- Nullstelle, 107
- Nullvektor, 3
  
- obere Dreiecksmatrix, 8
- Optimierungsproblem, 60
- orthogonal, 21
- Orthogonalbasis, 54
- orthogonale Basis, 54
- orthogonale Projektion, 57
  
- orthogonales Komplement, 51
- Orthogonalraum, 51
- Orthonormalbasis, 54
- Orthonormalmatrix, 55
  
- Permutation, 66
- Polynom, 107
- Polynomfunktion, 107
- Potenzmenge, 102
  
- quadratische, 7
  
- Rang, 36
- Realteil, 104
  
- Schnitt, 95
- Schnittmenge, 95
- Signum, 66
- skalare Multiplikation, 4
- Spaltenvektoren, 8
- Standardbasis, 27
- surjektiv, 100
- symmetrisch, 17
  
- Teilmenge, 95
- Transponierte, 16
- transponierte Matrix, 16
  
- Umkehrabbildung, 101
- unter Dreiecksmatrix, 9
- Untervektorraum, 26, 91
- Urbildmenge, 99
  
- Vektoren, 3
- Vektorraum, 26, 89
- Vereinigung, 95
- Verkettung, 100
- Vielfachheiten, 108

Vorzeichen, [66](#)

Wertebereich, [99](#)

Winkel, [21](#)

Zeilenstufenform, [43](#)

Zeilenvektoren, [8](#)

Zielbereich, [98](#)

überabzählbar, [102](#)