

Lineare Algebra für Lehramt

Vorlesung von Johann Linhart im SS 2005

Inhalt

0 Vorbemerkungen	5
1 Der Vektorraum \mathbb{R}^n	6
2 Matrizen und lineare Abbildungen	13
3 Addition und Multiplikation von Matrizen	20
4 Lineare Gleichungssysteme	25
4.1 Elementare Umformungen	25
4.2 Halbdiagonalform	26
4.3 Bestimmung der Lösungsmenge	30
5 Lineare und affine Teilräume des \mathbb{R}^n	37
5.1 Definitionen	37
5.2 Basis, lineare Unabhängigkeit und Dimension	38
5.2.1 Fundamentale Begriffe und Sätze	38
5.2.2 Berechnung der Koordinaten bezüglich einer gegebenen Basis	46
5.2.3 Dimension affiner Teilräume und lineare Gleichungssysteme	47
5.3 Ergänzungen zum Rang einer Matrix	49
5.4 Schnitt von affinen Teilräumen	52
5.4.1 Schnitt einer Geraden mit einer Hyperebene	52
5.4.2 Schnitt von zwei beliebigen affinen Teilräumen	53
5.4.3 Berechnung eines Gleichungssystems zu gegebener Parameterdarstellung	54
6 Inversion von Matrizen	57

INHALT	3
7 Inneres Produkt und Orthogonalität	61
7.1 Das innere Produkt und die Norm	61
7.2 Orthogonalität	67
7.2.1 Orthogonalsysteme	67
7.2.2 Orthogonale Projektion	70
7.2.3 Das Schmidt'sche Orthonormalisierungsverfahren . . .	71
7.2.4 Normalvektoren von Hyperebenen	74
7.2.5 Spiegelung an einem linearen oder affinen Teilraum .	75
7.2.6 Abstand eines Punktes von einem linearen oder affinen Teilraum	76
7.2.7 Orthogonale Matrizen	77
8 Koordinatentransformationen	80
8.1 Umrechnung der Koordinaten eines Punktes oder Vektors . . .	80
8.2 Umrechnung der Matrix einer linearen Abbildung	82
8.2.1 Matrix eines Endomorphismus in Bezug auf eine Basis	82
8.2.2 Drehung im \mathbb{R}^3 um eine gegebene Achse	84
9 Determinanten	86
9.1 Definition und grundlegende Eigenschaften	86
9.2 Geometrische Bedeutung der Determinante im \mathbb{R}^2	95
9.2.1 Orientierung zweier Vektoren im \mathbb{R}^2	95
9.2.2 Flächeninhalt eines Parallelogramms	96
10 Äußeres Produkt im \mathbb{R}^3	97
10.1 Definition und grundlegende Eigenschaften	97
10.2 Geometrische Bedeutung von Vektorprodukt und Determinante im \mathbb{R}^3	100
10.2.1 Orientierung im \mathbb{R}^3	100
10.2.2 Volumen eines Parallelepipedes	102

11 Eigenvektoren und Eigenwerte	103
12 Quadratische Formen	107
12.1 Definition	107
12.2 Koordinatentransformation	109
12.3 Definite quadratische Formen	110

0 Vorbemerkungen

Datum der letzten Änderung: July 6, 2005

Da sich diese Vorlesung an Lehramtskandidaten richtet, wird immer wieder auf die Lehrpläne an (österreichischen) "Höheren Schulen" Bezug genommen, und zwar vor allem auf die der AHS-Oberstufe, wo die lineare Algebra am stärksten vertreten ist. Es ist dabei allerdings zu beachten, dass sich die Lehrpläne immer wieder ändern. Wenn Sie sich also jetzt z.B. im 2. Semester Ihres Studiums befinden, so ist es sehr leicht möglich, dass die Lehrpläne bereits zu Beginn Ihrer Berufstätigkeit als Lehrer deutlich anders aussehen. Für die Lehrerausbildung ist daher zweierlei von Bedeutung:

1. Der Vorlesungsstoff muss über den Schulstoff hinausgehen.
2. Da aber nicht vorherzusehen ist, welche Inhalte in Zukunft im Schulunterricht vorkommen werden, muss der Hochschulunterricht die Fähigkeit vermitteln, sich selbst weitere (mathematische) Kenntnisse anzueignen.

Da die "Lineare Algebra für Lehramt" inklusive Übungen nur eine dreistündige Lehrveranstaltung ist, kann der erste Punkt hier natürlich nur sehr begrenzt realisiert werden. Dem zweiten Punkt wird u.a. dadurch Rechnung getragen, dass versucht wird, die mathematische Denkweise auf universitärer Ebene den Hörern nahezubringen. Dazu gehört z.B. ein gewisses Training in abstrakt-logischem Denken und die Unterscheidung zwischen diesem Denken und den damit verbundenen anschaulichen Vorstellungen.

Es gibt zahlreiche Bücher über lineare Algebra. Von den deutschsprachigen sind einige im Literaturverzeichnis angeführt ([3], [4], [5], [7]), von denen sich das letztgenannte speziell an Lehramtsstudenten und Lehrer wendet.

1 Der Vektorraum \mathbb{R}^n

Sei n eine natürliche Zahl. Mit \mathbb{R}^n bezeichnen wir die Menge aller n -Tupel von reellen Zahlen, d.h.

$$\mathbb{R}^n := \{(x_1, \dots, x_n) : x_i \in \mathbb{R} \text{ für alle } i \in \{1, \dots, n\}\}.$$

Wenn $x = (x_1, \dots, x_n)$, so nennen wir die Zahlen x_1, \dots, x_n auch *Koordinaten* von x .

Die Elemente des \mathbb{R}^n werden oft in besonderer Weise geschrieben, um sie deutlicher von Zahlen zu unterscheiden, z.B. \mathbf{x} , \underline{x} , \vec{x} , \mathfrak{x} (und analog mit anderen Buchstaben).

\mathbb{R}^3 ist ein Modell für den gewöhnlichen dreidimensionalen Raum, da bekanntlich nach Festlegung eines Koordinatensystems jeder Punkt des Raumes drei reellen Zahlen entspricht. Daher nennt man \mathbb{R}^n den n -dimensionalen Raum. Zur Unterscheidung von anderen Räumen, die vor allem in der Mathematik und Physik Bedeutung haben, spricht man auch genauer vom n -dimensionalen *euklidischen Raum*.

Die Elemente des \mathbb{R}^n kann man sich als *Punkte* oder als *Vektoren* ("gerichtete Strecken" oder "Pfeile") vorstellen, und daher werden die Elemente des \mathbb{R}^n oft auch so bezeichnet.

Von spezieller Bedeutung sind die folgenden Vektoren des \mathbb{R}^n :

$$\begin{aligned} e_1 &:= (1, 0, 0, \dots, 0) \\ e_2 &:= (0, 1, 0, \dots, 0) \\ &\dots \\ e_n &:= (0, 0, \dots, 0, 1) \end{aligned}$$

Sie heißen die *kanonischen Basisvektoren* des \mathbb{R}^n . (Später werden wir noch andere Basisvektoren kennenlernen.)

Eine besondere Rolle spielt der *Nullvektor* $o := (0, \dots, 0)$, der eigentlich keiner gerichteten Strecke entspricht. Stellt man sich o als Punkt vor, so nennt man ihn den *Ursprung*.

Seien p und q zwei Elemente des \mathbb{R}^n , $p = (p_1, \dots, p_n)$, $q = (q_1, \dots, q_n)$. Dann versteht man unter der *Summe* bzw. *Differenz* von p und q in natürlicher Weise folgendes:

$$\begin{aligned} p + q &:= (p_1 + q_1, \dots, p_n + q_n), \\ p - q &:= (p_1 - q_1, \dots, p_n - q_n). \end{aligned}$$

Oft ist für die Elemente des \mathbb{R}^n die *Spaltenschreibweise* zweckmäßiger. Dann sieht etwa $z = x + y$ so aus:

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

Stellen wir uns p und q als Punkte vor, so verstehen wir unter dem *Vektor mit Anfangspunkt p und Endpunkt q* das folgende Element des \mathbb{R}^n :

$$\overrightarrow{pq} := q - p = (q_1 - p_1, \dots, q_n - p_n)$$

Der Vektor mit Anfangspunkt o und Endpunkt p ist also gleich p :

$$\overrightarrow{op} = p.$$

Manchmal nennt man diesen Vektor auch den *Ortsvektor* des Punktes p .

Offensichtlich gilt (für beliebiges $r \in \mathbb{R}^n$)

$$\overrightarrow{pq} + \overrightarrow{qr} = \overrightarrow{pr},$$

und das ergibt eine anschauliche Interpretation der Addition von Vektoren.

Seien $a, p, q \in \mathbb{R}^n$, und $p' = p + a$, $q' = q + a$. Dann ist $\overrightarrow{p'q'} = \overrightarrow{pq}$ und $\overrightarrow{pp'} = \overrightarrow{qq'}$, und wir sagen: p, p', q, q' bilden die Ecken eines *Parallelogramms*. Zwei parallele, gleich lange und gleich gerichtete Strecken stellen also ein und denselben Vektor dar.

Die Vektoren \overrightarrow{pq} und $\overrightarrow{p'q}$ entsprechen den beiden Diagonalen dieses Parallelogramms. Bezeichnen wir den Vektor \overrightarrow{pq} mit b , so gilt

$$\overrightarrow{pq} = a + b,$$

und das ergibt eine zweite anschauliche Interpretation der Summe zweier Vektoren.

Man kann jeden Vektor a von jedem beliebigen Punkt p aus "abtragen", d.h. zu jedem $a \in \mathbb{R}^n$ und jedem $p \in \mathbb{R}^n$ gibt es einen eindeutig bestimmten Punkt $p' \in \mathbb{R}^n$, sodass $a = \overrightarrow{pp'}$, nämlich $p' = p + a$. Das ergibt eine etwas andere Interpretation der Addition im \mathbb{R}^n .

Jedem Vektor $a \in \mathbb{R}^n$ entspricht in umkehrbar eindeutiger Weise eine *Parallelverschiebung* oder *Schiebung*, die man *Translation* um den Vektor a nennt. Es handelt sich dabei um die folgende Abbildung:

$$\tau_a : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto x + a,$$

also $\tau_a(x) = x + a$. (Wir verwenden den gewöhnlichen Pfeil \rightarrow zur Angabe von Definitions- und Wertebereich einer Abbildung, dagegen den mit einem senkrechten Strich versehenen Pfeil \mapsto zur Angabe der Zuordnungsvorschrift. Die obige Definition von τ_a sprechen wir dann etwa so aus: " τ_a ist eine Abbildung vom \mathbb{R}^n in den \mathbb{R}^n , welche jeden Punkt x in $x + a$ überführt" oder etwas kürzer: " τ_a vom \mathbb{R}^n in den \mathbb{R}^n : x geht über in $x + a$." In der Vorlesung "Diskrete Mathematik" wird der Abbildungsbegriff ausführlicher behandelt.)

Die Zusammensetzung zweier Translationen entspricht der Addition der zugehörigen Vektoren:

$$\tau_{a+b} = \tau_a \circ \tau_b.$$

Das folgt aus dem Assoziativ- und Kommutativgesetz der Addition (siehe Seite 9): $\tau_{a+b}(x) = x + (a+b) = x + (b+a) = (x+b) + a = \tau_b(x) + a = \tau_a(\tau_b(x))$.

(Auf diese Weise erhalten wir eine vierte Möglichkeit, sich die Summe von zwei Vektoren vorzustellen.)

Elemente des \mathbb{R}^n kann man in natürlicher Weise mit reellen Zahlen multiplizieren:

$$\lambda x := (\lambda x_1, \dots, \lambda x_n)$$

für $\lambda \in \mathbb{R}$ und $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Wenn man sich x als Vektor vorstellt, so ist λx für $\lambda > 0$ ein Vektor mit der gleichen Richtung wie x , aber einer um den Faktor λ vergrößerten bzw. verkleinerten Länge. Für $\lambda < 0$ erhält man einen Vektor mit entgegengesetzter Richtung, und für $\lambda = 0$ ergibt sich der Nullvektor. Für $\lambda = -1$ erhält man den Vektor $-x := (-x_1, \dots, -x_n)$.

In Zusammenhang mit Vektoren nennt man reelle Zahlen auch *Skalare*.

Sind $x, y \in \mathbb{R}^n$ und $\lambda, \mu \in \mathbb{R}$, so nennt man $\lambda x + \mu y$ eine *Linearkombination* von x und y , und analog für drei oder mehr Vektoren bzw. Skalare.

Wir sehen jetzt z.B., dass man jeden Vektor $x = (x_1, \dots, x_n)$ des \mathbb{R}^n als Linearkombination der kanonischen Basisvektoren ausdrücken kann:

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

Jedem Skalar λ entspricht in folgender Weise eine Abbildung, die man für $\lambda > 1$ auch *Streckung* und für $0 < \lambda < 1$ *Stauchung* nennt:

$$s_\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto \lambda x.$$

Für $\lambda = -1$ nennt man diese Abbildung *Spiegelung am Ursprung*. Für beliebige Werte von λ heißt s_λ *Homothetie (mit Zentrum im Ursprung)*.

Für die Addition von Vektoren und die Multiplikation von Vektoren mit Skalaren gelten die folgenden leicht einzusehenden Rechenregeln:

$$\begin{aligned}
 (a + b) + c &= a + (b + c) && \text{(Assoziativgesetz der Addition)} \\
 a + o &= a \\
 a + (-a) &= o \\
 a + b &= b + a && \text{(Kommutativgesetz der Addition)}
 \end{aligned}$$

$$\begin{aligned}
 \lambda(a + b) &= \lambda a + \lambda b \\
 (\lambda + \mu)a &= \lambda a + \mu a \\
 \lambda(\mu a) &= (\lambda\mu)a \\
 1a &= a
 \end{aligned}$$

Versuchen Sie, diese Regeln durch Zeichnungen zu veranschaulichen!

Die ersten vier dieser Regeln besagen, dass $(\mathbb{R}^n, +)$ eine kommutative Gruppe ist (siehe Bemerkungen Seite 11). Alle zusammen besagen, dass \mathbb{R}^n ein *Vektorraum* ist. Allgemein versteht man unter einem *Vektorraum über einem Körper K* eine Menge V zusammen mit einer Verknüpfung " + " und einer Abbildung

$$s : K \times V \rightarrow V : (\lambda, x) \mapsto \lambda x,$$

sodass $(V, +)$ eine kommutative Gruppe ist und für s die obigen Regeln gelten. Dabei bedeutet natürlich o das neutrale Element von $(V, +)$, $-a$ das zu a inverse Element in $(V, +)$ und 1 das Einselement des Körpers K (das ist das neutrale Element der Gruppe $(K \setminus \{0\}, \cdot)$).

So ist z.B. \mathbb{C}^n analog zum \mathbb{R}^n ein Vektorraum über dem Körper der komplexen Zahlen. Weitere Beispiele für Vektorräume stellen die im Kapitel 5 behandelten linearen Teilräume des \mathbb{R}^n dar.

Einige weitere sehr einfache Regeln, die in jedem Vektorraum gelten (für beliebige $x \in V$ und $\lambda \in K$):

$$\begin{aligned}
 0x &= o \\
 \lambda o &= o \\
 (-1)x &= -x
 \end{aligned}$$

Der \mathbb{R}^n dient in vielen verschiedenen Anwendungsbereichen als Modell von Objekten, die man durch n Zahlen beschreiben kann. Zum Beispiel kann man eine Preisliste für n Waren als Element des \mathbb{R}^n auffassen. Die Multiplikation

dieses Vektors mit dem Skalar 1.05 entspricht dann einer Preiserhöhung von 5%.

In einem Computer werden die Elemente des \mathbb{R}^n meist durch eindimensionale *Arrays* (der Länge n) dargestellt, manchmal auch durch verkettete Listen.

Sehen wir uns zum Schluss noch Zusammensetzungen von Translationen mit Streckungen oder Stauchungen an. Sei $a \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ und $f := \tau_a \circ s_\lambda$, d.h.

$$f(x) = \lambda x + a.$$

Für $\lambda = 1$ ist das nichts anderes als die Translation τ_a . Für $\lambda \neq 1$ besitzt diese Abbildung jedoch einen eindeutig bestimmten *Fixpunkt*, d.h. einen Punkt, der durch f auf sich selbst abgebildet wird. So ein Fixpunkt ist nämlich eine Lösung der Gleichung

$$\lambda x + a = x,$$

und für $\lambda \neq 1$ ergibt das

$$x = \frac{1}{1 - \lambda}a.$$

Bezeichnen wir diesen Fixpunkt mit p , so gilt für jeden Punkt $x \in \mathbb{R}^n$:

$$p + \lambda(x - p) = \frac{1}{1 - \lambda}a + \lambda(x - \frac{1}{1 - \lambda}a) = \lambda x + a = f(x),$$

d.h.

$$f(x) - p = \lambda(x - p)$$

bzw.

$$\overrightarrow{p, f(x)} = \lambda \overrightarrow{px}.$$

Daher nennt man so eine Abbildung *Homothetie mit Zentrum p (und Streckungsfaktor λ)*. Für $\lambda = -1$ spricht man auch von einer Spiegelung am Punkt p .

Umgekehrt kann man zu gegebenem Zentrum p und Faktor λ leicht den Translationsvektor a ausrechnen:

$$f(x) = p + \lambda(x - p) = \lambda x + (1 - \lambda)p = \lambda x + a \text{ mit } a = (1 - \lambda)p.$$

Für $0 < \lambda < 1$ sieht man anschaulich, dass $f(x)$ ein Punkt auf der Verbindungsstrecke von p und x ist, der diese Strecke im Verhältnis λ zu $1 - \lambda$ teilt. Speziell für $\lambda = 1/2$ erhält man den *Mittelpunkt* von p und x :

$$M(p, x) := \frac{1}{2}(p + x).$$

Bemerkungen zu den Begriffen "Gruppe" und "Körper":

Unter einer *Gruppe* versteht man eine Menge G zusammen mit einer Abbildung $\circ : G \times G \rightarrow G : (a, b) \mapsto a \circ b$, sodass gilt:

1. $(a \circ b) \circ c = a \circ (b \circ c)$ für beliebige $a, b, c \in G$ (*Assoziativgesetz*).
2. Es gibt ein $e \in G$, sodass $a \circ e = e \circ a = a$ für alle $a \in G$.

(Dieses Element e ist eindeutig bestimmt und heißt das *neutrale Element* von G .)

3. Zu jedem $a \in G$ gibt es ein $\bar{a} \in G$, sodass $a \circ \bar{a} = \bar{a} \circ a = e$.

(Dieses \bar{a} ist ebenfalls eindeutig bestimmt und heißt das zu a *inverse Element*.)

Etwas exakter sagt man: Eine Gruppe ist ein geordnetes Paar (G, \circ) , bestehend aus einer Menge G und einer Abbildung $\circ : G \times G \rightarrow G$, sodass die obigen drei Eigenschaften erfüllt sind.

Wenn außerdem $a \circ b = b \circ a$ für beliebige $a, b \in G$, so spricht man von einer *kommutativen* oder *abelschen* Gruppe. (Niels Henrik Abel (1802-1829) ist ein berühmter norwegischer Mathematiker. Er hat insbesondere bewiesen, dass Gleichungen 5. Grades im Allgemeinen nicht auflösbar sind (ungefähr gleichzeitig mit dem französischen Mathematiker Evariste Galois (1811 - 1832), aber unabhängig von ihm).)

Die Abbildung \circ nennt man *Verknüpfung*. Sie wird oft auch mit einem Punkt bezeichnet oder nur durch Nebeneinanderschreiben angedeutet. Das zu a inverse Element \bar{a} bezeichnet man meist mit a^{-1} .

Bei kommutativen Gruppen wird die Verknüpfung oft mit $+$ bezeichnet. In diesem Fall bezeichnet man das neutrale Element meist mit 0 oder so ähnlich, und das zu a inverse Element mit $-a$.

Typische Beispiele für kommutative Gruppen sind \mathbb{R} und \mathbb{C} mit der üblichen Addition, aber auch z.B. $\mathbb{R} \setminus \{0\}$ und $\mathbb{C} \setminus \{0\}$ mit der üblichen Multiplikation.

Unter einem *Körper* versteht man eine Menge K zusammen mit zwei Verknüpfungen, die meist mit $+$ und \cdot bezeichnet werden, sodass sowohl $(K, +)$ als auch $(K \setminus \{0\}, \cdot)$ eine abelsche Gruppe bildet, und außerdem das *Distributivgesetz* $(a + b) \cdot c = a \cdot c + b \cdot c$ für alle $a, b, c \in K$ gilt. Das ist also eine Verallgemeinerung von \mathbb{R} bzw. \mathbb{C} mit der üblichen Addition und Multiplikation.

Bemerkungen zur Behandlung des Stoffes in der Schule:

1. In der Schule (und auch sonst) werden oft anschaulich-geometrische Begriffe mit exakten mathematischen Begriffen vermischt, z.B. bei Punkten und Vektoren, aber auch bei Geraden, Ebenen, Kreisen usw.. Wenn man Mathematik wirklich verstehen will, muss einem aber der Unterschied klar sein. Die mathematischen Objekte (wie z.B. der \mathbb{R}^3) sind nur *Modelle* für Objekte der (anschaulichen, physikalischen oder sonstwie verstandenen) Wirklichkeit und dürfen nicht mit diesen verwechselt werden. Es ist auch durchaus möglich und sinnvoll, für ein und dasselbe anschauliche Objekt verschiedene mathematische Modelle zu verwenden. Z.B. gibt es für die ebene Geometrie ein Axiomensystem, in dem die Begriffe "Punkt", "Gerade" und "liegt auf" zunächst undefiniert sind und nur durch die Gültigkeit gewisser Axiome eingeschränkt werden, z.B. "Zu je zwei verschiedenen Punkten gibt es genau eine Gerade, auf der diese Punkte liegen.". Dadurch entsteht ein mathematisches Modell für die Ebene, das zunächst ganz verschieden von unserem ist (d.h. vom \mathbb{R}^2). Bei geeigneter Wahl der Axiome kann man aber zeigen, dass die beiden Modelle doch im Wesentlichen gleich bzw. "isomorph" sind (siehe z.B. [1]).
2. Der an sich ganz grundlegende Begriff "Vektorraum" wird in der Schule normalerweise gar nicht oder nur am Rande behandelt. Das hängt einerseits damit zusammen, dass die Begriffe "Gruppe" und "Körper" in der Schule meist nicht vorkommen, beruht aber andererseits wohl auch auf der Ansicht, dass die einzigen Vektorräume, die in der Schule eine Rolle spielen, die Räume \mathbb{R}^n sind. Dabei übersieht man vielleicht, dass auch die linearen Teilräume des \mathbb{R}^n (das sind z.B. Geraden und Ebenen, die durch den Ursprung gehen) Vektorräume sind. Diese sind zwar isomorph zu \mathbb{R}^k mit $k < n$, unterscheiden sich aber doch in mancher Hinsicht davon, z.B. gibt es in diesen Teilräumen im Allgemeinen keine kanonische Basis (siehe Kapitel 5).

2 Matrizen und lineare Abbildungen

Beginnen wir mit der Betrachtung eines *linearen Gleichungssystems*, z.B.:

$$\begin{aligned} 3x_1 + x_2 - 2x_3 &= 0.5 \\ 2x_1 - x_2 + 4x_3 &= -6 \end{aligned}$$

Normalerweise geht es bei einem linearen Gleichungssystem darum, die *Unbekannten* x_1, x_2, \dots zu bestimmen. Das werden wir später noch ausführlich besprechen. Hier stellen wir nur einmal fest, dass das Gleichungssystem durch die Tabelle der Koeffizienten auf der linken Seite und die Liste der Zahlen auf der rechten Seite bestimmt ist:

$$A := \begin{pmatrix} 3 & 1 & -2 \\ 2 & -1 & 4 \end{pmatrix} \quad \text{und} \quad b := \begin{pmatrix} 0.5 \\ -6 \end{pmatrix}.$$

So eine Tabelle wie A nennt man eine *Matrix*. Allgemein versteht man unter einer (reellen) $m \times n$ -Matrix eine Tabelle von $m \cdot n$ (reellen) Zahlen, die in einem rechteckigen Schema mit m *Zeilen* und n *Spalten* angeordnet sind (oder zumindest so gedacht werden). Das Element in der i -ten Zeile und k -ten Spalte von A bezeichnen wir mit a_{ik} . In unserem Beispiel handelt es sich also um eine 2×3 -Matrix, und es ist z.B. $a_{21} = 2$.

Formal kann man $m \times n$ -Matrizen als Abbildungen

$$A : \{1, \dots, m\} \times \{1, \dots, n\} \rightarrow \mathbb{R} : (i, k) \mapsto a_{ik}$$

definieren. (Hier bedeutet \times natürlich das cartesische Produkt.) Die Menge aller reellen $m \times n$ -Matrizen bezeichnen wir mit $\mathbb{R}^{(m,n)}$.

Abkürzend schreiben wir manchmal auch $A = (a_{ik})$.

Wenn $m = n$ ist, spricht man von einer *quadratischen Matrix*.

In einem Computer realisiert man Matrizen normalerweise als *zweidimensionale Arrays*.

Sei A eine $m \times n$ -Matrix. Die Zeilen von A sind Elemente des \mathbb{R}^n , sie heißen deshalb auch *Zeilenvektoren* (von A). Die Spalten von A sind Elemente von \mathbb{R}^m , sie heißen auch *Spaltenvektoren* (von A).

Die rechte Seite eines linearen Gleichungssystems können wir als einen Vektor in Spaltenschreibweise auffassen. Es liegt nun nahe, die linke Seite als Pro-

dukt der Matrix A mit dem Vektor $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ anzusehen. Dann kann man

ein lineares Gleichungssystem einfach in der Form

$$Ax = b$$

schreiben. Die *Lösungsmenge* dieses Gleichungssystems ist dann die Menge

$$L := \{x \in \mathbb{R}^n : Ax = b\}.$$

Das *Produkt einer $m \times n$ -Matrix A mit einem Vektor $x \in \mathbb{R}^n$* wird also folgenderweise definiert:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{pmatrix}.$$

Wir können das auch so schreiben:

$$(Ax)_i = \sum_{k=1}^n a_{ik}x_k \quad \text{für } i \in \{1, \dots, m\}.$$

Beachten Sie, dass diese Multiplikation nur definiert ist, wenn die Anzahl der Spalten von A mit der Anzahl der Koordinaten von x übereinstimmt. Das Ergebnis der Multiplikation hat so viele Koordinaten, wie die Matrix A Zeilen hat.

Für die Multiplikation einer $m \times n$ -Matrix A mit Vektoren aus \mathbb{R}^n gelten folgende leicht zu überprüfenden Rechenregeln (für $x, y \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$):

$$\begin{aligned} A(x + y) &= Ax + Ay \\ A(\lambda x) &= \lambda (Ax) \end{aligned}$$

Jeder $m \times n$ -Matrix A entspricht nun in natürlicher Weise folgende Abbildung:

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m : x \mapsto Ax.$$

Auf Grund obiger Rechenregeln hat diese Abbildung folgende Eigenschaften:

$$\begin{aligned} f_A(x + y) &= f_A(x) + f_A(y) \quad (\text{"Additivität"}), \\ f_A(\lambda x) &= \lambda f_A(x) \quad (\text{"Homogenität"}). \end{aligned}$$

Wir könnten das auch so ausdrücken: " f_A ist mit der Summe und der Multiplikation mit einem Skalar vertauschbar", denn das f_A -Bild der Summe

zweier Vektoren ist gleich der Summe ihrer f_A -Bilder, und das f_A -Bild des λ -fachen von x ist gleich dem λ -fachen des f_A -Bilds von x .

Allgemein nennt man eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ *linear*, wenn sie diese beiden Eigenschaften hat, d.h. wenn $f(x + y) = f(x) + f(y)$ und $f(\lambda x) = \lambda f(x)$ für alle $x, y \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$. Jeder $m \times n$ -Matrix entspricht also eine *lineare Abbildung* vom \mathbb{R}^n in den \mathbb{R}^m .

Bemerkung 1 Eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist genau dann linear, wenn für je endlich viele Vektoren $v_1, \dots, v_s \in \mathbb{R}^n$ und ebensoviele Skalare $\lambda_1, \dots, \lambda_s \in \mathbb{R}$ gilt:

$$f(\lambda_1 v_1 + \dots + \lambda_s v_s) = \lambda_1 f(v_1) + \dots + \lambda_s f(v_s). \quad (*)$$

Beweis:

Da es sich um eine "genau dann"-Aussage handelt, sind zwei Richtungen zu beweisen.

1. Sei f eine Abbildung mit der angegebenen Eigenschaft (*). Dann gilt diese insbesondere für $s = 1$ und $s = 2$.

$s = 1$ bedeutet $f(\lambda_1 v_1) = \lambda_1 f(v_1)$, das ist die Homogenität.

$s = 2$ ergibt mit $\lambda_1 = \lambda_2 = 1$ die Bedingung $f(v_1 + v_2) = f(v_1) + f(v_2)$, das ist die Additivität.

Es folgt: f ist linear.

2. Sei jetzt f eine lineare Abbildung. Wir beweisen die angegebene Eigenschaft mit *vollständiger Induktion* nach s .

(Näheres zu diesem Beweisprinzip erfahren Sie in der Vorlesung "Diskrete Mathematik".)

$s = 1$: $f(\lambda_1 v_1) = \lambda_1 f(v_1)$ auf Grund der Homogenität von f .

Schluss von s auf $s + 1$:

Sei s eine beliebige natürliche Zahl. Wir nehmen an, dass

$$f(\lambda_1 v_1 + \dots + \lambda_s v_s) = \lambda_1 f(v_1) + \dots + \lambda_s f(v_s)$$

gilt (das ist die "Induktionsvoraussetzung"), und überlegen uns, dass daraus $f(\lambda_1 v_1 + \dots + \lambda_s v_s + \lambda_{s+1} v_{s+1}) = \lambda_1 f(v_1) + \dots + \lambda_s f(v_s) + \lambda_{s+1} f(v_{s+1})$

(die "Induktionsbehauptung") folgt.

Auf Grund der Additivität von f gilt:

$$f((\lambda_1 v_1 + \dots + \lambda_s v_s) + \lambda_{s+1} v_{s+1}) = f(\lambda_1 v_1 + \dots + \lambda_s v_s) + f(\lambda_{s+1} v_{s+1}).$$

Auf den ersten dieser beiden Summanden wenden wir die Induktionsvoraussetzung an und auf den zweiten die Homogenität von f . Das ergibt:

$$(\lambda_1 f(v_1) + \dots + \lambda_s f(v_s)) + \lambda_{s+1} f(v_{s+1}),$$

und damit folgt die Induktionsbehauptung. \square

(Das Zeichen \square zeigt das Ende eines Beweises an.)

Bemerkung 2 *Jede lineare Abbildung f bildet den Nullvektor auf den Nullvektor ab.*

Beweis: $f(o) = f(0x) = 0 f(x) = o$ (wobei x beliebig $\in \mathbb{R}^n$). \square

Überlegen wir uns, was die Abbildung f_A mit den kanonischen Basisvektoren macht:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix},$$

d.h. Ae_1 ist gleich der ersten Spalte von A . Analog sieht man, dass gilt:

$$f_A(e_k) = Ae_k = k\text{-te Spalte von } A.$$

Das ergibt einen sehr einprägsamen (und wichtigen) Zusammenhang zwischen einer Matrix und der zugehörigen linearen Abbildung: die Bilder der kanonischen Basisvektoren sind genau die Spaltenvektoren von A .

Auf Grund der Linearität von f_A gilt für $x = (x_1, \dots, x_n)$:

$$f_A(x) = f_A(x_1 e_1 + \dots + x_n e_n) = x_1 f_A(e_1) + \dots + x_n f_A(e_n).$$

Bezeichnen wir die k -te Spalte von A mit a_k , so können wir das auch so schreiben:

$$Ax = x_1 a_1 + \dots + x_n a_n.$$

(Das folgt übrigens auch direkt aus der Definition von Ax .)

Ax ist also eine Linearkombination der Spaltenvektoren von A , und zwar mit den Koordinaten von x als Koeffizienten.

Wir können uns jetzt leicht überlegen, dass der folgende grundlegende Satz gilt:

Satz 3 Zu jeder linearen Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ gibt es eine eindeutig bestimmte $m \times n$ -Matrix A , sodass $f = f_A$ ist.

Beweis: Sei $a_i = f(e_i)$ für alle $i \in \{1, \dots, n\}$, und A die Matrix mit den Spaltenvektoren a_1, \dots, a_n .

Dann gilt für jedes $x = (x_1, \dots, x_n) \in \mathbb{R}^n$:

$$f(x) = f(x_1 e_1 + \dots + x_n e_n) = x_1 f(e_1) + \dots + x_n f(e_n) = x_1 a_1 + \dots + x_n a_n = Ax = f_A(x),$$

also $f = f_A$.

Die Eindeutigkeit der Matrix A folgt daraus, dass $f(e_k)$ die k -te Spalte von A sein muss. \square

Sehen wir uns nun einige Spezialfälle etwas näher an.

a) $m = n = 1$:

In diesem sehr einfachen Fall ist $A = (a_{11})$, daher können wir A mit der reellen Zahl a_{11} identifizieren.

Ax ist dann nichts anderes als das Produkt zweier reeller Zahlen, und die Abbildung f_A ist eine Homothetie von $\mathbb{R}^1 = \mathbb{R}$ mit Streckungsfaktor A und Zentrum 0.

b) $m = 1, n > 1$:

Hier hat A nur eine Zeile. Daher können wir den Zeilenindex weglassen und $A = (a_1, \dots, a_n)$ schreiben, mit $a_i \in \mathbb{R}$.

$$f_A(x) = Ax = a_1 x_1 + \dots + a_n x_n.$$

f_A heißt in diesem Fall auch *Linearform* (auf dem \mathbb{R}^n).

Insbesondere für $n = 2$ kann man zur Veranschaulichung einer Linearform die *Urbilder* $f_A^{-1}(\{b\})$ für $b \in \mathbb{R}$ betrachten:

$$f_A^{-1}(\{b\}) = \{x \in \mathbb{R}^2 : f_A(x) = b\} = \{(x_1, x_2) \in \mathbb{R}^2 : a_1 x_1 + a_2 x_2 = b\},$$

das sind zueinander parallele Geraden, die man auch als "Niveaulinien" der Funktion f_A ansehen kann. (Vgl. Beispiel 3 im Kapitel 4.3.)

c) $m > 1, n = 1$:

Hier hat A nur eine Spalte, und wir können $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$ schreiben, mit $a_i \in \mathbb{R}$. Für $x \in \mathbb{R}$ ist dann $Ax = \begin{pmatrix} a_1 x \\ \vdots \\ a_m x \end{pmatrix}$. Fassen wir A als einen Vektor $\in \mathbb{R}^m$ auf, so ist also Ax ein skalares Vielfaches dieses Vektors. Der *Bildraum* $f_A(\mathbb{R}) = \{f_A(x) : x \in \mathbb{R}\} = \{Ax : x \in \mathbb{R}\}$ ist also eine Gerade mit Richtungsvektor A . (Vgl. Kapitel 5.3.)

d) $m = n = 2$:

Seien a_1, a_2 die beiden Spaltenvektoren von A , also $a_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$, $a_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$. Dann ist $Ax = x_1 a_1 + x_2 a_2$.

Sehen wir uns hier das Bild des Einheitsquadrats $Q := [0, 1]^2$ unter der Abbildung f_A an.

$Q = \{(x_1, x_2) : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$, also

$f_A(Q) = \{f_A(x) : x \in Q\} = \{x_1 a_1 + x_2 a_2 : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$.

$f_A(Q)$ besteht also aus allen Linearkombinationen der beiden Vektoren a_1, a_2 mit Koeffizienten aus dem Einheitsintervall. Das nennt man "das von a_1 und a_2 aufgespannte *Parallelogramm*" (mit den Ecken $o, a_1, a_2, a_1 + a_2$).

e) $m = n$ beliebig:

Hier sehen wir uns insbesondere den Spezialfall an, wo $a_{ik} = 0$ für $i \neq k$. Mit $\lambda_k := a_{kk}$ ist dann

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Eine solche Matrix nennt man *Diagonalmatrix*, und schreibt sie auch so:

$$A = \begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{pmatrix}.$$

Wir sehen:

$$Ax = \begin{pmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \vdots \\ \lambda_n x_n \end{pmatrix}.$$

Das kann man so interpretieren, dass durch f_A die k -te Koordinate um den Faktor λ_k gestreckt bzw. gestaucht wird (für $\lambda_k > 0$). Falls alle $\lambda_k = 1$ sind, ist f_A die *identische Abbildung*, d.h. $f_A(x) = x$ für alle $x \in \mathbb{R}^n$. In diesem Fall nennt man A die $n \times n$ -Einheitsmatrix und bezeichnet sie meist mit E oder I (oder genauer E_n bzw. I_n):

$$E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

Es gilt also $Ex = x$ für alle $x \in \mathbb{R}^n$.

Die Elemente a_{ik} mit $i = k$ nennt man auch die *Hauptdiagonale* der Matrix. Bei einer Diagonalmatrix sind also alle Elemente außerhalb der Hauptdiagonalen gleich Null. Eine Diagonalmatrix, bei der alle Diagonalelemente gleich einer Zahl λ sind, entspricht der Homothetie s_λ .

f) $m = n - 1 \geq 1$:

Sei P eine $m \times n$ -Matrix, die aus der $n \times n$ -Einheitsmatrix durch Weglassen der letzten Zeile entsteht (oder aus der $m \times m$ -Einheitsmatrix durch Hinzufügung einer Nullspalte), d.h.

$$P = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

Dann ist $Px = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \end{pmatrix}$, und die Abbildung f_P heißt *Normalprojektion* (vom \mathbb{R}^n auf den \mathbb{R}^{n-1}). Ein besonders wichtiger Fall ist natürlich $m = 2$, $n = 3$ zur Erzeugung 2-dimensionaler Bilder von 3-dimensionalen Objekten (Stichwort: *Computergeometrie*).

Zum Schluss sehen wir uns wieder die Zusammensetzungen mit Translationen an. Sei A eine $m \times n$ -Matrix, $b \in \mathbb{R}^m$ und $f = \tau_b \circ f_A$, d.h. $f(x) = Ax + b$. So eine Abbildung heißt *affine Abbildung*. Wenn $A = \lambda E$, dann ist $f(x) = \lambda x + b$, d.h. f ist eine Homothetie. Wenn $b = o$, dann ist f eine lineare Abbildung. Allgemein gilt: $f(o) = b$.

Für $m = n = 2$ etwa bildet f das Einheitsquadrat auf ein Parallelogramm mit den Ecken b , $b + a_1$, $b + a_2$, $b + a_1 + a_2$ ab.

(In manchen Anwendungsgebieten werden affine Abbildungen auch linear genannt.)

Bemerkung zur Behandlung des Stoffes in der Schule:

Matrizen gehören nur in gewissen Schultypen (Realgymnasien) zum Standardstoff. Lineare Abbildungen werden daher meist nur im \mathbb{R}^1 behandelt, und die Schüler bekommen im Allgemeinen keine Vorstellung davon, wie man z.B. Drehungen, Spiegelungen und Projektionen rechnerisch bzw. mit einem Computer durchführen kann (siehe Kapitel 7.2.5 und 8.2.2). Das scheint auch im Hinblick auf allgemeinbildende Aspekte des Unterrichts ein Mangel zu sein.

3 Addition und Multiplikation von Matrizen

Die *Addition* von zwei Matrizen ist sehr einfach: Seien A und B zwei $m \times n$ -Matrizen mit Elementen a_{ik} , b_{ik} . Dann versteht man unter $A + B$ die Matrix C mit den Elementen $c_{ik} := a_{ik} + b_{ik}$. Ebenso einfach ist die *Multiplikation einer Matrix mit einem Skalar* λ definiert: λA ist die Matrix mit Elementen λa_{ik} .

Beispiel: Sei $A = \begin{pmatrix} 2 & 0 & -1 \\ 1 & 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -2 & 0 \\ 1 & 2 & -3 \end{pmatrix}$.

Dann ist $\frac{1}{2}A + \frac{1}{2}B = \begin{pmatrix} \frac{3}{2} & -1 & -\frac{1}{2} \\ 1 & \frac{5}{2} & \frac{1}{2} \end{pmatrix}$.

Eine $m \times n$ -Matrix, die aus lauter Nullen besteht, heißt *Nullmatrix* und wird mit O bezeichnet. Die Matrix $(-1)A$ wird auch mit $-A$ bezeichnet.

Die folgenden Rechenregeln sind praktisch unmittelbar klar:

$$\begin{aligned}
 (A + B) + C &= A + (B + C) \\
 A + O &= A \\
 A + (-A) &= O \\
 A + B &= B + A \\
 \lambda(A + B) &= \lambda A + \lambda B \\
 (\lambda + \mu)A &= \lambda A + \mu A \\
 \lambda(\mu A) &= (\lambda\mu)A \\
 1A &= A.
 \end{aligned}$$

Für feste Zahlen m und n bilden die (reellen) $m \times n$ -Matrizen also einen Vektorraum.

Außerdem gilt für beliebige $x \in \mathbb{R}^n$

$$\begin{aligned}
 (A + B)x &= Ax + Bx, \\
 (\lambda A)x &= \lambda(Ax).
 \end{aligned}$$

Die Matrix $A + B$ entspricht also der Abbildung $x \mapsto Ax + Bx$, das heißt $f_{A+B} = f_A + f_B$. In ähnlicher Weise entspricht die Matrix λA der Abbildung $x \mapsto \lambda Ax$, also $f_{\lambda A} = \lambda(f_A)$.

Für Matrizen gibt es auch eine natürliche *Multiplikation*. Diese wird aber nicht analog zur Addition definiert, sondern so, dass AB der Zusammensetzung der zu A und B gehörigen linearen Abbildungen entspricht, d.h.

$$f_{AB} = f_A \circ f_B,$$

mit $f_B : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $f_A : \mathbb{R}^m \rightarrow \mathbb{R}^s$, d.h. A ist eine $s \times m$ -Matrix und B eine $m \times n$ -Matrix. Es ist sehr leicht einzusehen, dass die Zusammensetzung zweier linearer Abbildungen wieder linear ist und daher tatsächlich einer Matrix entspricht:

$$f_A(f_B(x + y)) = f_A(f_B(x) + f_B(y)) = f_A(f_B(x)) + f_A(f_B(y)),$$

$$f_A(f_B(\lambda x)) = f_A(\lambda f_B(x)) = \lambda f_A(f_B(x)).$$

Das Produkt AB ist also nur definiert, wenn gilt:

$$\text{Spaltenanzahl von } A = \text{Zeilenanzahl von } B.$$

Wir sehen außerdem, dass $f_A \circ f_B$ eine Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^s$ ist, d.h. die zugehörige Matrix muss s Zeilen und n Spalten haben, also

$$\begin{aligned} \text{Zeilenanzahl von } AB &= \text{Zeilenanzahl von } A, \\ \text{Spaltenanzahl von } AB &= \text{Spaltenanzahl von } B. \end{aligned}$$

Wie kann man nun das Produkt AB berechnen?

Sei $x \in \mathbb{R}^n$, $y = f_B(x) = Bx \in \mathbb{R}^m$ und $z = f_A(f_B(x)) = f_A(y) = Ay \in \mathbb{R}^s$. Das heißt:

$$\begin{aligned} y_j &= \sum_{k=1}^n b_{jk} x_k \quad \text{für } j \in \{1, \dots, m\}, \\ z_i &= \sum_{j=1}^m a_{ij} y_j \quad \text{für } i \in \{1, \dots, s\}. \end{aligned}$$

Setzen wir y_j aus der ersten Gleichung in die zweite ein, so erhalten wir

$$z_i = \sum_{j=1}^m a_{ij} \left(\sum_{k=1}^n b_{jk} x_k \right).$$

Durch Vertauschung der Summationsreihenfolge ergibt sich

$$z_i = \sum_{k=1}^n \left(\sum_{j=1}^m a_{ij} b_{jk} \right) x_k.$$

Für die Elemente c_{ik} der Produktmatrix $C = AB$ erhalten wir somit folgende Formel:

$$c_{ik} = \sum_{j=1}^m a_{ij} b_{jk}.$$

c_{ik} entsteht also durch Multiplikation der Elemente der i -ten Zeile von A mit den Elementen der k -ten Spalte von B und anschließende Summation.

Bemerkungen:

1. Sei A eine $m \times n$ -Matrix und $x \in \mathbb{R}^n$. Fassen wir x als $n \times 1$ -Matrix auf, so stimmt das Matrizenprodukt von A und x mit dem im vorigen Kapitel definierten Produkt Ax überein.
2. Wenn b_1, \dots, b_n die Spaltenvektoren von B sind, so sind Ab_1, \dots, Ab_n die Spaltenvektoren von AB . Wir können das auch so schreiben:

$$A(b_1, \dots, b_n) = (Ab_1, \dots, Ab_n).$$

Insbesondere können wir auf diese Weise sehen, dass $EB = B$ gilt (wobei E die $m \times m$ -Einheitsmatrix ist).

Rechenregeln:

Für die Matrizenmultiplikation gelten die folgenden beiden *Assoziativgesetze*:

$$\begin{aligned} A(BC) &= (AB)C \\ \lambda(AB) &= (\lambda A)B = A(\lambda B) \end{aligned}$$

Das erste folgt unmittelbar aus der Assoziativität der Zusammensetzung von Abbildungen, für das zweite benötigt man die Homogenität von linearen Abbildungen.

Auch die folgenden beiden *Distributivgesetze* sind leicht zu beweisen:

$$\begin{aligned} A(B + C) &= AB + AC \\ (A + B)C &= AC + BC \end{aligned}$$

Das erste folgt aus der Additivität der Abbildung f_A , das zweite aus der Regel $(A + B)x = Ax + Bx$.

Beispiele:

- Seien A und B wie in obigem Beispiel zur Addition von Matrizen. Dann ist AB nicht definiert, da die Spaltenanzahl von A nicht mit der Zeilenanzahl von B übereinstimmt. Aus demselben Grund ist auch BA nicht definiert.

- Sei $A = \begin{pmatrix} 2 & 0 & -1 \\ 1 & 3 & 4 \end{pmatrix}$ und $B = \begin{pmatrix} 3 & 0 \\ 1 & -1 \\ 2 & 1 \end{pmatrix}$.

Dann ist $AB = \begin{pmatrix} 4 & -1 \\ 14 & 1 \end{pmatrix}$ und $BA = \begin{pmatrix} 6 & 0 & -3 \\ 1 & -3 & -5 \\ 5 & 3 & 2 \end{pmatrix}$.

- Sei $Q = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Dann ist AQ nicht definiert, aber $QA = \begin{pmatrix} 4 & 6 & 7 \\ 10 & 12 & 13 \end{pmatrix}$.

- Auch bei zwei gleich großen quadratischen Matrizen kann das Ergebnis von der Reihenfolge abhängen, muss es aber nicht:

Sei $R = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ und Q wie oben. $QR = \begin{pmatrix} 1 & 5 \\ 3 & 11 \end{pmatrix}$, $RQ = \begin{pmatrix} 4 & 6 \\ 6 & 8 \end{pmatrix}$.

Mit $S = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$ gilt aber $RS = \begin{pmatrix} 2 & 4 \\ 0 & 6 \end{pmatrix} = SR$.

Da man leicht ähnliche Beispiele für $n \times n$ -Matrizen mit beliebigem $n > 1$ finden kann (z.B. durch Ergänzung der obigen Matrizen R und Q mit Nullen), folgt:

Bemerkung 4 Für festes $n > 1$ ist die Matrizenmultiplikation eine nichtkommutative Verknüpfung auf der Menge $\mathbb{R}^{(n,n)}$ aller (reellen) $n \times n$ -Matrizen (mit neutralem Element E).

5. Sei D eine $n \times n$ -Diagonalmatrix, $D = \begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{pmatrix}$, und A eine beliebige $m \times n$ -Matrix mit Elementen a_{ik} . Dann ist

$$AD = \begin{pmatrix} \lambda_1 a_{11} & \dots & \lambda_n a_{1n} \\ \vdots & \vdots & \vdots \\ \lambda_1 a_{m1} & \dots & \lambda_n a_{mn} \end{pmatrix},$$

d.h. durch diese Multiplikation wird die k -te Spalte von A mit λ_k multipliziert, für alle $k \in \{1, \dots, n\}$.

Mit der $n \times n$ -Einheitsmatrix an Stelle von D erhalten wir:

$$AE = A.$$

Bemerkung zur Behandlung des Stoffes in der Schule:

Im Realgymnasium kommen zwar Matrizen meistens vor, aber lineare Abbildungen nicht. Daher muss dann die Matrizenmultiplikation anders motiviert werden. Das ist zwar möglich (z.B. mit Graphen, siehe etwa [6]), erscheint aber doch irgendwie unnatürlich. Das wäre also ein weiteres Argument für die Aufnahme von linearen Abbildungen (vom \mathbb{R}^n in den \mathbb{R}^m) in den Lehrplan, zumindest für kleine Werte von n und m .

4 Lineare Gleichungssysteme

Sei A eine $m \times n$ -Matrix und $b \in \mathbb{R}^m$. Dann nennt man

$$Ax = b$$

ein *lineares Gleichungssystem*. Es geht dabei darum, Vektoren $x \in \mathbb{R}^n$ zu finden, sodass $Ax = b$ gilt. Mit anderen Worten, es geht um die nähere Bestimmung der Menge

$$L = \{x \in \mathbb{R}^n : Ax = b\},$$

das ist die *Lösungsmenge* des betrachteten Gleichungssystems. Jedes Element der Lösungsmenge heißt *Lösung* des Gleichungssystems. Wir werden sehen, dass es grundsätzlich drei Möglichkeiten gibt: Die Lösungsmenge kann leer sein, sie kann aus genau einem Punkt bestehen, oder sie kann eine Menge mit unendlich vielen Punkten sein.

Ein lineares Gleichungssystem $Ax = b$ mit A und b wie oben besteht also aus m Gleichungen. Die i -te Zeile des Gleichungssystems (d.h. die i -te Gleichung) sieht so aus:

$$a_{i1}x_1 + \dots + a_{in}x_n = b_i.$$

Die x_i heißen die *Unbekannten*.

4.1 Elementare Umformungen

Um zu einem systematischen Verfahren zur Bestimmung der Lösungsmenge L zu kommen, überlegen wir uns zunächst, dass L bei den folgenden *elementaren Umformungen* unverändert bleibt:

1. *Addition eines skalaren Vielfachen einer Zeile zu einer anderen Zeile*:

Addieren wir das λ -fache der r -te Zeile des Gleichungssystems zur s -ten Zeile (mit $r \neq s$), so erhalten wir als neue s -te Zeile:

$$(a_{s1} + \lambda a_{r1})x_1 + \dots + (a_{sn} + \lambda a_{rn})x_n = b_s + \lambda b_r. \quad (*)$$

Die anderen Zeilen bleiben unverändert.

Wenn nun $x^\circ = \begin{pmatrix} x_1^\circ \\ \vdots \\ x_n^\circ \end{pmatrix}$ eine Lösung des gegebenen Gleichungssystems ist, dann ist x° offensichtlich auch eine Lösung der neuen Gleichung (*). Wenn umgekehrt $x' = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$ eine Lösung des neuen Gleichungssystems ist, dann erfüllt x' insbesondere das λ -fache der (unveränderten) r -ten Gleichung und die neue s -te Gleichung (*), d.h. es gilt

$$\lambda a_{r1}x_1' + \dots + \lambda a_{rn}x_n' = \lambda b_r$$

und

$$(a_{s1} + \lambda a_{r1})x_1' + \dots + (a_{sn} + \lambda a_{rn})x_n' = b_s + \lambda b_r.$$

Durch Subtraktion folgt daraus

$$a_{s1}x_1' + \dots + a_{sn}x_n' = b_s,$$

und daher ist x' auch eine Lösung des gegebenen Gleichungssystems.

2. *Multiplikation einer Zeile mit einem Skalarfaktor $\lambda \neq 0$:*

Hier ist die Invarianz der Lösungsmenge praktisch unmittelbar klar.

3. *Vertauschung zweier Zeilen:*

Auch das erfordert keine Begründung.

In gewissen Fällen erweist es sich als zweckmäßig, auch noch eine weitere elementare Umformung zu verwenden, bei der die Lösungsmenge zwar nicht gleich bleibt, sich aber nur die Nummerierung der Unbekannten x_i ändert:

4. *Vertauschung zweier Spalten:*

Vertauschen wir die r -te Spalte der Matrix A mit der s -ten Spalte, so unterscheidet sich die neue Lösungsmenge von der alten nur durch eine Vertauschung der r -ten mit der s -ten Koordinate.

4.2 Halbdiagonalform

Wir überlegen uns nun zunächst, dass wir mit Hilfe dieser elementaren Umformungen erreichen können, dass die Matrix A eines beliebigen linearen

Gleichungssystems folgende *Halbdiagonalform* annimmt:

$$\tilde{A} = \begin{pmatrix} 1 & \tilde{a}_{12} & \dots & \dots & \tilde{a}_{1,r} & \tilde{a}_{1,r+1} & \dots & \tilde{a}_{1n} \\ 0 & 1 & \tilde{a}_{23} & \dots & \tilde{a}_{2r} & \tilde{a}_{2,r+1} & \dots & \tilde{a}_{2n} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & 1 & \tilde{a}_{r-1,r} & \tilde{a}_{r-1,r+1} & \dots & \tilde{a}_{r-1,n} \\ 0 & \dots & \dots & 0 & 1 & \tilde{a}_{r,r+1} & \dots & \tilde{a}_{rn} \\ 0 & \dots & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

mit einer bestimmten Zahl $r \in \{0, \dots, \min(m, n)\}$. Diese Zahl r hat eine besondere Bedeutung, sie heißt der *Rang* der Matrix A (engl. *rank of A* , Bezeichnung: $\text{rank } A$). Die ersten r Zeilenvektoren der Halbdiagonalform sind $\neq 0$, die anderen sind $= 0$. Die Anzahl der Einsen in der Hauptdiagonalen ist gleich r .

Wenn A die Nullmatrix ist, haben wir nichts zu tun, denn dann haben wir eine Halbdiagonalform mit $r = 0$ vor uns. Sonst können wir annehmen, dass $a_{11} \neq 0$ ist. Denn sollte das nicht der Fall sein, gibt es sicher mindestens ein Element $a_{ik} \neq 0$, und durch Zeilen- und/oder Spaltenvertauschung können wir erreichen, dass dieses Element in die linke obere Ecke der Matrix kommt.

Wir multiplizieren nun die 1. Zeile des Gleichungssystems mit $1/a_{11}$. Die erste Zeile der Matrix A lautet dann

$$(1, a'_{12}, \dots, a'_{1n})$$

mit $a'_{1k} = \frac{a_{1k}}{a_{11}}$ für $k \in \{2, \dots, n\}$, und die rechte Seite der ersten Gleichung ist $b'_1 = \frac{b_1}{a_{11}}$.

Als Nächstes addieren wir für $i \in \{2, \dots, m\}$ das $(-a_{i1})$ -fache der (neuen) ersten Gleichung zur i -ten Gleichung. Das wird durch folgende Schreibweise gut ausgedrückt, die allerdings nicht allgemein üblich ist:

$$z_i \leftarrow z_i - a_{i1}z_1 \quad \text{für } i = 2, \dots, m.$$

Der Linkspfeil " \leftarrow " bedeutet dabei so etwas wie eine Zuweisung beim Programmieren. z_i bezeichnet natürlich die i -te Zeile des Gleichungssystems.

Die Matrix und die rechte Seite des Gleichungssystems haben nach diesen

Umformungen folgende Gestalt:

$$A' = \begin{pmatrix} 1 & a'_{12} & \dots & a'_{1n} \\ 0 & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a'_{m2} & \dots & a'_{mn} \end{pmatrix}, \quad b' = \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_m \end{pmatrix}$$

mit

$$a'_{ik} = a_{ik} - a_{i1}a'_{1k} \quad \text{für } 2 \leq i \leq m, 2 \leq k \leq n$$

und

$$b'_i = b_i - a_{i1}b'_1 \quad \text{für } 2 \leq i \leq m.$$

Wenn jetzt $a'_{ik} = 0$ ist für alle $i \geq 2$ und $k \geq 2$, dann haben wir eine Halbdiagonalform mit $r = 1$ erreicht. Dasselbe gilt, wenn es gar keine a'_{ik} mit $i \geq 2$ gibt, d.h. wenn $m = 1$ ist. Andernfalls setzen wir das Verfahren fort: Wegen der Möglichkeit von Zeilen- und Spaltenvertauschungen können wir $a'_{22} \neq 0$ annehmen. Wir multiplizieren die 2. Zeile mit $1/a'_{22}$ und addieren anschließend für $i \in \{3, \dots, m\}$ das $(-a'_{i2})$ -fache der 2. Zeile zur i -ten Zeile. Dadurch erhalten wir eine Matrix der folgenden Gestalt:

$$A'' = \begin{pmatrix} 1 & a'_{12} & a'_{13} & \dots & a'_{1n} \\ 0 & 1 & a''_{23} & \dots & a''_{2n} \\ 0 & 0 & a''_{33} & \dots & a''_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a''_{m3} & \dots & a''_{mn} \end{pmatrix},$$

wo also in den ersten beiden Spalten unterhalb der Hauptdiagonalen nur mehr Nullen stehen. Wenn $m = 2$ oder $a''_{ik} = 0$ für alle $i \geq 3$ und $k \geq 3$, haben wir eine Halbdiagonalform mit $r = 2$. Sonst setzen wir das Verfahren in analoger Weise fort, und erhalten spätestens nach $\min(m, n)$ Schritten eine Halbdiagonalform.

Die rechte Seite des auf Halbdiagonalform gebrachten Gleichungssystems bezeichnen wir mit

$$\tilde{b} = \begin{pmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_r \\ \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{pmatrix}.$$

Beispiel: Sei

$$A = \begin{pmatrix} 2 & 4 & 3 & 2 \\ 3 & 2 & 1 & 3 \\ 1 & 2 & 3 & 0 \\ 6 & 8 & 7 & 5 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 6 \end{pmatrix}.$$

Wir wollen das Gleichungssystem $Ax = b$ auf Halbdiagonalf orm bringen und führen dazu folgende Umformungen durch:

$$z_1 \leftarrow \frac{1}{2}z_1:$$

$$\begin{array}{rrrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 3 & 2 & 1 & 3 & 3 \\ 1 & 2 & 3 & 0 & 2 \\ 6 & 8 & 7 & 5 & 6 \end{array}$$

$$z_2 \leftarrow z_2 - 3z_1,$$

$$z_3 \leftarrow z_3 - z_1,$$

$$z_4 \leftarrow z_4 - 6z_1:$$

$$\begin{array}{rrrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & -4 & -\frac{7}{2} & 0 & \frac{3}{2} \\ 0 & 0 & \frac{3}{2} & -1 & \frac{3}{2} \\ 0 & -4 & -2 & -1 & 3 \end{array}$$

$$z_2 \leftarrow -\frac{1}{4}z_2:$$

$$\begin{array}{rrrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & \frac{7}{8} & 0 & -\frac{3}{8} \\ 0 & 0 & \frac{3}{2} & -1 & \frac{3}{2} \\ 0 & -4 & -2 & -1 & 3 \end{array}$$

$$z_4 \leftarrow z_4 + 4z_2:$$

$$\begin{array}{rrrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & \frac{7}{8} & 0 & -\frac{3}{8} \\ 0 & 0 & \frac{3}{2} & -1 & \frac{3}{2} \\ 0 & 0 & \frac{3}{2} & -1 & \frac{3}{2} \end{array}$$

$$z_3 \leftarrow \frac{2}{3}z_3:$$

$$\begin{array}{rrrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & \frac{7}{8} & 0 & -\frac{3}{8} \\ 0 & 0 & 1 & -\frac{2}{3} & 1 \\ 0 & 0 & \frac{3}{2} & -1 & \frac{3}{2} \end{array}$$

$z_4 \leftarrow z_4 - \frac{3}{2}z_3$:

$$\begin{array}{rrrrr} 1 & 2 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{8} & 0 & -\frac{3}{8} \\ 0 & 0 & 1 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

Wir sehen also, dass die Matrix Rang 3 hat.

Bemerkungen zum Begriff "Rang":

1. Wenn man nur den Rang einer Matrix bestimmen will, ohne ein entsprechendes Gleichungssystem zu lösen, dann führt man die angegebenen Umformungen natürlich ohne eine rechte Seite durch.
2. Angenommen, wir haben eine Matrix A durch elementare Umformungen auf Halbdiagonalform gebracht, und dabei hat sich ein bestimmter Rang r ergeben. Es wäre nun vorstellbar, dass man durch andere elementare Umformungen eine Halbdiagonalform mit einem anderen Rang erhalten kann. Dann wäre aber der Begriff "Rang einer Matrix" ziemlich unsinnig. In Wirklichkeit kann so etwas nicht passieren. Man kann nämlich zeigen, dass der (durch eine Halbdiagonalform ermittelte) Rang einer Matrix A immer gleich der Dimension des Bildraums der zugehörigen linearen Abbildung f_A ist (siehe Satz 20).

4.3 Bestimmung der Lösungsmenge

Wir unterscheiden zwei Fälle:

1. Fall: Es gibt ein $s \in \{r+1, \dots, m\}$ mit $\tilde{b}_s \neq 0$.

Die s -te Zeile des Gleichungssystems lautet dann:

$$0x_1 + \dots + 0x_n = \tilde{b}_s,$$

und diese Gleichung hat wegen $\tilde{b}_s \neq 0$ keine Lösung. Daher hat auch das ganze Gleichungssystem keine Lösung, d.h. die Lösungsmenge ist leer und wir können das Verfahren abbrechen.

2. Fall: $\tilde{b}_s = 0$ für alle $s \in \{r+1, \dots, m\}$.

In diesem Fall addieren wir für alle $i \in \{1, \dots, r-1\}$ das $(-\tilde{a}_{ir})$ -fache der r -ten Zeile zur i -ten Zeile. Dadurch erhalten wir in der r -ten Spalte auch oberhalb der Hauptdiagonalen lauter Nullen, also insgesamt den kanonischen Basisvektor e_r . Die Elemente der so veränderten Matrix nennen wir

\tilde{a}'_{ik} . Anschließend addieren wir für $i \in \{1, \dots, r-2\}$ das $(-\tilde{a}'_{i,r-1})$ -fache der $(r-1)$ -ten Zeile zur i -ten Zeile usw., bis wir schließlich ein Gleichungssystem erhalten, dessen Matrix und rechte Seite folgende Gestalt haben:

$$\hat{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & \hat{a}_{1,r+1} & \dots & \hat{a}_{1n} \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \dots & 0 & 1 & \hat{a}_{r,r+1} & \dots & \hat{a}_{rn} \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \hat{b} = \begin{pmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Das Gleichungssystem selbst sieht folgendermaßen aus, wenn wir die irrelevannten Null-Zeilen weglassen:

$$\begin{aligned} x_1 + \hat{a}_{1,r+1}x_{r+1} + \dots + \hat{a}_{1n}x_n &= \hat{b}_1 \\ &\dots \\ x_r + \hat{a}_{r,r+1}x_{r+1} + \dots + \hat{a}_{rn}x_n &= \hat{b}_r \end{aligned}$$

Wir sehen, dass man hier x_{r+1}, \dots, x_n beliebig wählen kann, sagen wir

$$\begin{aligned} x_{r+1} &= \lambda_1, \\ &\dots \\ x_n &= \lambda_{n-r}. \end{aligned}$$

x_1, \dots, x_r ergeben sich dann in eindeutiger Weise daraus:

$$\begin{aligned} x_1 &= -\hat{a}_{1,r+1}\lambda_1 - \dots - \hat{a}_{1n}\lambda_{n-r} + \hat{b}_1 \\ &\dots \\ x_r &= -\hat{a}_{r,r+1}\lambda_1 - \dots - \hat{a}_{rn}\lambda_{n-r} + \hat{b}_r \end{aligned}$$

Die "allgemeine Lösung" des Gleichungssystems können wir daher so schreiben:

$$\begin{aligned} x_1 &= \hat{b}_1 - \lambda_1 \hat{a}_{1,r+1} - \dots - \lambda_{n-r} \hat{a}_{1n} \\ &\dots \\ x_r &= \hat{b}_r - \lambda_1 \hat{a}_{r,r+1} - \dots - \lambda_{n-r} \hat{a}_{rn} \\ x_{r+1} &= \lambda_1 \\ &\dots \\ x_n &= \lambda_{n-r} \end{aligned}$$

Die Lösungsmenge hat also folgende Gestalt:

$$L = \{x \in \mathbb{R}^n : x = \hat{b}' + \lambda_1 u_1 + \dots + \lambda_{n-r} u_{n-r} \text{ mit } \lambda_1, \dots, \lambda_{n-r} \in \mathbb{R}\},$$

wobei

$$\hat{b}' = \begin{pmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad u_1 = \begin{pmatrix} -\hat{a}_{1,r+1} \\ \vdots \\ -\hat{a}_{r,r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, u_{n-r} = \begin{pmatrix} -\hat{a}_{1,n} \\ \vdots \\ -\hat{a}_{r,n} \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

(\hat{b}' unterscheidet sich für $m \neq n$ von \hat{b} durch die Anzahl der Nullen: $\hat{b}' \in \mathbb{R}^n$, $\hat{b} \in \mathbb{R}^m$.)

Das hier beschriebene Verfahren zur Lösung von linearen Gleichungssystemen heißt *Gauß'sches Eliminationsverfahren*. (Carl Friedrich Gauß (1777-1855, Braunschweig, Göttingen) ist einer der berühmtesten deutschen Mathematiker.)

Wählen wir $\lambda_1 = \dots = \lambda_{n-r} = 0$, so sehen wir, dass der Vektor \hat{b}' auch eine Lösung ist. Jede andere Lösung erhält man also, indem man zu der "partikulären Lösung" \hat{b}' eine beliebige Linearkombination der Vektoren u_1, \dots, u_{n-r} addiert.

Das Gleichungssystem $Ax = b$ heißt *homogen*, wenn $b = o$ ist. (Das ist natürlich gleichbedeutend mit $\hat{b}' = o$.) Andernfalls heißt es *inhomogen*, und $Ax = o$ heißt das *zugehörige homogene Gleichungssystem*. Wir können also sagen: Man erhält die allgemeine Lösung eines inhomogenen linearen Gleichungssystems, indem man zu der partikulären Lösung \hat{b}' dieses Systems die allgemeine Lösung des zugehörigen homogenen Gleichungssystems addiert.

Die Zahlen λ_i nennt man hier *Parameter*, und die Darstellung

$$x = \hat{b}' + \lambda_1 u_1 + \dots + \lambda_{n-r} u_{n-r}$$

heißt *Parameterdarstellung der Lösungsmenge*.

Bei einem homogenen linearen Gleichungssystem kann der Fall nicht eintreten, dass die Lösungsmenge leer ist, denn der Nullvektor ist jedenfalls

eine Lösung: $Ao = o$. In diesem Zusammenhang nennt man den Nullvektor die *triviale Lösung* des Gleichungssystems.

Die Lösungsmenge eines homogenen linearen Gleichungssystems ist also mit den obigen Bezeichnungen die Menge aller Linearkombinationen von u_1, \dots, u_{n-r} .

Beim Beispiel des vorigen Abschnitts sehen wir, dass die Lösungsmenge nicht leer ist, und wir können folgenderweise eine Parameterdarstellung berechnen:

$$z_1 \leftarrow z_1 - \frac{3}{2}z_3,$$

$$z_2 \leftarrow z_2 - \frac{7}{8}z_3:$$

$$\begin{array}{rrrrr} 1 & 2 & 0 & 2 & -1 \\ 0 & 1 & 0 & \frac{7}{12} & -\frac{10}{8} \\ 0 & 0 & 1 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

$$z_1 \leftarrow z_1 - 2z_2:$$

$$\begin{array}{rrrrr} 1 & 0 & 0 & \frac{5}{6} & \frac{3}{2} \\ 0 & 1 & 0 & \frac{7}{12} & -\frac{5}{4} \\ 0 & 0 & 1 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

Wir erhalten also folgende Parameterdarstellung der Lösungsmenge:

$$x = \begin{pmatrix} \frac{3}{2} \\ -\frac{5}{4} \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -\frac{5}{6} \\ -\frac{7}{12} \\ \frac{2}{3} \\ 1 \end{pmatrix}.$$

Probe:

$$\begin{pmatrix} 2 & 4 & 3 & 2 \\ 3 & 2 & 1 & 3 \\ 1 & 2 & 3 & 0 \\ 6 & 8 & 7 & 5 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ -\frac{5}{4} \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 6 \end{pmatrix},$$

$$\begin{pmatrix} 2 & 4 & 3 & 2 \\ 3 & 2 & 1 & 3 \\ 1 & 2 & 3 & 0 \\ 6 & 8 & 7 & 5 \end{pmatrix} \begin{pmatrix} -\frac{5}{6} \\ -\frac{7}{12} \\ \frac{2}{3} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Weitere Beispiele:

1. Wir wollen das folgende lineare Gleichungssystem lösen:

$$\begin{aligned} x_1 + x_2 + 2x_3 + 3x_4 &= 4 \\ 2x_1 + 2x_2 + 3x_3 + 4x_4 &= 5 \end{aligned}$$

$z_2 \leftarrow z_2 - 2z_1$:

$$\begin{array}{ccccc} 1 & 1 & 2 & 3 & 4 \\ 0 & 0 & -1 & -2 & -3 \end{array}$$

Jetzt ist eine Spaltenvertauschung unumgänglich.

$a_3 \longleftrightarrow a_2$:

$$\begin{array}{ccccc} 1 & 2 & 1 & 3 & 4 \\ 0 & -1 & 0 & -2 & -3 \end{array}$$

$z_2 \leftarrow (-1)z_2$:

$$\begin{array}{ccccc} 1 & 2 & 1 & 3 & 4 \\ 0 & 1 & 0 & 2 & 3 \end{array}$$

Damit haben wir die Halbdiagonalf orm mit $r = 2$ erreicht, und zur Bestimmung der Lösungsmenge fehlt nur noch ein Schritt:

$z_1 \leftarrow z_1 - 2z_2$:

$$\begin{array}{ccccc} 1 & 0 & 1 & -1 & -2 \\ 0 & 1 & 0 & 2 & 3 \end{array}$$

Wir erhalten also:

$$\hat{b}' = \begin{pmatrix} -2 \\ 3 \\ 0 \\ 0 \end{pmatrix}, \quad u_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix}.$$

Unter Berücksichtigung der durchgeführten Spaltenvertauschung können wir jetzt die Parameterdarstellung der Lösungsmenge ablesen:

$$x = \begin{pmatrix} -2 \\ 0 \\ 3 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix}.$$

2. Sei $A = \begin{pmatrix} 4 & 1 & -1 \\ -2 & 0 & 3 \\ 2 & 1 & 2 \end{pmatrix}$ und $b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, und wir wollen das Gleichungssystem $Ax = b$ lösen.

$z_1 \leftarrow \frac{1}{4}z_1$:

$$\begin{array}{cccc} 1 & \frac{1}{4} & -\frac{1}{4} & 0 \\ -2 & 0 & 3 & 0 \\ 2 & 1 & 2 & 1 \end{array}$$

$z_2 \leftarrow z_2 + 2z_1, z_3 \leftarrow z_3 - 2z_1$:

$$\begin{array}{cccc} 1 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{5}{2} & 0 \\ 0 & \frac{1}{2} & \frac{5}{2} & 1 \end{array}$$

$z_2 \leftarrow 2z_2$:

$$\begin{array}{cccc} 1 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & 1 & 5 & 0 \\ 0 & \frac{1}{2} & \frac{5}{2} & 1 \end{array}$$

$z_3 \leftarrow z_3 - \frac{1}{2}z_2$:

$$\begin{array}{cccc} 1 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{array}$$

Das ist eine Halbdiagonalform mit $r = 2$. Da die 3. Koordinate der rechten Seite $\neq 0$ ist, gibt es keine Lösung.

3. Betrachten wir ein lineares Gleichungssystem mit nur einer Zeile:

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b$$

mit reellen Zahlen a_1, \dots, a_n, b . Wenn $a_1 \neq 0$ ist, brauchen wir nur mit $1/a_1$ zu multiplizieren und haben schon die Gestalt, bei der wir die Lösungsmenge ablesen können:

$$1 \quad \frac{a_2}{a_1} \quad \dots \quad \frac{a_n}{a_1} \quad \frac{b}{a_1}$$

Die Parameterdarstellung der Lösungsmenge lautet:

$$x = \begin{pmatrix} \frac{b}{a_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} -\frac{a_2}{a_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + \lambda_{n-1} \begin{pmatrix} -\frac{a_n}{a_1} \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Wenn alle $a_i = 0$ sind, so ist die Lösungsmenge entweder der ganze \mathbb{R}^n oder die leere Menge, je nachdem $b = 0$ ist oder nicht.

Natürlich kann man hier die Vektoren u_i mit a_1 multiplizieren (aber nicht den Vektor \hat{b}' !) und erhält dann:

$$x = \begin{pmatrix} \frac{b}{a_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} -a_2 \\ a_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + \lambda_{n-1} \begin{pmatrix} -a_n \\ 0 \\ \vdots \\ 0 \\ a_1 \end{pmatrix}.$$

Historische Bemerkung:

Obwohl das hier behandelte Eliminationsverfahren nach Gauß benannt wird, war es in China schon sehr viel früher bekannt. Unter dem Namen "Fang-Cheng-Regel" wird es in der "Mathematik in neun Büchern" praktisch in der heutigen Form erklärt. Dieses Werk ist angeblich bereits im 2. Jahrhundert v. Chr. vom chinesischen Kanzler Chang Ts'ang überarbeitet worden und wurde im Jahre 1084 erstmals gedruckt (vgl. [3]).

Bemerkung zur numerischen Lösung linearer Gleichungssysteme:

Da reelle Zahlen auf einem Computer im Allgemeinen nur näherungsweise dargestellt werden können, treten beim Rechnen damit notwendigerweise gewisse Fehler auf. Besonders große Fehler können sich ergeben, wenn im Laufe einer Rechnung durch Zahlen dividiert wird, die sehr nahe bei Null liegen. Um die Gefahr solcher *numerischer Fehler* möglichst gering zu halten, geht man bei der Umformung auf Halbdiagonalform folgenderweise vor: Im ersten Schritt dividiert man auch für $a_{11} \neq 0$ nicht einfach die 1. Zeile durch a_{11} , sondern man sucht zuerst entweder in der ersten Spalte oder in der ganzen Matrix ein Element a_{ik} mit größtmöglichen Absolutbetrag (dieses Element heißt *Pivot*) und vertauscht dann (für $i \neq 1$) die 1. Zeile mit der i -ten Zeile und (für $k \neq 1$) die 1. Spalte mit der k -ten Spalte. Im zweiten Schritt geht man analog vor, wobei sich die Suche aber nur auf Elemente a_{ik} mit $i \geq 2$ und $k \geq 2$ beschränkt, usw.. Natürlich muss man die durchgeführten Spaltenvertauschungen am Schluss berücksichtigen. Je nachdem, ob man jeweils nur in der betrachteten Spalte oder im ganzen in Frage kommenden Bereich sucht, spricht man von *Spaltenpivotsuche* oder *vollständiger Pivotsuche*. Näheres dazu erfährt man in Vorlesungen oder Büchern über Numerische Mathematik.

Bemerkung zur Behandlung von linearen Gleichungssystemen in der Schule:

Meistens werden in der Schule nur Gleichungssysteme mit zwei oder drei Unbekannten besprochen. Im Realgymnasium sind aber auch solche mit mehr Unbekannten vorgesehen. Das hat vor allem den Vorteil, dass die Systematik des Lösungsverfahrens und die allgemeine Gestalt der Lösungsmenge klarer sichtbar wird. Für größere Gleichungssysteme ist es allerdings wegen des Rechenaufwandes empfehlenswert, Taschenrechner oder Computer zu verwenden. Im Lehrplan ist das unter "allenfalls" angeführt. Das würde jedenfalls auch Gelegenheit bieten, auf die numerischen Probleme aufmerksam zu machen.

5 Lineare und affine Teilräume des \mathbb{R}^n

5.1 Definitionen

Eine nicht leere Teilmenge U des \mathbb{R}^n heißt *linearer Teilraum* oder *Untervektorraum* (des \mathbb{R}^n), wenn gilt:

$$\begin{aligned} x, y \in U \Rightarrow x + y \in U, \\ x \in U, \lambda \in \mathbb{R} \Rightarrow \lambda x \in U. \end{aligned}$$

Jeder lineare Teilraum U des \mathbb{R}^n ist selbst wieder ein Vektorraum, wie man sich leicht überlegen kann. Z.B. gilt $o \in U$ wegen $0x = o$, und für jedes $x \in U$ ist auch $-x = (-1)x \in U$.

Wichtiges *Beispiel*: Die Lösungsmenge eines homogenen linearen Gleichungssystems $Ax = o$ mit n Unbekannten ist ein linearer Teilraum des \mathbb{R}^n , denn wenn x° und y° Lösungen sind, dann gilt $Ax^\circ = o$ und $Ay^\circ = o$, und daher $A(x^\circ + y^\circ) = Ax^\circ + Ay^\circ = o + o = o$ und $A(\lambda x^\circ) = \lambda A(x^\circ) = \lambda o = o$.

Mit vollständiger Induktion sieht man leicht, dass ein linearer Teilraum mit je endlich vielen Vektoren auch alle Linearkombinationen davon enthält.

Sind andererseits u_1, \dots, u_s feste Vektoren des \mathbb{R}^n , so ist die Menge aller ihrer Linearkombinationen ein linearer Teilraum: Sei $U = \{x \in \mathbb{R}^n : x = \lambda_1 u_1 + \dots + \lambda_s u_s, \lambda_i \in \mathbb{R}\}$. Mit $x = \lambda_1 u_1 + \dots + \lambda_s u_s$ und $y = \mu_1 u_1 + \dots + \mu_s u_s$ gilt:

$$\begin{aligned} x + y &= (\lambda_1 + \mu_1) u_1 + \dots + (\lambda_s + \mu_s) u_s, \\ \lambda x &= (\lambda \lambda_1) u_1 + \dots + (\lambda \lambda_s) u_s. \end{aligned}$$

Die Menge aller Linearkombinationen von u_1, \dots, u_s nennt man auch die *lineare Hülle* dieser Vektoren oder den von diesen Vektoren *erzeugten* oder *aufgespannten linearen Teilraum*. Bezeichnung: $\text{lin}(u_1, \dots, u_s)$.

Ist U ein linearer Teilraum und p ein beliebiger Punkt des \mathbb{R}^n , so nennt man

$$p + U := \{p + u : u \in U\}$$

einen *affinen Teilraum* des \mathbb{R}^n . Das ist also ein translatierter linearer Teilraum: $p + U = \tau_p(U)$. Außerdem wird die leere Menge zu den affinen Teilräumen gerechnet. Die Lösungsmenge eines linearen Gleichungssystems $Ax = b$ mit einer $m \times n$ -Matrix A ist also in jedem Fall ein affiner Teilraum des \mathbb{R}^n .

Die Definition eines affinen Teilraums des \mathbb{R}^n kann man etwas ausführlicher auch so schreiben:

$$p + U := \{x \in \mathbb{R}^n : \text{es gibt ein } u \in U, \text{ sodass } x = p + u\}.$$

Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lineare oder affine Abbildung. Dann nennt man das Urbild des Nullvektors,

$$f^{-1}(\{o\}) = \{x \in \mathbb{R}^n : f(x) = o\},$$

den *Kern* von f (Bezeichnung: $\ker f$). Der Kern einer linearen bzw. affinen Abbildung ist ein linearer bzw. affiner Teilraum, denn für $f(x) = Ax + b$ ist $\ker f$ die Lösungsmenge des Gleichungssystems $Ax = -b$.

Der *Nullraum* $\{o\}$ und der ganze \mathbb{R}^n sind nach unserer Definition ebenfalls lineare Teilräume des \mathbb{R}^n .

Achtung: Auf englisch wird der Kern einer linearen Abbildung entweder *kernel* oder *null space* genannt, der Nullraum heißt dagegen *zero subspace*.

5.2 Basis, lineare Unabhängigkeit und Dimension

5.2.1 Fundamentale Begriffe und Sätze

Sei U ein linearer Teilraum des \mathbb{R}^n (eventuell $U = \mathbb{R}^n$). Eine endliche Teilmenge $B = \{b_1, \dots, b_s\}$ von U heißt *Basis* von U , wenn sich jeder Vektor aus U in eindeutiger Weise als Linearkombination von b_1, \dots, b_s darstellen lässt. Das heißt dann also: Zu jedem $x \in U$ gibt es eindeutig bestimmte Zahlen $\lambda_1, \dots, \lambda_s$, sodass $x = \lambda_1 b_1 + \dots + \lambda_s b_s$. Diese Zahlen $\lambda_1, \dots, \lambda_s$ heißen die

Koordinaten von x bezüglich der Basis B . Da es dabei natürlich auf die Reihenfolge ankommt, sollte man in diesem Zusammenhang besser eine *geordnete Basis* verwenden, also ein geordnetes s -Tupel von Vektoren (b_1, \dots, b_s) mit der oben genannten Eigenschaft.

Insbesondere bilden die kanonischen Basisvektoren e_1, \dots, e_n eine Basis des \mathbb{R}^n , und die diesbezüglichen Koordinaten stimmen mit den früher definierten gewöhnlichen Koordinaten überein, da ja $(x_1, \dots, x_n) = x_1e_1 + \dots + x_ne_n$.

Löst man ein homogenes lineares Gleichungssystem mit der oben angegebenen Methode, so bilden die Vektoren u_1, \dots, u_{n-r} eine Basis des Lösungsraums L , denn in diesem Fall ist $\hat{b} = b = o$, also

$$L = \{x \in \mathbb{R}^n : x = \lambda_1u_1 + \dots + \lambda_{n-r}u_{n-r} \text{ mit } \lambda_i \in \mathbb{R}\},$$

und die Koeffizienten λ_i sind durch x eindeutig bestimmt, da ja $\lambda_1 = x_{r+1}, \dots, \lambda_{n-r} = x_n$.

Wir halten fest: Die Lösungsmenge eines homogenen linearen Gleichungssystems mit n Unbekannten ist ein linearer Teilraum des \mathbb{R}^n , der eine Basis aus $n - r$ Vektoren besitzt, wobei r der Rang der zugehörigen Matrix ist. Für $r = n$ heißt das, dass die Lösungsmenge der Nullraum $\{o\}$ ist. Ein solches Gleichungssystem besitzt also genau dann nicht-triviale Lösungen, wenn $r < n$ ist.

Die Eindeutigkeit der Koeffizienten bei einer Linearkombination hängt eng mit folgendem Begriff zusammen: Die Vektoren a_1, \dots, a_s heißen *linear unabhängig*, wenn gilt:

$$(o = \lambda_1a_1 + \dots + \lambda_s a_s \text{ mit } \lambda_i \in \mathbb{R}) \Rightarrow (\lambda_1 = \dots = \lambda_s = 0),$$

das heißt, wenn sich der Nullvektor nur in *trivialer* Weise als Linearkombination von a_1, \dots, a_s darstellen lässt. Andernfalls heißen diese Vektoren *linear abhängig*. a_1, \dots, a_s sind also genau dann linear abhängig, wenn es reelle Zahlen $\lambda_1, \dots, \lambda_s$ gibt, welche nicht alle gleich Null sind, sodass $o = \lambda_1a_1 + \dots + \lambda_s a_s$, d.h., wenn sich der Nullvektor in *nicht-trivialer* Weise als Linearkombination von a_1, \dots, a_s darstellen lässt.

Es ist nun leicht einzusehen, dass s Vektoren genau dann linear unabhängig sind, wenn bei jeder Linearkombination dieser Vektoren die Koeffizienten eindeutig bestimmt sind:

Angenommen, a_1, \dots, a_s sind linear unabhängig und $\lambda_1a_1 + \dots + \lambda_s a_s = \mu_1a_1 + \dots + \mu_s a_s$. Dann folgt $o = (\lambda_1 - \mu_1)a_1 + \dots + (\lambda_s - \mu_s)a_s$, und daher $\lambda_1 - \mu_1 = \dots = \lambda_s - \mu_s = 0$, d.h. $\lambda_i = \mu_i$ für alle i . Wenn umgekehrt

jede Linearkombination eindeutig ist, so insbesondere dann, wenn sie den Nullvektor ergibt, und daraus folgt die lineare Unabhängigkeit.

Es gilt also:

Satz 5 Eine Teilmenge B eines Untervektorraums U des \mathbb{R}^n ist genau dann eine Basis von U , wenn sie linear unabhängig ist und U erzeugt.

Wie schon bemerkt wurde, kann U natürlich auch der ganze \mathbb{R}^n sein.

Bemerkung 6 Wenn die Vektoren a_1, \dots, a_s linear unabhängig sind (mit $s \geq 2$), dann sind sie paarweise verschieden.

Beweis: Nehmen wir o.B.d.A. an, dass $a_1 = a_2$ ist. Dann ist $o = 1a_1 - 1a_2 + 0a_3 + \dots + 0a_s$, und das ist eine nicht-triviale Linearkombination. \square

Eine endliche Teilmenge $\{a_1, \dots, a_s\}$ des \mathbb{R}^n heißt *linear unabhängig*, wenn ihre Elemente a_1, \dots, a_s linear unabhängig sind. Dabei nehmen wir natürlich an, dass keines der Elemente mehrfach aufgezählt ist, das heißt, dass a_1, \dots, a_s paarweise verschieden sind.

Beispiele:

1. Jede Teilmenge einer Basis eines (Unter-)Vektorraums ist linear unabhängig. Allgemeiner ist jede Teilmenge einer linear unabhängigen Menge von Vektoren linear unabhängig. (Das sieht man am besten mit einem indirekten Beweis.)
2. Zwei Vektoren a, b mit $b \neq o$ sind genau dann linear abhängig, wenn es ein $\lambda \in \mathbb{R}$ gibt, sodass $a = \lambda b$:

Angenommen, $o = \lambda_1 a + \lambda_2 b$, und λ_1, λ_2 sind nicht beide gleich Null. Es ist nicht möglich, dass $\lambda_1 = 0$ und $\lambda_2 \neq 0$ ist, denn dann wäre $o = \lambda_2 b$ im Widerspruch zu $b \neq o$. Also muss $\lambda_1 \neq 0$ sein, und daher können wir durch λ_1 dividieren und erhalten $a = -\frac{\lambda_2}{\lambda_1} b$.

Wenn umgekehrt $a = \lambda b$ ist, dann ist $o = 1a + (-\lambda)b$. Das ist eine nicht-triviale Linearkombination (da ja $1 \neq 0$), und daher sind a und b linear abhängig.

Anschaulich gesprochen, bedeutet die lineare Abhängigkeit von zwei Vektoren, dass sie *parallel* sind.

3. Ein einzelner Vektor ist genau dann linear abhängig, wenn er $= o$ ist: Aus $\lambda a = o$ mit $\lambda \neq 0$ folgt nach Division durch λ sofort $a = o$.
4. Wenn a_1, \dots, a_s linear unabhängig sind, dann sind alle $a_i \neq o$: Wäre z.B. $a_1 = o$, dann wäre $o = 1o + 0a_2 + \dots + 0a_s$ eine nicht-triviale Darstellung von o als Linearkombination von a_1, \dots, a_s .
5. Wenn a_1, \dots, a_s linear abhängig sind (mit $s \geq 2$), dann lässt sich wenigstens einer dieser Vektoren als Linearkombination der anderen ausdrücken, und umgekehrt. (Das entspricht vielleicht besser der intuitiven Vorstellung von Abhängigkeit als die Definition von "linear abhängig".)

Beweis: Sei $\lambda_1 a_1 + \dots + \lambda_s a_s = o$ und nicht alle $\lambda_i = 0$. Dann gibt es mindestens einen Index i_0 , sodass $\lambda_{i_0} \neq 0$. Wir können daher durch λ_{i_0} dividieren und erhalten mit $\mu_i := -\frac{\lambda_i}{\lambda_{i_0}}$:

$$a_{i_0} = \sum_{i \neq i_0} \mu_i a_i.$$

Wenn umgekehrt a_{i_0} so darstellbar ist, dann folgt $\mu_1 a_1 + \dots + \mu_s a_s = o$ mit $\mu_{i_0} := -1 \neq 0$, und daher sind a_1, \dots, a_s linear abhängig. \square

Will man zu gegebenen Vektoren a_1, \dots, a_s feststellen, ob sie linear abhängig sind, so geht es eigentlich um die Frage, ob das homogene Gleichungssystem mit Matrix $A = (a_1, \dots, a_s)$ und den Unbekannten $\lambda_1, \dots, \lambda_s$ eine nicht-triviale Lösung hat: $\lambda_1 a_1 + \dots + \lambda_s a_s = o$ ist gleichbedeutend mit

$$(a_1, \dots, a_s) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_s \end{pmatrix} = o.$$

Die Frage kann man also mit dem Gauß'schen Eliminationsverfahren beantworten.

Beispiel:

Sei

$$a_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \quad a_2 = \begin{pmatrix} -2 \\ 5 \\ 4 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 6 \\ 11 \\ 20 \end{pmatrix}.$$

Die Matrix des zugehörigen homogenen Gleichungssystems ist

$$A = \begin{pmatrix} 3 & -2 & 6 \\ -1 & 5 & 11 \\ 2 & 4 & 20 \end{pmatrix}.$$

Wir bringen sie auf zunächst auf Halbdiagonalform:

$$\begin{matrix} 1 & -\frac{2}{3} & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{matrix}$$

Hier sehen wir schon, dass die Vektoren a_1, a_2, a_3 linear abhängig sind, denn der Rang der Matrix ist gleich $2 < 3$, und daher hat das Gleichungssystem nicht-triviale Lösungen.

Wenn wir die Skalarfaktoren λ_i bestimmen wollen, so brauchen wir nur noch das Element $-\frac{2}{3}$ oberhalb der Diagonalen zu eliminieren:

$$\begin{matrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{matrix}$$

Die Lösungsmenge des Gleichungssystems besteht also aus den skalaren Vielfachen von $\begin{pmatrix} -4 \\ -3 \\ 1 \end{pmatrix}$. Wir können daher z.B. $\lambda_1 = 4, \lambda_2 = 3, \lambda_3 = -1$ nehmen. Zur Probe kann man leicht nachrechnen, dass tatsächlich $4a_1 + 3a_2 - a_3 = o$ ist.

Wenn ein linearer Teilraum U eine Basis aus s Vektoren besitzt, so sagt man, seine *Dimension* ist gleich s . Diese Definition ist allerdings insoferne etwas problematisch, als nicht von vornherein klar ist, ob U nicht auch eine Basis mit einer anderen Anzahl von Elementen besitzen kann. Zur Klärung dieser Frage überlegen wir uns zunächst folgenden Satz.

Satz 7 *Je $s + 1$ Linearkombinationen von s gegebenen Vektoren sind linear abhängig.*

Beweis: Seien c_1, \dots, c_{s+1} Vektoren, die sich als Linearkombinationen von b_1, \dots, b_s darstellen lassen. D.h. es gibt reelle Zahlen a_{ik} , sodass

$$c_k = a_{1k}b_1 + \dots + a_{sk}b_s \quad \text{für } k = 1, \dots, s+1.$$

Wir suchen nun reelle λ_k , sodass

$$\lambda_1 c_1 + \dots + \lambda_{s+1} c_{s+1} = o.$$

Setzen wir hier die obigen Ausdrücke für c_k ein und fassen die Terme mit gleichen b_i zusammen, so heißt das

$$\left(\sum_{k=1}^{s+1} a_{1k} \lambda_k \right) b_1 + \dots + \left(\sum_{k=1}^{s+1} a_{sk} \lambda_k \right) b_s = o.$$

Wenn es uns gelingt, reelle λ_k zu finden, die nicht alle gleich Null sind, sodass

$$\sum_{k=1}^{s+1} a_{1k} \lambda_k = 0,$$

⋮

$$\sum_{k=1}^{s+1} a_{sk} \lambda_k = 0,$$

dann sind wir fertig. Es geht also um eine nicht-triviale Lösung des homogenen Gleichungssystems $Ax = o$, wobei A die $s \times (s+1)$ -Matrix mit den Elementen a_{ik} ist. Sei r der Rang von A . Auf Grund des Gauß'schen Eliminationsverfahrens wissen wir, dass $r \leq s$ ist, und dass das Gleichungssystem $(s+1) - r$ linear unabhängige Lösungen besitzt. $(s+1) - r \geq 1$, d.h. es gibt mindestens einen linear unabhängigen Lösungsvektor, und der muss $\neq o$ sein. \square

Satz 8 *Je zwei Basen eines linearen Teilraums des \mathbb{R}^n haben gleich viele Elemente.*

Beweis: Angenommen, es gibt zwei Basen $\{b_1, \dots, b_s\}$ und $\{c_1, \dots, c_t\}$ von U mit verschieden vielen Elementen. Ohne Beschränkung der Allgemeinheit sei $s < t$.

Da $\{b_1, \dots, b_s\}$ eine Basis von U ist, lassen sich die c_i als Linearkombinationen von b_1, \dots, b_s darstellen. Dann sind aber c_1, \dots, c_{s+1} $s+1$ Linearkombinationen von s Vektoren und somit nach dem vorigen Satz linear abhängig. Das steht im Widerspruch dazu, dass jede Teilmenge einer Basis linear unabhängig ist. \square

Wir wissen jetzt also, dass die Anzahl der Elemente einer Basis eines linearen Teilraums eindeutig bestimmt ist. Sind wir aber sicher, dass es in jedem linearen Teilraum überhaupt eine Basis gibt? Darauf gibt der folgende Satz Antwort.

Satz 9 Außer dem Nullraum besitzt jeder lineare Teilraum des \mathbb{R}^n eine Basis.

Beweis: Sei $U \neq \{o\}$ ein linearer Teilraum des \mathbb{R}^n . Nach Satz 7 hat eine linear unabhängige Teilmenge von U sicher nicht mehr als n Elemente. Andererseits gibt es jedenfalls linear unabhängige Teilmengen von U , z.B. ist jede einelementige Teilmenge $\{a\}$ mit $a \neq o$ linear unabhängig. Jetzt kommt der entscheidende Gedanke: Unter den linear unabhängigen Teilmengen von U gibt es sicher (mindestens) eine mit einer *maximalen Anzahl* von Elementen, sagen wir $B = \{u_1, \dots, u_s\}$ mit $s \leq n$. Wir überlegen uns, dass B eine Basis ist. Dazu ist wegen Satz 5 nur zu zeigen, dass sich jeder Vektor $x \in U$ als Linearkombination der Vektoren aus B darstellen lässt. Für $x \in B$ ist das trivial, z.B. gilt für $x = u_1$:

$$x = 1u_1 + 0u_2 + \dots + 0u_s.$$

Sei nun $x \notin B$. Dann ist wegen der Maximalität von s die Menge $\{u_1, \dots, u_s, x\}$ linear abhängig, d.h. es gibt $\lambda_1, \dots, \lambda_{s+1}$ (nicht alle = 0), sodass

$$\lambda_1 u_1 + \dots + \lambda_s u_s + \lambda_{s+1} x = o.$$

$\lambda_{s+1} \neq 0$, denn sonst wären ja u_1, \dots, u_s linear abhängig. Wir können daher durch λ_{s+1} dividieren und erhalten

$$x = \left(\frac{-\lambda_1}{\lambda_{s+1}} \right) u_1 + \dots + \left(\frac{-\lambda_s}{\lambda_{s+1}} \right) u_s,$$

was zu zeigen war. \square

Bemerkung: Die in diesem Beweis verwendete Methode wird später noch mehrmals angewendet.

Wir können also jetzt wirklich sagen: Unter der *Dimension eines linearen Teilraums U* verstehen wir die Anzahl der Elemente einer (beliebigen) Basis von U . Bezeichnung: $\dim U$.

Insbesondere ergibt sich: $\dim \mathbb{R}^n = n$. Natürlicherweise setzt man $\dim \{o\} := 0$.

Wir können jetzt auch sofort die Gültigkeit der folgenden beiden Sätze erkennen:

Satz 10 Die Dimension eines linearen Teilraums U des \mathbb{R}^n ist gleich der maximalen Anzahl von linear unabhängigen Vektoren aus U .

Satz 11 Sei U ein s -dimensionaler linearer Teilraum des \mathbb{R}^n . Dann ist jede s -elementige linear unabhängige Teilmenge von U eine Basis von U .

Wenn man also z.B. wissen will, ob eine n -elementige Teilmenge des \mathbb{R}^n eine Basis des \mathbb{R}^n bildet, so braucht man nur deren lineare Unabhängigkeit zu überprüfen.

Später werden wir gelegentlich die folgende Verschärfung von Satz 9 brauchen, die auch für den Fall $U = \mathbb{R}^n$ interessant ist.

Satz 12 (Basisergänzungssatz) Sei $\{u_1, \dots, u_k\}$ eine linear unabhängige Teilmenge eines s -dimensionalen linearen Teilraums U des \mathbb{R}^n . Dann gibt es $s-k$ Vektoren u_{k+1}, \dots, u_s , sodass $\{u_1, \dots, u_k, u_{k+1}, \dots, u_s\}$ eine Basis von U ist.

(Das heißt also: Man kann jede linear unabhängige Teilmenge von U zu einer Basis von U ergänzen.)

Beweis: Sei B eine linear unabhängige Teilmenge von U , welche $\{u_1, \dots, u_k\}$ enthält und eine maximale Anzahl von Elementen besitzt. Dann können wir wie bei Satz 9 schließen, dass B eine Basis ist, und B enthält natürlich die Vektoren u_1, \dots, u_k . \square

Bemerkung zu Verallgemeinerungen der Begriffe und Sätze dieses Abschnitts:

Alles, was hier über lineare Teilräume des \mathbb{R}^n gesagt wurde, lässt sich ohne weiteres auf beliebige *endlichdimensionale* Vektorräume übertragen. Z.B. ist jede n -elementige linear unabhängige Teilmenge eines n -dimensionalen Vektorraums V (über einem beliebigen Körper) eine Basis von V (vgl. Satz 11).

Es gibt aber auch *unendlichdimensionale* Vektorräume, das sind also solche, in denen es für jede natürliche Zahl n linear unabhängige Teilmengen mit n Elementen gibt. Z.B. bildet die Menge aller Funktionen $\mathbb{R} \rightarrow \mathbb{R}$ einen solchen Vektorraum: Die Addition und Skalarmultiplikation wird ganz einfach "punktweise" definiert:

$$\begin{aligned} f + g : x &\mapsto f(x) + g(x) \\ \lambda f : x &\mapsto \lambda (f(x)). \end{aligned}$$

Der Nullvektor ist natürlich die durch $o(x) := 0$ für alle $x \in \mathbb{R}$ definierte Funktion o . In diesem Vektorraum sind z.B. je endlich viele (verschiedene) Funktionen der folgenden Art linear unabhängig:

$$g_a(x) := \begin{cases} 1 & \text{für } x = a, \\ 0 & \text{für } x \neq a. \end{cases}$$

Beweis: Angenommen, $\lambda_1 g_{a_1} + \dots + \lambda_n g_{a_n} = o$, d.h. $\lambda_1 g_{a_1}(x) + \dots + \lambda_n g_{a_n}(x) = 0$ für alle $x \in \mathbb{R}$. Setzen wir hier a_i für x ein, so folgt $\lambda_i = 0$. \square

In unendlichdimensionalen Vektorräumen ist schon die Existenz einer Basis nicht ganz einfach zu beweisen, man benötigt dazu das sogenannte "Auswahlaxiom" (oder einen dazu äquivalenten Satz). Näheres dazu kann man z.B. in Vorlesungen über Mengenlehre oder Funktionalanalysis erfahren.

5.2.2 Berechnung der Koordinaten bezüglich einer gegebenen Basis

Ein linearer Teilraum U des \mathbb{R}^n sei durch eine geordnete Basis $B = (b_1, \dots, b_s)$ gegeben. Wie kann man die Koordinaten eines gegebenen Punktes $x \in U$ bezüglich dieser Basis berechnen? Das lässt sich, ähnlich wie die Überprüfung der linearen Unabhängigkeit, auf die Lösung eines linearen Gleichungssystems zurückführen.

Sei also $x = (x_1, \dots, x_n)$. Wir suchen $x'_1, \dots, x'_s \in \mathbb{R}$, sodass $x = x'_1 b_1 + \dots + x'_s b_s$. Fassen wir B als Matrix mit den Spaltenvektoren b_1, \dots, b_s auf, so heißt das mit $x' := (x'_1, \dots, x'_s)$

$$Bx' = x.$$

Das ist also ein lineares Gleichungssystem mit den Unbekannten x'_1, \dots, x'_s und gegebener rechter Seite x . Auf diese Weise kann man auch überprüfen, ob x wirklich in U liegt: Das ist genau dann der Fall, wenn dieses Gleichungssystem eine Lösung hat.

Beispiel:

Sei $U = \text{lin}((1, 0, 2), (-1, 1, 3))$ im \mathbb{R}^3 und $x = (-5, 2, 0)$.

Die Matrix und die rechte Seite des entsprechenden Gleichungssystems sehen so aus:

$$B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 2 & 3 \end{pmatrix}, \quad x = \begin{pmatrix} -5 \\ 2 \\ 0 \end{pmatrix}$$

Mit dem Gauß'schen Eliminationsverfahren erhalten wir: $x'_1 = -3$, $x'_2 = 2$. Wir können zur Probe nachrechnen: $(-3)(1, 0, 2) + 2(-1, 1, 3) = (-5, 2, 0)$.

Mehr zu diesem Thema findet sich in Kapitel 8.

5.2.3 Dimension affiner Teilräume und lineare Gleichungssysteme

Unter der *Dimension eines affinen Teilraums* $L = p + U$ versteht man die Dimension des zugehörigen linearen Teilraums U . Insbesondere ist jede einelementige Teilmenge $\{p\}$ des \mathbb{R}^n ein affiner Teilraum mit Dimension 0, denn $\{p\} = p + \{o\}$. Zusätzlich definiert man $\dim \emptyset := -1$.

Auf Grund der folgenden Bemerkung ist die Definition der Dimension eines affinen Teilraums wirklich sinnvoll.

Bemerkung 13 *Der zu einem affinen Teilraum gehörige lineare Teilraum ist eindeutig bestimmt.*

Beweis: Aus $L = p + U = \tau_p(U)$ folgt ja $U = \tau_p^{-1}(L)$. (Wir können auch schreiben: $U = \tau_{-p}(L) = L - p$.)

(Zum hier verwendeten Begriff der Umkehrfunktion siehe Kapitel 6.)

Im Gegensatz dazu ist der Punkt p bei $L = p + U$ für $\dim L > 0$ nicht eindeutig bestimmt. Statt p kann man nämlich jeden beliebigen Punkt von L nehmen:

Bemerkung 14 *Sei $L = p + U$. Dann gilt $L = q + U$ für alle $q \in L$.*

Beweis: Sei $q \in L = p + U$. Dann gibt es ein $u \in U$, sodass $q = p + u$. Wir zeigen nun, dass $L \subset q + U$ und $q + U \subset L$ gilt, woraus natürlich $L = q + U$ folgt.

$L \subset q + U$: Zu jedem $x \in L$ gibt es ein $v \in U$, sodass $x = p + v$. Dann ist aber $x = (q - u) + v = q + (v - u) \in q + U$.

$q + U \subset L$: Wenn $y \in q + U$, so gibt es ein $w \in U$ mit $y = q + w$, also $y = p + u + w \in p + U = L$. \square

Unser Wissen über die Lösungsmenge eines linearer Gleichungssystems können wir nun folgendermaßen ausdrücken:

Satz 15 *Die Lösungsmenge eines linearen Gleichungssystems $Ax = b$ mit n Unbekannten ist entweder leer oder ein affiner Teilraum des \mathbb{R}^n mit Dimension $n - r$, wobei r der Rang von A ist.*

Die Unterscheidung dieser beiden Fälle wird noch etwas klarer, wenn man die Matrix betrachtet, welche aus A durch Hinzufügung der Spalte b entsteht. Man nennt das die *erweiterte Matrix* des Gleichungssystems. Wir bezeichnen sie mit (A, b) . Beim Gauß'schen Eliminationsverfahren wird eigentlich diese erweiterte Matrix umgeformt. Vertauschen wir bei der Halbdiagonalform die $(r+1)$ -te Spalte mit der letzten Spalte (das ist \tilde{b}), so sehen wir, dass folgendes gilt:

Satz 16 *Die Lösungsmenge eines linearen Gleichungssystems $Ax = b$ ist genau dann nicht leer, wenn der Rang von A mit dem Rang der erweiterten Matrix (A, b) übereinstimmt. Andernfalls ist $\text{rank}(A, b) = \text{rank } A + 1$.*

Da ein homogenes lineares Gleichungssystem immer zumindest die triviale Lösung hat, gilt:

Satz 17 *Die Lösungsmenge eines homogenen linearen Gleichungssystems $Ax = 0$ mit n Unbekannten ist stets ein linearer Teilraum mit Dimension $n - \text{rank } A$.*

Anders ausgedrückt, heißt das:

$$\dim(\ker f_A) = n - \text{rank } A$$

bzw. $\text{rank } A = n - \dim(\ker f_A)$. Daraus folgt auf Grund der Eindeutigkeit der Dimension auch die Eindeutigkeit des Rangs. Die Zahl r in der Halbdiagonalform ist also insbesondere unabhängig von der Art der Pivotsuche.

Wir können jetzt auch leicht sagen, wann ein lineares Gleichungssystem genau eine Lösung hat. Das ist nämlich gleichbedeutend damit, dass die Dimension der Lösungsmenge 0 ist. Wir sagen in diesem Fall, die Lösung ist *eindeutig*. Wir haben also:

Satz 18 *Ein lineares Gleichungssystem $Ax = b$ mit n Unbekannten hat genau dann eine eindeutige Lösung, wenn $\text{rank } A = \text{rank}(A, b) = n$ ist.*

Wenn die Anzahl der Gleichungen mit der Anzahl der Unbekannten übereinstimmt, können wir das noch etwas vereinfachen, denn in diesem Fall kann der Rang von (A, b) nicht größer als n sein:

Satz 19 Ein lineares Gleichungssystem $Ax = b$ mit n Gleichungen und n Unbekannten hat genau dann eine eindeutige Lösung, wenn $\text{rank } A = n$ ist.

Ein affiner Teilraum L des \mathbb{R}^n heißt

- | | |
|--------------------|-------------------------|
| <i>Punkt,</i> | wenn $\dim L = 0$; |
| <i>Gerade,</i> | wenn $\dim L = 1$; |
| <i>Ebene,</i> | wenn $\dim L = 2$; |
| <i>Hyperebene,</i> | wenn $\dim L = n - 1$. |

Die Bezeichnung "Punkt" wird also sowohl für die Elemente des \mathbb{R}^n als auch für seine einelementigen Teilmengen verwendet.

Eine Hyperebene im \mathbb{R}^1 ist ein Punkt, im \mathbb{R}^2 eine Gerade und im \mathbb{R}^3 eine Ebene.

Die Lösungsmenge einer einzelnen linearen Gleichung $a_1x_1 + \dots + a_nx_n = b$, bei der nicht alle Koeffizienten a_i gleich Null sind, ist eine Hyperebene (siehe Beispiel 3 in Kapitel 4.3).

Eine Parameterdarstellung einer Ebene sieht so aus:

$$x = p + \lambda u + \mu v.$$

Dabei ist p ein Punkt auf der Ebene, und u, v sind linear unabhängige Vektoren, die den zugehörigen linearen Teilraum aufspannen.

5.3 Ergänzungen zum Rang einer Matrix

Der folgende Satz zeigt besonders klar, um was es sich beim Rang einer Matrix handelt. Außerdem wird die schon vorhin erwähnte Eindeutigkeit des Rangs nochmals bestätigt.

Satz 20 Der Rang einer $m \times n$ -Matrix A ist gleich der Dimension von $f_A(\mathbb{R}^n)$.

Beweis: Sei $f := f_A$ und $\{u_1, \dots, u_s\}$ eine Basis von $\ker f$. Nach dem Basisergänzungssatz (Satz 12) kann diese zu einer Basis $\{u_1, \dots, u_s, u_{s+1}, \dots, u_n\}$ des \mathbb{R}^n ergänzt werden. Zu jedem $x \in \mathbb{R}^n$ gibt es also eindeutig bestimmte $\xi_1, \dots, \xi_n \in \mathbb{R}$, sodass

$$x = \xi_1 u_1 + \dots + \xi_n u_n.$$

Wegen $f(u_1) = \dots = f(u_s) = o$ folgt daraus

$$f(x) = \xi_{s+1}f(u_{s+1}) + \dots + \xi_nf(u_n).$$

Da hier für ξ_{s+1}, \dots, ξ_n beliebige reelle Zahlen auftreten können, sehen wir: $f(\mathbb{R}^n) = \text{lin}(f(u_{s+1}), \dots, f(u_n))$.

$f(\mathbb{R}^n)$ ist also jedenfalls ein linearer Teilraum des \mathbb{R}^m (da f eine Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^m$ ist), und wenn wir zeigen können, dass die Vektoren $f(u_{s+1}), \dots, f(u_n)$ linear unabhängig sind, so folgt $\dim f(\mathbb{R}^n) = n - s = n - \dim(\ker f_A) = \text{rank } A$.

Angenommen, $\lambda_{s+1}f(u_{s+1}) + \dots + \lambda_nf(u_n) = o$, d.h. $f(\lambda_{s+1}u_{s+1} + \dots + \lambda_nu_n) = o$. Dann ist $\lambda_{s+1}u_{s+1} + \dots + \lambda_nu_n \in \ker f$, also gibt es $\lambda_1, \dots, \lambda_s \in \mathbb{R}$, sodass

$$\lambda_{s+1}u_{s+1} + \dots + \lambda_nu_n = \lambda_1u_1 + \dots + \lambda_su_s,$$

d.h.

$$\lambda_1u_1 + \dots + \lambda_su_s - \lambda_{s+1}u_{s+1} - \dots - \lambda_nu_n = o.$$

Da $\{u_1, \dots, u_n\}$ eine Basis ist, folgt daraus: alle λ_i sind = 0, was zu zeigen war. \square

Bemerkung: $f(\mathbb{R}^n)$ heißt *Bildraum* von f oder kurz *Bild* von f (engl. *image of f*) und wird auch mit $\text{im } f$ bezeichnet.

Die Bemerkung im Anschluss an Satz 17 kann man daher auch so ausdrücken: Für jede auf dem \mathbb{R}^n definierte lineare Abbildung f gilt:

$$\dim(\ker f) + \dim(\text{im } f) = \dim \mathbb{R}^n.$$

$\dim(\text{im } f)$ nennt man auch den *Rang von f* (Bezeichnung: $\text{rank } f$). Es gilt also: $\text{rank } f_A = \text{rank } A$.

Die letzte Formel kann man auch so interpretieren: Die Dimension des Kerns von f gibt an, um wieviel die Dimension von \mathbb{R}^n bei Anwendung der Abbildung f kleiner wird. Das erklärt vielleicht die manchmal verwendete Bezeichnung "Defekt von f " für $\dim(\ker f)$.

In der Halbdiagonalform \hat{A} von Kapitel 4.3 sehen wir, dass es r linear unabhängige Zeilen gibt, nämlich die ersten r Zeilen, und dass je $r+1$ Zeilen linear abhängig sind (da mindestens eine davon eine Nullzeile ist). Ebenso sieht man, dass es r linear unabhängige Spalten gibt und dass je $r+1$ Spalten linear abhängig sind (da es sich ja um Linearkombinationen der ersten r Spalten handelt). Man kann nun zeigen, dass diese Eigenschaften unabhängig von den elementaren Umformungen sind, durch die \hat{A} aus A entstanden ist. Es gelten also die folgenden beiden Sätze, von denen wir nur den zweiten beweisen:

Satz 21 Der Rang einer Matrix A ist gleich der maximalen Anzahl von linear unabhängigen Zeilen von A .

Satz 22 Der Rang einer Matrix A ist gleich der maximalen Anzahl von linear unabhängigen Spalten von A .

Beweis:

Seien a_{i_1}, \dots, a_{i_s} s linear unabhängige Spalten der $m \times n$ -Matrix A , und je $s+1$ Spalten von A seien linear abhängig. Wir überlegen uns, dass $B := \{a_{i_1}, \dots, a_{i_s}\}$ eine Basis von $\text{im } f_A$ ist. Dann folgt nach Satz 20 $s = \text{rank } A$.

Wegen der linearen Unabhängigkeit der a_{i_k} ist nur zu zeigen, dass sie den Bildraum erzeugen. Sei y ein beliebiges Element von $\text{im } f_A$. Dann gibt es ein $x \in \mathbb{R}^n$, sodass $y = Ax = x_1 a_1 + \dots + x_n a_n$.

Sei a_k eine Spalte von A , die nicht zu B gehört. Wegen der Maximalität von s sind die Vektoren $a_{i_1}, \dots, a_{i_s}, a_k$ linear abhängig. Wie beim Beweis von Satz 9 folgt daraus, dass a_k eine Linearkombination von a_{i_1}, \dots, a_{i_s} ist. Wir sehen daher, dass jeder Spaltenvektor von A entweder zu B gehört oder eine Linearkombination von Elementen aus B ist. Daher ist auch y eine Linearkombination von Elementen aus B , was zu zeigen war. \square

Beispiel: Sei

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

Der Rang dieser Matrix ist gleich 2. Das kann man auf verschiedene Weise begründen:

1. Die Halbdiagonalform lautet:

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

(Hier wurde eine Spaltenvertauschung durchgeführt!)

2. Die maximale Anzahl von linear unabhängigen Zeilen ist 2, denn es sind sogar je 2 Zeilen linear unabhängig, aber alle drei Zeilen z_1, z_2, z_3 sind linear abhängig, da z.B. $z_1 + z_2 - z_3 = o$.

3. Die maximale Anzahl von linear unabhängigen Spalten ist 2, denn z.B. sind die 1. und 3. Spalte linear unabhängig (ebenso die 2. und 3. Spalte, aber

keine anderen zwei Spalten), aber je drei Spalten sind linear abhängig: Die ersten drei Spalten sind linear abhängig, da es bereits die ersten beiden sind, und je drei andere Spalten sind linear abhängig, da eine davon die Nullspalte ist.

5.4 Schnitt von affinen Teilräumen

Wenn zwei affine Teileräume L_1, L_2 des \mathbb{R}^n als Lösungsmengen von linearen Gleichungssystemen $A_1x = b_1$ und $A_2x = b_2$ gegeben sind, so ist der Durchschnitt $L_1 \cap L_2$ die Lösungsmenge der "Zusammenfassung" der beiden Gleichungssysteme, d.h. von $Ax = b$ mit

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Wir haben hier erstmals eine Schreibweise verwendet, bei der eine Matrix durch Zusammenfassung von Teilmatrizen angegeben wird. Man spricht in diesem Zusammenhang auch von *Blockmatrizen*.

Insbesondere können wir also die Lösungsmenge eines Gleichungssystems mit m Gleichungen und n Unbekannten als Durchschnitt der m Hyperebenen ansehen, die den einzelnen Gleichungen entsprechen (vorausgesetzt, dass bei keiner Gleichung alle Koeffizienten gleich Null sind).

Häufig ist von zwei affinen Teileräumen einer in Parameterdarstellung gegeben und der andere durch ein lineares Gleichungssystem. Wir behandeln zunächst einen wichtigen Spezialfall:

5.4.1 Schnitt einer Geraden mit einer Hyperebene

Sei G eine Gerade mit Parameterdarstellung $x = p + \lambda u$, und H eine Hyperebene, die durch die Gleichung $ax = b$ mit einer einzeiligen Matrix $a = (a_1, \dots, a_n)$ und einer reellen Zahl b gegeben ist. Dann besteht der Durchschnitt $G \cap H$ genau aus den Punkten $p + \lambda u \in \mathbb{R}^n$, für die gilt

$$a(p + \lambda u) = b,$$

also

$$\lambda au = b - ap.$$

(Bei au und ap handelt es sich um Produkte einer einzeiligen Matrix mit einem Vektor. Das ergibt jeweils einen Vektor aus dem \mathbb{R}^1 , der mit der entsprechenden Zahl identifiziert werden kann.)

Wir unterscheiden jetzt zwei Fälle:

1. Fall: $au \neq 0$.

In diesem Fall erhalten wir für λ eine eindeutige Lösung λ_0 , nämlich

$$\lambda_0 = \frac{b - ap}{au}.$$

Der Durchschnitt von G und H besteht daher aus genau einem Punkt:

$$G \cap H = \{x_0\} \quad \text{mit } x_0 = p + \lambda_0 u.$$

Dieser Punkt heißt der *Schnittpunkt* von G mit H .

2. Fall: $au = 0$.

In diesem Fall ist die Gleichung $\lambda au = b - ap$ entweder für alle $\lambda \in \mathbb{R}$ erfüllt oder für keines, je nachdem $b - ap = 0$ ist oder nicht, d.h. je nachdem der Punkt p auf der Hyperebene H liegt oder nicht. Das heißt also, $G \cap H$ ist entweder die ganze Gerade G (d.h. $G \subset H$) oder die leere Menge.

Beispiel:

Sei G die Gerade im \mathbb{R}^3 mit Parameterdarstellung $x = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$

und H die Ebene mit Gleichung $x_1 + 4x_2 = 1$, also $a = (1, 4, 0)$ und $b = 1$. Hier ist $au = 2$, also liegt der 1. Fall vor, und wir erhalten $\lambda_0 = \frac{1-5}{2} = -2$

und damit $x_0 = \begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix}$.

Betrachten wir die Ebene H' mit Gleichung $x_1 + 2x_3 = 0$, also $a' = (1, 0, 2)$, so ist $a'u = 0$, und es liegt der 2. Fall vor. Da $p = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ nicht auf dieser Ebene liegt, ist $G \cap H' = \emptyset$.

5.4.2 Schnitt von zwei beliebigen affinen Teilmengen

Zur Bestimmung des Durchschnitts von zwei affinen Teilmengen, die nicht beide durch ein lineares Gleichungssystem gegeben sind, kann man die im Folgenden beschriebene Methode verwenden.

5.4.3 Berechnung eines Gleichungssystems zu gegebener Parameterdarstellung

Sei L ein s -dimensionaler affiner Teilraum des \mathbb{R}^n mit Parameterdarstellung $x = p + \lambda_1 c_1 + \dots + \lambda_s c_s$ (also jedenfalls $L \neq \emptyset$). Wir versuchen, ein lineares Gleichungssystem $Ax = b$ zu finden, dessen Lösungsmenge genau L ist. Der Rang von A muss dann $n-s$ sein, daher probieren wir es mit einer $(n-s) \times n$ -Matrix.

Wir wollen erreichen, dass $Ac_i = o$ ist für alle $i \in \{1, \dots, s\}$. Bezeichnen wir mit $C = (c_1, \dots, c_s)$ die $n \times s$ -Matrix mit Spaltenvektoren c_i , so heißt das

$$AC = O.$$

Bezeichnen wir die Zeilenvektoren von A mit a_1^*, \dots, a_{n-s}^* , so können wir das auch so schreiben:

$$\begin{pmatrix} a_1^* \\ \vdots \\ a_{n-s}^* \end{pmatrix} (c_1, \dots, c_s) = O,$$

das heißt aber

$$a_i^* c_k = 0 \quad \text{für alle } i \in \{1, \dots, n-s\} \text{ und } k \in \{1, \dots, s\}.$$

Bezeichnen wir mit c'_k den Vektor c_k , aber als Zeile geschrieben, und andererseits mit a_i^{*l} den Vektor a_i^* als Spalte geschrieben, so heißt das: Wir suchen linear unabhängige (Spalten-)Vektoren a_i^{*l} , sodass

$$c'_k a_i^{*l} = 0 \quad \text{für alle } k \in \{1, \dots, s\},$$

also

$$\begin{pmatrix} c'_1 \\ \vdots \\ c'_s \end{pmatrix} a_i^{*l} = o.$$

Die a_i^{*l} sind also Lösungen des homogenen Gleichungssystems mit der Matrix

$$C^T = \begin{pmatrix} c'_1 \\ \vdots \\ c'_s \end{pmatrix}.$$

Diese Matrix heißt die zu C transponierte Matrix. Die Zeilen von C^T entsprechen genau den Spalten von C , und man sieht leicht, dass für die Elemente c_{ik}^T von C^T gilt:

$$c_{ik}^T = c_{ki}.$$

Daraus erkennt man auch, dass die Spalten von C^T genau den Zeilen von C entsprechen. Man kann auch sagen: C^T entsteht aus C durch Spiegelung an der Hauptdiagonalen.

Da die Vektoren c'_1, \dots, c'_s linear unabhängig sind, hat C^T den Rang s , und daher hat das zugehörige homogene lineare Gleichungssystem $n-s$ linear unabhängige Lösungen, die wir für $a_1^{*'}, \dots, a_{n-s}^{*'}$ nehmen können. Damit haben wir dann schon eine geeignete Matrix A gefunden. Jetzt brauchen wir nur noch b so zu bestimmen, dass der Punkt p eine Lösung von $Ax = b$ ist. Das ist aber ganz einfach: wir setzen

$$b := Ap,$$

und damit ist die Aufgabe im Wesentlichen gelöst. Genaueres ergibt sich aus dem folgenden Satz und seinem Beweis.

Satz 23 *Jeder s -dimensionale affine Teilraum des \mathbb{R}^n (mit $s \geq 0$) lässt sich als Lösungsmenge eines linearen Gleichungssystems mit $n-s$ Gleichungen darstellen.*

Beweis: Sei L ein s -dimensionaler affiner Teilraum des \mathbb{R}^n mit Parameterdarstellung $x = p + \lambda_1 c_1 + \dots + \lambda_s c_s$. Wie wir soeben gesehen haben, gibt es eine $(n-s) \times n$ -Matrix A mit Rang $n-s$, sodass $Ac_i = o$ ist für alle $i \in \{1, \dots, s\}$. Bezeichnen wir die Lösungsmenge von $Ax = o$ mit U , so heißt das: $\text{lin}(c_1, \dots, c_s) \subset U$. Hier gilt aber sogar Gleichheit, denn $\dim \text{lin}(c_1, \dots, c_s) = s = n - \text{rank } A = \dim U$, und daher ist $\{c_1, \dots, c_s\}$ eine Basis von U (siehe Satz 11).

Sei nun $b := Ap$. Da p eine partikuläre Lösung von $Ax = b$ ist, ist $p + U = p + \text{lin}(c_1, \dots, c_s) = L$ die gesamte Lösungsmenge. \square

Bemerkung: Zu einem affinen Teilraum L gibt es im Allgemeinen viele verschiedene lineare Gleichungssysteme, deren Lösungsmenge gleich L ist. Im Falle einer Hyperebene ist die Gleichung allerdings bis auf einen Skalarfaktor eindeutig bestimmt, denn in diesem Fall ist C^T eine $(n-1) \times n$ -Matrix mit Rang $n-1$, und daher ist die Lösungsmenge von $C^T x = o$ eindimensional.

Beispiel:

Sei L eine Ebene des \mathbb{R}^4 , die folgenderweise durch eine Parameterdarstellung gegeben ist:

$$x = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix},$$

also

$$C^T = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 \end{pmatrix},$$

und wir lösen das Gleichungssystem $C^T x = o$.

Da hier C^T bereits Halbdiagonalf orm hat, brauchen wir nur noch einen Eliminationsschritt: wir subtrahieren das 2-fache der 2. Zeile von der 1. Zeile und erhalten

$$\begin{pmatrix} 1 & 0 & -1 & -4 \\ 0 & 1 & 1 & 2 \end{pmatrix}$$

Jetzt können wir die Lösungsvektoren bereits ablesen:

$$a_1^{*\prime} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad a_2^{*\prime} = \begin{pmatrix} 4 \\ -2 \\ 0 \\ 1 \end{pmatrix},$$

und wir erhalten

$$A = \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 4 & -2 & 0 & 1 \end{pmatrix}$$

und

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

Wenn man also z.B. die Ebene L mit der Hyperebene H schneiden will, die durch die Gleichung $x_1 + x_2 + x_3 + x_4 = 5$ gegeben ist, so löst man jetzt das Gleichungssystem

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

Man erhält die Gerade mit folgender Parameterdarstellung:

$$x = \frac{1}{4} \begin{pmatrix} 5 \\ 0 \\ 11 \\ 4 \end{pmatrix} + \nu \begin{pmatrix} 1 \\ 1 \\ 0 \\ -2 \end{pmatrix}.$$

Zur Transposition von Matrizen:

1. *Rechenregel bezüglich des Matrizenprodukts:*

$$(AB)^T = B^T A^T.$$

Beweis: Sei $C = AB$, also $c_{ik} = \sum_j a_{ij}b_{jk}$. Dann gilt

$$c_{ik}^T = c_{ki} = \sum_j a_{kj}b_{ji} = \sum_j a_{jk}^T b_{ij}^T = \sum_j b_{ij}^T a_{jk}^T. \quad \square$$

2. *Rang der transponierten Matrix:* Da der Rang einer Matrix sowohl durch die Zeilen wie durch die Spalten charakterisiert werden kann (siehe Sätze 21 und 22), gilt

$$\text{rank } A^T = \text{rank } A.$$

Der Rang bleibt daher auch bei allen *elementaren Spaltenumformungen* unverändert, welche analog zu den *elementaren Zeilenumformungen* von Abschnitt 4.1 definiert sind. Das kann manchmal zu einer schnelleren Bestimmung des Rangs verwendet werden.

Bemerkung zur Behandlung des Stoffes in der Schule:

Von dem im letzten Abschnitt behandelten Thema wird in der Schule normalerweise nur der Fall diskutiert, dass eine Ebene im \mathbb{R}^3 durch eine Parameterdarstellung $x = p + \lambda u + \mu v$ gegeben und eine zugehörige Gleichung gesucht ist. In diesem Fall braucht man nicht die oben angegebene Methode. Man kann statt dessen mit dem Vektorprodukt einen Normalvektor $n = u \times v$ der Ebene berechnen und erhält dann folgende Gleichung: $n \cdot x = n \cdot p$. (Siehe Kapitel 10.) Als Lehrer sollte man aber doch wissen, wie man allgemein zu einem in Parameterdarstellung gegebenen affinen Teilraum ein zugehöriges Gleichungssystem berechnen kann.

6 Inversion von Matrizen

Seien M, N beliebige Mengen. Eine Abbildung $f : M \rightarrow N$ heißt *bijektiv* oder *umkehrbar eindeutig*, wenn es zu jedem Element $y \in N$ genau ein $x \in M$ gibt, das durch f auf y abgebildet wird, sodass also $y = f(x)$ ist. Zu einer Abbildung gibt es daher eine *Umkehrabbildung* $f^{-1} : N \rightarrow M$, die dadurch gekennzeichnet ist, dass $f^{-1}(f(x)) = x$ für alle $x \in M$ und $f(f^{-1}(y)) = y$ für alle $y \in N$. Das heißt, $f^{-1} \circ f$ ist die identische Abbildung auf M , und $f \circ f^{-1}$ die identische Abbildung auf N . (Näheres: siehe Vorlesung "Diskrete Mathematik").

Einfaches Beispiel: Die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto 2x$ ist bijektiv, und die Umkehrabbildung lautet: $f^{-1} : \mathbb{R} \rightarrow \mathbb{R} : y \mapsto \frac{1}{2}y$. (Natürlich kann man bei der Umkehrabbildung auch $x \mapsto \frac{1}{2}x$ schreiben.)

Sei nun A eine quadratische Matrix mit n Zeilen und Spalten. Wenn $\text{rank } A = n$ ist, so hat das Gleichungssystem $Ax = b$ nach Satz 19 für jedes $b \in \mathbb{R}^n$ genau eine Lösung. Wenn $\text{rank } A < n$ ist, so gibt es entweder keine Lösung oder unendlich viele Lösungen (nämlich einen affinen Teilraum der Dimension $n - \text{rank } A$). Wir sehen also, dass die lineare Abbildung f_A genau dann bijektiv ist, wenn der Rang von A gleich n ist. In diesem Fall gibt es also eine Umkehrabbildung. Diese ist wieder linear, wie man folgenderweise sehen kann:

Seien $x', y' \in \mathbb{R}^n$. Wegen der Bijektivität von f_A gibt es eindeutig bestimmte $x, y \in \mathbb{R}^n$, sodass $x' = f_A(x)$ und $y' = f_A(y)$. Aus $f_A(x+y) = f_A(x) + f_A(y) = x' + y'$ folgt $x+y = f_A^{-1}(x'+y')$, und das heißt $f_A^{-1}(x') + f_A^{-1}(y') = f_A^{-1}(x'+y')$, womit die Additivität von f_A^{-1} bewiesen ist. Die Homogenität sieht man ebenso leicht.

Wir überlegen uns nun, wie man die Matrix von f_A^{-1} berechnen kann.

Da f_A^{-1} linear ist, gibt es eine (eindeutig bestimmte) Matrix A' , sodass $f_A^{-1} = f_{A'}$ und somit $AA' = E$. Wir versuchen nun, diese Matrix A' zu finden, d.h. wir suchen Spaltenvektoren $a'_1, \dots, a'_n \in \mathbb{R}^n$, sodass $A(a'_1, \dots, a'_n) = (e_1, \dots, e_n)$, also

$$Aa'_i = e_i \text{ für alle } i \in \{1, \dots, n\}.$$

Wir sehen: für jedes i ist a'_i die eindeutige Lösung des Gleichungssystems $Ax = e_i$.

Wir bezeichnen die Matrix A' mit A^{-1} und nennen sie die zu A *inverse Matrix*. Außerdem haben wir auch gleich ein Verfahren zur Berechnung von A^{-1} gewonnen: Wir lösen die n linearen Gleichungssysteme $Ax = e_i$, und die Lösungsvektoren ergeben die Spalten von A^{-1} . Da diese Gleichungssysteme alle dieselbe Matrix A haben, kann man sie simultan lösen, wie in folgendem Beispiel gezeigt wird.

Beispiel: Sei $A = \begin{pmatrix} 2 & 2 & 3 \\ 3 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

Wir schreiben A und E folgenderweise nebeneinander und wenden dann die

angegebenen elementaren Umformungen an:

$$\begin{array}{ccc|ccc} 2 & 2 & 3 & 1 & 0 & 0 \\ 3 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}$$

$z_1 \leftarrow \frac{1}{2}z_1 :$

$$\begin{array}{ccc|ccc} 1 & 1 & \frac{3}{2} & \frac{1}{2} & 0 & 0 \\ 3 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}$$

$z_2 \leftarrow z_2 - 3z_1 :$

$$\begin{array}{ccc|ccc} 1 & 1 & \frac{3}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{7}{2} & -\frac{3}{2} & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}$$

$z_2 \longleftrightarrow z_3 :$

$$\begin{array}{ccc|ccc} 1 & 1 & \frac{3}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & -\frac{7}{2} & -\frac{3}{2} & 1 & 0 \end{array}$$

$z_3 \leftarrow -\frac{2}{7}z_3 :$

$$\begin{array}{ccc|ccc} 1 & 1 & \frac{3}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{3}{7} & -\frac{2}{7} & 0 \end{array}$$

$z_1 \leftarrow z_1 - \frac{3}{2}z_3, z_2 \leftarrow z_2 - z_3 :$

$$\begin{array}{ccc|ccc} 1 & 1 & 0 & -\frac{1}{7} & \frac{3}{7} & 0 \\ 0 & 1 & 0 & -\frac{3}{7} & \frac{2}{7} & 1 \\ 0 & 0 & 1 & \frac{3}{7} & -\frac{2}{7} & 0 \end{array}$$

$z_1 \leftarrow z_1 - z_2 :$

$$\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{2}{7} & \frac{1}{7} & -1 \\ 0 & 1 & 0 & -\frac{3}{7} & \frac{2}{7} & 1 \\ 0 & 0 & 1 & \frac{3}{7} & -\frac{2}{7} & 0 \end{array}$$

Wir erhalten also $A^{-1} = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} & -1 \\ -\frac{3}{7} & \frac{2}{7} & 1 \\ \frac{3}{7} & -\frac{2}{7} & 0 \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 2 & 1 & -7 \\ -3 & 2 & 7 \\ 3 & -2 & 0 \end{pmatrix}.$

Wie wir gesehen haben, existiert die Inverse einer $n \times n$ -Matrix A genau dann, wenn $\text{rank } A = n$. In diesem Fall nennt man A *regulär* oder *invertierbar*. Andernfalls heißt A *singulär*.

Regeln für die Inversion von Matrizen:

a) $AA^{-1} = A^{-1}A = E$, und daher gilt $(A^{-1})^{-1} = A$.

(Zum Beweis: $f_{A^{-1}} = f_A^{-1}$.)

b) Wenn A und B regulär sind, dann ist auch AB regulär, und es gilt

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

(*Beweis:* $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AEA^{-1} = AA^{-1} = E$. Analog kann man $(B^{-1}A^{-1})(AB) = E$ zeigen. Nach der folgenden Bemerkung ist das aber eigentlich gar nicht nötig.)

Bemerkung 24 Wenn A und B zwei $n \times n$ -Matrizen mit der Eigenschaft $AB = E$ sind, dann sind A und B beide regulär, und es gilt $A^{-1} = B$ und $B^{-1} = A$.

Beweis: Angenommen, A wäre nicht regulär, d.h. $\text{rank } A < n$.

Wegen $f_B(\mathbb{R}^n) \subset \mathbb{R}^n$ wäre dann $\dim f_A(f_B(\mathbb{R}^n)) \leq \dim f_A(\mathbb{R}^n) < n$ (siehe Satz 20).

Andererseits ist aber $f_A(f_B(\mathbb{R}^n)) = f_A \circ f_B(\mathbb{R}^n) = f_E(\mathbb{R}^n) = \mathbb{R}^n$, und das ergibt einen Widerspruch.

Wir können daher die Gleichung $AB = E$ von links mit A^{-1} multiplizieren und erhalten $B = A^{-1}$. Daraus folgt, dass auch B regulär ist, und $B^{-1} = (A^{-1})^{-1} = A$. \square

Wir sehen außerdem:

Bemerkung 25 Für jedes feste $n \in \mathbb{N}$ ist die Menge aller regulären (reellen) $n \times n$ -Matrizen bezüglich der Matrizenmultiplikation eine Gruppe (mit neuem Element E).

Diese Gruppe ist für $n > 1$ nicht kommutativ (siehe Bemerkung 4).

Bemerkung zur Lösung von linearen Gleichungssystemen mit Hilfe der Matrizeninversion:

Wenn A eine reguläre $n \times n$ -Matrix ist, so kann man das Gleichungssystem $Ax = b$ für jedes $b \in \mathbb{R}^n$ einfach dadurch lösen, dass man beide Seiten

von links mit A^{-1} multipliziert. Man erhält: $x = A^{-1}b$. Da die Inversion einer Matrix aber etwas aufwendiger als das Eliminationsverfahren ist, lohnt sich diese Methode nur, wenn man mehrere Gleichungssysteme mit derselben Matrix und verschiedenen rechten Seiten zu lösen hat.

Bemerkung zur Behandlung des Stoffes in der Schule:

Die Inversion von Matrizen gehört (derzeit) nicht zum verpflichtenden Schulstoff. Wegen des engen Zusammenhangs mit linearen Gleichungen bzw. elementaren Umformungen und der Multiplikation von Matrizen stellt sie jedoch einen idealen Erweiterungsstoff dar und wird deshalb auch in gewissen Schulbüchern behandelt (z.B. [2]).

Wenn man aus irgendeinem Grund in der Schule den Begriff "Gruppe" einführt, dann bieten sich die regulären $n \times n$ -Matrizen (mit $n \geq 2$) mit der Matrizenmultiplikation als interessante Beispiele von nichtkommutativen Gruppen an: Es gilt das Assoziativgesetz, das neutrale Element ist die Einheitsmatrix E , und das zu A inverse Element ist die inverse Matrix A^{-1} . Um den Aufwand in Grenzen zu halten, kann man sich dabei auf den Fall $n = 2$ beschränken. In diesem Fall kann man die inverse Matrix sehr einfach berechnen:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Diese Formel kann man z.B. mit Hilfe von Determinanten begründen (siehe Kapitel 9, Satz 49) oder einfach durch direktes Nachrechnen der Beziehung $AA^{-1} = E$ (siehe oben) beweisen.

7 Inneres Produkt und Orthogonalität

7.1 Das innere Produkt und die Norm

Unter dem *inneren* oder *skalaren Produkt* zweier Vektoren $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ versteht man die Zahl

$$x \cdot y := x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_{i=1}^n x_iy_i.$$

Wegen der Schreibweise mit einem Punkt wird dieses Produkt auf englisch oft auch *dot product* genannt, zum Unterschied vom später erklärten "äußereren

Produkt" (siehe Kapitel 10), das mit einem Kreuz (\times) geschrieben wird und daher auch *cross product* heißt.

Häufig wird für das innere Produkt auch folgende Schreibweise verwendet: $\langle x, y \rangle$.

Dieses Produkt dient insbesondere zur Längen- und Winkelberechnung, wie später genauer ausgeführt wird.

Das innere Produkt hängt eng mit dem Matrizenprodukt zusammen:

1.

$$(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = (x_1, \dots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Wenn wir x und y mit den entsprechenden einspaltigen Matrizen identifizieren, dann heißt das:

$$x \cdot y = x^T y.$$

2. Seien a_1^*, \dots, a_s^* die Zeilenvektoren der Matrix A und b_1, \dots, b_n die Spaltenvektoren von B . Dann gilt für die Elemente c_{ik} der Produktmatrix $C = AB$:

$$c_{ik} = a_i^* \cdot b_k.$$

Satz 26 Für beliebige Vektoren $x, x', y, y' \in \mathbb{R}^n$ und Skalare $\lambda, \mu \in \mathbb{R}$ gilt

- a) $(\lambda x + \mu x') \cdot y = \lambda(x \cdot y) + \mu(x' \cdot y),$
- b) $x \cdot (\lambda y + \mu y') = \lambda(x \cdot y) + \mu(x \cdot y'),$
- c) $x \cdot y = y \cdot x \quad (\text{Symmetrie}),$
- d) $x \cdot x > 0 \text{ für } x \neq 0 \quad (\text{positive Definitheit}).$

Der Beweis dieses Satzes ist eine einfache Übungsaufgabe und wird daher hier weggelassen.

Bemerkungen dazu: Die erste Regel drückt aus, dass für jedes feste y die Abbildung

$$f_y : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto x \cdot y$$

linear ist. Analog drückt die zweite Regel die Linearität der Abbildung

$$g_x : \mathbb{R}^n \rightarrow \mathbb{R} : y \mapsto x \cdot y$$

für festes x aus. Allgemein nennt man eine Abbildung $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ mit diesen Eigenschaften eine *Bilinearform*. Das innere Produkt ist also eine positiv definite symmetrische Bilinearform.

Mit $\mu = 0$ ergibt sich insbesondere folgendes "Assoziativgesetz":

$$(\lambda x) \cdot y = \lambda(x \cdot y) = x \cdot (\lambda y),$$

sodass man ohne Zweideutigkeit $\lambda x \cdot y$ schreiben kann. Zur vierten Regel können wir ergänzend anmerken:

$$x \cdot x = 0 \Leftrightarrow x = o.$$

ACHTUNG: Das innere Produkt von drei oder mehr Vektoren hat keinen Sinn! Dagegen sind die Ausdrücke $(x \cdot y)z$ und $x(y \cdot z)$ schon beide sinnvoll, wenn man bei der Multiplikation eines Skalars λ mit einem Vektor x die Schreibweise $x\lambda$ zulässt. Im Allgemeinen ist aber $(x \cdot y)z \neq x(y \cdot z)$: Das eine ist nämlich ein skalares Vielfaches von z , das andere von x .

Beispiel: Sei $x = (4, 0, 0)$, $y = (1, 1, 2)$, $z = (0, 0, 3)$. Dann ist $(x \cdot y)z = 4z = (0, 0, 12)$, aber $x(y \cdot z) = x6 = 6x = (24, 0, 0)$.

Unter der *Länge* oder *Norm* eines Vektors $x \in \mathbb{R}^n$ versteht man die Zahl

$$\|x\| := \sqrt{x \cdot x} = \sqrt{x_1^2 + \dots + x_n^2}.$$

Im \mathbb{R}^2 und \mathbb{R}^3 entspricht das der elementargeometrischen Länge, wie man sich mit Hilfe des bekannten Satzes von Pythagoras leicht überlegen kann.

Der folgende Satz gibt eine grundlegende Eigenschaft des inneren Produkts an:

Satz 27 ("Cauchy-Schwarz'sche Ungleichung") Für beliebige Vektoren $x, y \in \mathbb{R}^n$ gilt

$$|x \cdot y| \leq \|x\| \|y\|$$

mit Gleichheit genau dann, wenn x und y linear abhängig sind.

Beweis:

Für $y = o$ ist die Aussage des Satzes trivial: In diesem Fall sind beide Seiten der Ungleichung gleich Null, und die Vektoren x, y sind linear abhängig.

Sei nun $y \neq o$. Auf Grund der positiven Definitheit gilt

$$0 \leq (x + \lambda y) \cdot (x + \lambda y)$$

für jedes $\lambda \in \mathbb{R}$. Multiplizieren wir die rechte Seite gemäß den Regeln von Satz 26 aus, so ergibt das

$$0 \leq x \cdot x + 2\lambda(x \cdot y) + \lambda^2(y \cdot y).$$

Setzen wir speziell $\lambda = -\frac{x \cdot y}{y \cdot y}$, so erhalten wir

$$0 \leq x \cdot x - 2\frac{x \cdot y}{y \cdot y}(x \cdot y) + \frac{(x \cdot y)^2}{(y \cdot y)^2}(y \cdot y).$$

Das ergibt nach Multiplikation mit $y \cdot y$

$$0 \leq (x \cdot x)(y \cdot y) - (x \cdot y)^2,$$

Addieren wir auf beiden Seiten $(x \cdot y)^2$ und ziehen anschließend die Wurzel, so haben wir schon die behauptete Ungleichung.

Wenn hier Gleichheit eintritt, so gilt (mit dem speziellen λ)

$$0 = (x + \lambda y) \cdot (x + \lambda y)$$

und daher $x + \lambda y = o$. Die Vektoren x, y sind dann also linear abhängig.

Wenn wir umgekehrt annehmen, dass x und y linear abhängig sind, dann gibt es (wegen $y \neq o$) ein μ , sodass $x = \mu y$, und es folgt:

$(x \cdot y)^2 = (\mu y \cdot y)(\mu y \cdot y) = \mu^2(y \cdot y)(y \cdot y) = (\mu y \cdot \mu y)(y \cdot y) = (x \cdot x)(y \cdot y)$, was zu zeigen war. \square

Aus der Cauchy-Schwarz'schen Ungleichung folgt für $x, y \neq o$:

$$-1 \leq \frac{x \cdot y}{\|x\| \|y\|} \leq 1.$$

Unter dem *Winkel* zwischen zwei Vektoren $x, y \in \mathbb{R}^n \setminus \{o\}$ versteht man nun die (eindeutig bestimmte) Zahl $\gamma \in [0, \pi]$ mit

$$\cos \gamma = \frac{x \cdot y}{\|x\| \|y\|}.$$

Auch hier kann man sich überlegen, dass diese Zahl γ im \mathbb{R}^2 und \mathbb{R}^3 mit dem elementargeometrischen Winkel übereinstimmt (natürlich im Bogenmaß). (Siehe Bemerkung zur Behandlung des Stoffes in der Schule am Ende dieses Abschnitts.)

Wir erhalten damit die folgende Formel, die eine anschauliche Interpretation des inneren Produkts ermöglicht:

$$x \cdot y = \|x\| \|y\| \cos \gamma \quad \text{für } x, y \neq o.$$

$x \cdot y$ ist also gleich der Länge von x , multipliziert mit der (mit dem "richtigen" Vorzeichen versehenen) Länge der senkrechten Projektion von y auf x (oder umgekehrt).

Mit Hilfe der Norm kann man auch den *Abstand zweier Punkte* $x, y \in \mathbb{R}^n$ definieren:

$$d(x, y) := \|x - y\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}.$$

Satz 28 Für beliebige Vektoren $x, y \in \mathbb{R}^n$ und Skalare $\lambda \in \mathbb{R}$ gilt

- a) $\|x + y\| \leq \|x\| + \|y\|$ ("Dreiecksungleichung"),
- b) $\|\lambda x\| = |\lambda| \|x\|$,
- c) $\|x\| = 0 \Leftrightarrow x = o$.

Beweis:

a) Da die Norm stets nicht-negativ ist, können wir die Ungleichung in quadratischer Form beweisen.

$$\|x + y\|^2 = (x + y) \cdot (x + y) = x \cdot x + 2(x \cdot y) + y \cdot y = \|x\|^2 + 2(x \cdot y) + \|y\|^2,$$

und das ist nach der Cauchy-Schwarz'schen Ungleichung

$$\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

$$\text{b) } \|\lambda x\|^2 = \lambda x \cdot \lambda x = \lambda^2 x \cdot x = \lambda^2 \|x\|^2.$$

$$\text{c) } \|x\| = 0 \Leftrightarrow x \cdot x = 0 \Leftrightarrow x = o. \quad \square$$

Aus der obigen Dreiecksungleichung folgt mit $x - y$ an Stelle von x und $y - z$ an Stelle von y :

$$\|x - z\| \leq \|x - y\| + \|y - z\|,$$

d.h.

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

Dadurch wird die Bezeichnung "Dreiecksungleichung" besser verständlich, denn die letzte Ungleichung bedeutet ja: In einem Dreieck ist die Länge einer Seite höchstens so groß wie die Summe der Längen der beiden anderen Seiten.

Einen Vektor mit Länge 1 nennt man *normiert*. Man kann jeden Vektor $x \neq o$ *normieren*, indem man ihn mit dem Kehrwert seiner Norm multipliziert: Sei $x' = \frac{1}{\|x\|}x$. Dann ist $\|x'\| = \left| \frac{1}{\|x\|} \right| \|x\| = 1$.

Bemerkung zum Beweis der Cauchy-Schwarz'schen Ungleichung:

Beim Beweis wird die Definition des inneren Produkts im \mathbb{R}^n gar nicht benutzt, sondern nur die Tatsache, dass es sich um eine positiv definite symmetrische Bilinearform handelt. Es gilt daher eine analoge Ungleichung für jede solche Bilinearform in einem beliebigen (auch unendlichdimensionalen) reellen Vektorraum. Z.B. kann man sich leicht überlegen, dass die Menge $\mathcal{C}[a, b]$ der stetigen Funktionen auf dem Intervall $[a, b]$ in natürlicher Weise ein Vektorraum ist, in dem man in folgender Weise ein "inneres Produkt", d.h. eine positiv definite symmetrische Bilinearform definieren kann:

$$f \cdot g := \int_a^b f(x)g(x) dx.$$

Man erhält dann als Spezialfall der allgemeinen Cauchy-Schwarz'schen Ungleichung:

$$\left(\int_a^b f(x)g(x) dx \right)^2 \leq \left(\int_a^b f(x)^2 dx \right) \left(\int_a^b g(x)^2 dx \right).$$

Historische Bemerkung: Die hier so genannte "Cauchy-Schwarz'sche Ungleichung" wird oft auch nur "Schwarz'sche Ungleichung" genannt, nach Hermann Schwarz (1843-1921). Sie war jedoch bereits 1859, 25 Jahre vor Schwarz, von Viktor Bunyakovsky (1804-1889, St. Petersburg) veröffentlicht worden. Schwarz stieß in Zusammenhang mit seinen Studien der Minimalflächen auf die obige Ungleichung für Integrale.

Augustin Louis Cauchy (1789-1857, Paris) ist einer der Begründer der modernen Analysis. Ihm wird die Ungleichung im \mathbb{R}^n zugeschrieben, die ja nach der obigen Bemerkung auch nur einen Spezialfall der allgemeinen Ungleichung darstellt.

Bemerkung zur Behandlung des Stoffes in der Schule:

In der Schule wird man natürlich den Begriff "Winkel" zunächst anschaulich definieren, ebenso den Kosinus eines Winkels. Mit Hilfe des Pythagoräischen Lehrsatzes kann man dann den Kosinussatz der ebenen Trigonometrie herleiten. Die Beziehung zu der obigen Definition eines Winkels ergibt sich dann folgenderweise daraus:

Auf Grund der Definition der Norm und der Rechenregeln für das innere Produkt gilt

$$\|a - b\|^2 = (a - b) \cdot (a - b) = a \cdot a + b \cdot b - 2a \cdot b = \|a\|^2 + \|b\|^2 - 2a \cdot b.$$

Betrachten wir das Dreieck mit den Eckpunkten o, a, b und bezeichnen den "elementargeometrischen" Winkel bei o mit γ , so gilt auf Grund des Kosinussatzes:

$$\|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\| \cdot \|b\| \cos \gamma.$$

Durch Vergleich sehen wir: $a \cdot b = \|a\| \cdot \|b\| \cos \gamma$, also $\cos \gamma = \frac{a \cdot b}{\|a\| \|b\|}$. Das heißt, der elementargeometrische Winkel stimmt mit dem Winkel gemäß obiger Definition überein.

7.2 Orthogonalität

Zwei Vektoren $x, y \in \mathbb{R}^n$ heißen *orthogonal* (zueinander), wenn $x \cdot y = 0$ ist. Wenn x und y vom Nullvektor verschieden sind, so ist das gleichbedeutend damit, dass der Winkel zwischen x und y gleich $\pi/2$ ist. Man sagt dann auch, dass die beiden Vektoren *senkrecht* oder *normal* aufeinander stehen. Bezeichnung: $x \perp y$.

Der folgenden sehr einfache Satz entspricht dem bekannten *Satz von Pythagoras*:

Satz 29 Seien x, y zueinander orthogonale Vektoren des \mathbb{R}^n . Dann gilt:

$$\|x\|^2 + \|y\|^2 = \|x + y\|^2.$$

Beweis: Wegen $x \cdot y = 0$ gilt

$$\|x + y\|^2 = (x + y) \cdot (x + y) = x \cdot x + 2x \cdot y + y \cdot y = x \cdot x + y \cdot y = \|x\|^2 + \|y\|^2.$$

□

7.2.1 Orthogonalsysteme

Eine (nicht-leere) Teilmenge M von $\mathbb{R}^n \setminus \{o\}$ heißt *Orthogonalsystem*, wenn je zwei Vektoren aus M zueinander orthogonal sind. Wenn ein Orthogonalsystem Basis eines linearen Teilraums U ist (es kann natürlich auch $U = \mathbb{R}^n$ sein), so nennt man es eine *Orthogonalbasis* von U .

Beispiele:

1. Sei $x = (x_1, x_2) \in \mathbb{R}^2$. Dann ist der Vektor $x' := (-x_2, x_1)$ orthogonal zu x . Für $x \neq o$ ist daher $\{x, x'\}$ eine Orthogonalbasis des \mathbb{R}^2 .
2. Die kanonische Basis des \mathbb{R}^n ist eine Orthogonalbasis, denn $e_i \cdot e_k = 0$ für $i \neq k$.
3. Die folgenden vier Vektoren bilden eine interessante Orthogonalbasis des \mathbb{R}^4 :

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

Bemerkung: Aus dem "Satz von Pythagoras" folgt durch Induktion, dass für jedes Orthogonalsystem $\{u_1, \dots, u_s\}$ gilt:

$$\|u_1\|^2 + \dots + \|u_s\|^2 = \|u_1 + \dots + u_s\|^2.$$

Satz 30 *Jedes Orthogonalsystem ist linear unabhängig.*

Beweis: Sei $\{u_1, \dots, u_s\}$ ein Orthogonalsystem. Angenommen, $\lambda_1 u_1 + \dots + \lambda_s u_s = o$. Dann ist $(\lambda_1 u_1 + \dots + \lambda_s u_s) \cdot u_k = 0$ für jedes $k \in \{1, \dots, s\}$, d.h. aber

$$0 = \sum_{i=1}^s \lambda_i (u_i \cdot u_k) = \lambda_k (u_k \cdot u_k), \text{ also } \lambda_k = 0 \text{ für alle } k,$$

und daher ist $\{u_1, \dots, u_s\}$ linear unabhängig. \square

Besondere Bedeutung haben Orthogonalsysteme oder -basen, bei denen alle Vektoren normiert sind. Man nennt sie dann *Orthonormalsysteme* bzw. *Orthonormalbasen* (ONB). Für ein Orthonormalsystem $\{u_1, \dots, u_s\}$ gilt also

$$u_i \cdot u_k = \begin{cases} 0 & \text{für } i \neq k, \\ 1 & \text{für } i = k. \end{cases}$$

Man schreibt dafür oft auch

$$u_i \cdot u_k = \delta_{ik},$$

wobei δ_{ik} das sogenannte *Kronecker-Symbol* ist, das eben durch

$$\delta_{ik} := \begin{cases} 0 & \text{für } i \neq k, \\ 1 & \text{für } i = k. \end{cases}$$

definiert ist. Wir können auch sagen: δ_{ik} ist das Element in der i -ten Zeile und k -ten Spalte der Einheitsmatrix.

Insbesondere ist also die kanonische Basis des \mathbb{R}^n eine Orthonormalbasis.

Satz 31 Seien ξ_1, \dots, ξ_s die Koordinaten von $x \in U$ bezüglich einer (geordneten) Orthonormalbasis (u_1, \dots, u_s) von U , und analog η_1, \dots, η_s die Koordinaten von $y \in U$. Dann gilt:

- a) $\xi_k = x \cdot u_k$ für alle $k \in \{1, \dots, s\}$,
- b) $x \cdot y = \xi_1 \eta_1 + \dots + \xi_s \eta_s$,
- c) $\|x\| = \sqrt{\xi_1^2 + \dots + \xi_s^2}$.

Beweis:

a) Sei $x = \xi_1 u_1 + \dots + \xi_s u_s = \sum_{i=1}^s \xi_i u_i$. Dann gilt auf Grund der Rechenregeln für das innere Produkt:

$$x \cdot u_k = \left(\sum_{i=1}^s \xi_i u_i \right) \cdot u_k = \sum_{i=1}^s \xi_i (u_i \cdot u_k) = \sum_{i=1}^s \xi_i \delta_{ik} = \xi_k.$$

b)

$$\begin{aligned} x \cdot y &= \left(\sum_{i=1}^s \xi_i u_i \right) \cdot \left(\sum_{k=1}^s \eta_k u_k \right) = \sum_{i,k} \xi_i \eta_k (u_i \cdot u_k) = \\ &= \sum_{i,k} \xi_i \eta_k \delta_{ik} = \sum_{i=1}^s \xi_i \eta_i. \end{aligned}$$

c) $\|x\|^2 = x \cdot x$, und das ist nach Teil b) gleich $\xi_1^2 + \dots + \xi_s^2$. \square

Bemerkung: Nach a) gilt (für $x \in U$):

$$x = \sum_{i=1}^s \xi_i u_i = \sum_{i=1}^s (x \cdot u_i) u_i.$$

Beispiel:

Sei U die Ebene mit Gleichung $x_2 + x_3 = 0$ im \mathbb{R}^3 . Dann bilden die Vektoren $u_1 = (1, 0, 0)$ und $u_2 = \frac{1}{\sqrt{2}}(0, 1, -1)$ eine Orthonormalbasis B von U .

Sei $x = (2, 3, -3)$. Offensichtlich ist $x \in U$, und es gilt $x = 2u_1 + 3\sqrt{2}u_2$, d.h. die Koordinaten von x bezüglich B sind $\xi_1 = 2$ und $\xi_2 = 3\sqrt{2}$.

$$x \cdot u_1 = (2, 3, -3) \cdot (1, 0, 0) = 2, \text{ und}$$

$$x \cdot u_2 = (2, 3, -3) \cdot \frac{1}{\sqrt{2}}(0, 1, -1) = \frac{1}{\sqrt{2}}(3 + 3) = \frac{6}{\sqrt{2}} = 3\sqrt{2}.$$

$$\|x\|^2 = 2^2 + 3^2 + 3^2 = 22,$$

$$\xi_1^2 + \xi_2^2 = 2^2 + (3\sqrt{2})^2 = 4 + 18 = 22.$$

7.2.2 Orthogonale Projektion

Sei U ein echter linearer Teilraum des \mathbb{R}^n , und $\{u_1, \dots, u_s\}$ eine Orthonormalbasis von U . Dann können wir jeden Punkt $x \in \mathbb{R}^n$ durch folgende Abbildung auf U "projizieren":

$$\pi_U : \mathbb{R}^n \rightarrow U : x \mapsto \sum_{i=1}^s (x \cdot u_i) u_i.$$

Diese Abbildung hat folgende interessante Eigenschaften:

- 1) π_U ist eine lineare Abbildung.
- 2) Die Punkte von U (und nur diese) werden auf sich selbst abgebildet, d.h. $\pi_U(x) = x$ für alle $x \in U$.
- 3) Für alle $x \in \mathbb{R}^n$ ist $x - \pi_U(x)$ orthogonal zu jedem Vektor $u \in U$.

Beweis:

- 1) Das folgt aus der Linearität des inneren Produkts bezüglich des ersten Faktors.
- 2) Das ergibt sich aus der Bemerkung nach dem letzten Satz.
- 3) Sei $y = \pi_U(x)$. Es genügt natürlich, zu zeigen, dass $x - y$ zu jedem Basisvektor von U orthogonal ist.

$$\begin{aligned} (x - y) \cdot u_k &= (x - \sum_{i=1}^s (x \cdot u_i) u_i) \cdot u_k = x \cdot u_k - \sum_{i=1}^s (x \cdot u_i) (u_i \cdot u_k) = \\ &= x \cdot u_k - \sum_{i=1}^s (x \cdot u_i) \delta_{ik} = x \cdot u_k - x \cdot u_k = 0. \quad \square \end{aligned}$$

Auf Grund dieser Eigenschaften heißt die Abbildung π_U die *orthogonale Projektion* oder *Normalprojektion* auf U , und $\pi_U(x)$ nennt man manchmal den *Lotfußpunkt* von x auf U .

Wichtige Spezialfälle:

1. Sei U eine Gerade durch den Ursprung, $U = \text{lin}(u)$ mit $u \neq o$. Hier ist $\{u'\}$ mit $u' = \frac{1}{\|u\|}u$ eine Orthonormalbasis von U , und $\pi_U(x) = (x \cdot u')u' = \frac{1}{\|u\|}(x \cdot u)u'$, also $\|\pi_U(x)\| = \frac{1}{\|u\|}|x \cdot u|$, und es folgt: $|x \cdot u| = \|u\|\|\pi_U(x)\|$. Das entspricht der anschaulichen Interpretation des inneren Produkts.
2. Sei U eine Hyperebene durch den Ursprung. In diesem Fall kann man sich das so vorstellen: Man legt durch x eine Gerade, die senkrecht auf U steht, und schneidet sie mit U . Der Schnittpunkt ist dann $\pi_U(x)$.

Bemerkung:

Für die Norm von $\pi_U(x)$ ergibt sich nach Satz 31, c):

$$\|\pi_U(x)\|^2 = \sum_{i=1}^s |x \cdot u_i|^2.$$

7.2.3 Das Schmidt'sche Orthonormalisierungsverfahren

Der Beweis des nächsten Satzes ist ein typisches Beispiel eines *konstruktiven Existenzbeweises*: Die Existenz einer Orthonormalbasis wird durch Angabe und Analyse eines entsprechenden Algorithmus bewiesen.

Satz 32 *Jeder lineare Teilraum $U \neq \{o\}$ des \mathbb{R}^n besitzt eine Orthonormalbasis.*

Beweis (Schmidt'sches Orthonormalisierungsverfahren): Sei $\{a_1, \dots, a_s\}$ irgendeine Basis von U . (Wir wissen bereits, dass es immer eine Basis gibt: siehe Satz 9). Wir konstruieren schrittweise Vektoren u_1, \dots, u_s , sodass für jedes $k \in \{1, \dots, s\}$ die Menge $\{u_1, \dots, u_k\}$ eine ONB von $U_k := \text{lin}(a_1, \dots, a_k)$ ist.

Wir beginnen, indem wir a_1 normieren:

$$u_1 := \frac{1}{\|a_1\|}a_1.$$

Angenommen, für ein $k < s$ sind die Vektoren u_1, \dots, u_k bereits so konstruiert, dass sie eine ONB von U_k bilden. Dann setzen wir

$$u'_{k+1} := a_{k+1} - \pi_{U_k}(a_{k+1}) = a_{k+1} - \sum_{i=1}^k (a_{k+1} \cdot u_i)u_i$$

und

$$u_{k+1} := \frac{1}{\|u'_{k+1}\|} u'_{k+1}.$$

Auf Grund der Eigenschaften einer orthogonalen Projektion ist dann u_{k+1} orthogonal zu u_1, \dots, u_k , und daher ist $\{u_1, \dots, u_k, u_{k+1}\}$ ein Orthonormalsystem in U_{k+1} . Da jedes Orthonormalsystem linear unabhängig ist und $\dim U_{k+1} = k + 1$, folgt, dass $\{u_1, \dots, u_k, u_{k+1}\}$ sogar eine ONB von U_{k+1} ist. \square

Beispiel:

$$\text{Sei } U = \text{lin}(a_1, a_2) \text{ mit } a_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ und } a_2 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

$$\|a_1\| = \sqrt{14}, \text{ also } u_1 = \frac{1}{\sqrt{14}} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

$$\begin{aligned} u'_2 &= a_2 - (a_2 \cdot u_1)u_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{14} \left(\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right) \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \\ &= \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} - \frac{7}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \text{ also } u_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}. \end{aligned}$$

Damit können wir nun auch die orthogonale Projektion auf U leicht berechnen: Sei z.B. $x = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$. Dann ist $\pi_U(x) = (x \cdot u_1)u_1 + (x \cdot u_2)u_2 = \frac{4}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \frac{6}{10} \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{35} \begin{pmatrix} 73 \\ 20 \\ 9 \end{pmatrix}$.

Historische Bemerkung: Dieses Verfahren ist nach Erhard Schmidt benannt (1876 - 1959), manchmal wird es auch *Gram-Schmidt-Orthonormalisierung* genannt, nach Jorgen Pedersen Gram (1850- 1916, Dänemark). Es war jedoch bereits vorher von Pierre-Simon Laplace (1749 - 1827) angegeben worden [8].

Satz 33 *Jedes Orthonormalsystem in einem linearen Teilraum U des \mathbb{R}^n lässt sich zu einer Orthonormalbasis von U ergänzen.*

Beweis: Sei $\dim U = s$, und $\{u_1, \dots, u_r\}$ ein in U enthaltenes Orthonormalsystem. Nach dem Basisergänzungssatz (Satz 12) gibt es Vektoren a_{r+1}, \dots, a_s , sodass $\{u_1, \dots, u_r, a_{r+1}, \dots, a_s\}$ eine Basis von U ist. Wenden wir nun das Schmidt'sche Orthonormalisierungsverfahren darauf an, so bleiben u_1, \dots, u_r unverändert, und wir erhalten eine ONB, die $\{u_1, \dots, u_r\}$ enthält. \square

Wenden wir diesen Satz für den \mathbb{R}^n an Stelle von U an, so sehen wir: Man kann jede Orthonormalbasis $\{u_1, \dots, u_s\}$ eines Teilraums U des \mathbb{R}^n zu einer Orthonormalbasis $B = \{u_1, \dots, u_s, u_{s+1}, \dots, u_n\}$ des ganzen \mathbb{R}^n erweitern. Die lineare Hülle von $\{u_{s+1}, \dots, u_n\}$ nennt man dann das *orthogonale Komplement* von U . Bezeichnung: U^\perp .

Einige Eigenschaften von U^\perp :

1. $U^\perp = \{x \in \mathbb{R}^n : x \cdot u = 0 \text{ für alle } u \in U\}$.

(Daraus folgt, dass U^\perp eindeutig definiert ist. Oft wird überhaupt diese Eigenschaft zur Definition von U^\perp verwendet.)

2. $\dim U^\perp = n - \dim U$.

3. $(U^\perp)^\perp = U$.

4. $\pi_U(x) + \pi_{U^\perp}(x) = x$.

Beweis:

1. $u \in U, x \in U^\perp \Rightarrow u = \lambda_1 u_1 + \dots + \lambda_s u_s, x = \lambda_{s+1} u_{s+1} + \dots + \lambda_n u_n$ für gewisse $\lambda_i \in \mathbb{R}$.

Wegen $u_i \cdot u_k = 0$ für $i \neq k$ folgt sofort $x \cdot u = 0$.

Wenn umgekehrt $x \cdot u = 0$ für alle $u \in U$, dann folgt insbesondere $x \cdot u_i = 0$ für alle $i \in \{1, \dots, s\}$. Das heißt nach Satz 31, dass die ersten s Koordinaten von x bezüglich der Basis B gleich Null sind. Also ist x eine Linearkombination von u_{s+1}, \dots, u_n , d.h. $x \in U^\perp$.

2. und 3. sind trivial.

4. Nach Definition der orthogonalen Projektion gilt

$$\pi_U(x) = \sum_{i=1}^s (x \cdot u_i) u_i, \quad \pi_{U^\perp}(x) = \sum_{i=s+1}^n (x \cdot u_i) u_i,$$

und auf Grund der Bemerkung nach Satz 31 (mit n statt s) ist

$$x = \sum_{i=1}^n (x \cdot u_i) u_i. \quad \square$$

Die 4. Eigenschaft kann man benützen, um die Berechnung der orthogonalen Projektion auf U zu vereinfachen, wenn $\dim U > n/2$ ist, denn

$$\pi_U(x) = x - \pi_{U^\perp}(x) = x - \sum_{i=s+1}^n (x \cdot u_i) u_i,$$

und in dieser Form ist der Berechnungsaufwand für $s > n/2$ kleiner als nach der Definition von $\pi_U(x)$. Das hat natürlich nur einen Sinn, wenn man schon eine Orthonormalbasis von U^\perp kennt.

Beispiel:

Wenn wir die ONB von U aus dem letzten Beispiel zu einer ONB des \mathbb{R}^3 ergänzen wollen, so können wir den dortigen Vektor $x = (2, 1, 0)$ als dritten Basisvektor a_3 nehmen, denn die Vektoren u_1, u_2, x sind linear unabhängig, da $x \notin U$ ist (sonst müsste ja $x = \pi_U(x)$ sein). Wir erhalten dann:

$$u'_3 = x - \pi_U(x) = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{35} \begin{pmatrix} 73 \\ 20 \\ 9 \end{pmatrix} = \frac{1}{35} \begin{pmatrix} -3 \\ 15 \\ -9 \end{pmatrix} = \frac{3}{35} \begin{pmatrix} -1 \\ 5 \\ -3 \end{pmatrix},$$

somit $u_3 = \frac{1}{\sqrt{35}} \begin{pmatrix} -1 \\ 5 \\ -3 \end{pmatrix}$.

$\pi_U(x)$ kann man daher auch so berechnen:

$$\pi_U(x) = x - (x \cdot u_3) u_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \frac{3}{35} \begin{pmatrix} -1 \\ 5 \\ -3 \end{pmatrix} = \frac{1}{35} \begin{pmatrix} 73 \\ 20 \\ 9 \end{pmatrix}.$$

7.2.4 Normalvektoren von Hyperebenen

Wenn U eine Hyperebene durch den Ursprung ist, so kann U durch eine lineare Gleichung der Form $ax = 0$ beschrieben werden, wobei $a = (a_1, \dots, a_n)$ eine einzeilige Matrix ist. Identifizieren wir diese Matrix mit dem entsprechenden Vektor $\in \mathbb{R}^n$, so können wir statt ax auch $a \cdot x$ schreiben, und sehen:

$$U = \{x \in \mathbb{R}^n : x \perp a\}.$$

Der Vektor a heißt deshalb *Normalvektor* von U , und wir sehen $U^\perp = \text{lin}(a)$. Wenn a normiert ist, d.h. $\|a\| = 1$, kann die orthogonale Projektion auf U daher folgenderweise berechnet werden:

$$\pi_U(x) = x - (x \cdot a)a.$$

Sei nun H eine beliebige Hyperebene mit Gleichung $ax = b$ bzw. $a \cdot x = b$. Wenn wir eine (partikuläre) Lösung dieser Gleichung kennen, sagen wir p , so ist $H = p + U$, wobei U die Hyperebene mit der Gleichung $a \cdot x = 0$ ist. Wir nennen daher auch in diesem Fall a einen Normalvektor der Hyperebene H .

7.2.5 Spiegelung an einem linearen oder affinen Teilraum

Sei U ein linearer Teilraum des \mathbb{R}^n . Die *Spiegelung* an U wird folgenderweise definiert:

$$\sigma_U : \mathbb{R}^n \rightarrow \mathbb{R}^n : x \mapsto 2\pi_U(x) - x.$$

Es gilt also $\sigma_U(x) - \pi_U(x) = -(x - \pi_U(x))$, und daraus ist die anschauliche Bedeutung leicht zu erkennen.

Für den Spezialfall, dass U eine Hyperebene ist, können wir das Ergebnis des vorigen Abschnitts verwenden und erhalten:

$$\sigma_U(x) = 2(x - (x \cdot a)a) - x = x - 2(x \cdot a)a,$$

wobei a wieder ein normierter Normalvektor der Hyperebene ist.

Wenn L ein affiner Teilraum ist, sagen wir $L = p + U$, so führen wir zuerst die Translation τ_{-p} durch, welche L in U überführt, dann spiegeln wir an U , und schließlich machen wir die Translation wieder rückgängig. Die Spiegelung an L wird daher so definiert:

$$\sigma_L := \tau_p \circ \sigma_U \circ \tau_{-p},$$

d.h.

$$\sigma_L(x) = \sigma_U(x - p) + p.$$

Beispiel:

Sei H die Ebene im \mathbb{R}^3 mit Gleichung $x_1 + 2x_2 + x_3 = 4$. Wir finden sehr leicht einen Punkt p auf dieser Ebene, nämlich $p = (1, 1, 1)$. Der normierte Normalvektor ist $a = \frac{1}{\sqrt{6}}(1, 2, 1)$. Die Spiegelung an H sieht daher so aus:

$$\sigma_H(x) = (x - p) - 2((x - p) \cdot a)a + p = x - 2((x - p) \cdot a)a.$$

Konkret erhalten wir etwa für $x = (1, 2, 3)$ den Punkt $\sigma_H(1, 2, 3) = (1, 2, 3) - \frac{8}{6}(1, 2, 1) = \frac{1}{3}(-1, -2, 5)$.

7.2.6 Abstand eines Punktes von einem linearen oder affinen Teilraum

Satz 34 Sei U ein (nicht leerer) linearer Teilraum des \mathbb{R}^n und x ein beliebiger Punkt des \mathbb{R}^n . Dann gilt für alle $u \in U$:

$$\|x - u\| \geq \|x - \pi_U(x)\|,$$

und das Gleichheitszeichen gilt nur für $u = \pi_U(x)$.

Beweis: $x - \pi_U(x)$ ist orthogonal zu $\pi_U(x) - u$, und daher folgt aus dem Satz von Pythagoras:

$$\|x - u\|^2 = \|x - \pi_U(x)\|^2 + \|\pi_U(x) - u\|^2 \geq \|x - \pi_U(x)\|^2. \quad \square$$

$\|\pi_U(x)\|$ ist also der kleinste Abstand eines Punktes $u \in U$ von dem Punkt x . Diese Zahl nennt man daher auch den *Abstand des Punktes x von dem Teilraum U* . Bezeichnung: $d(x, U)$.

Wenn $L = p + U$ ein affiner Teilraum ist, so sehen wir:

$$\|x - (p + u)\| = \|(x - p) - u\| \geq \|(x - p) - \pi_U(x - p)\| = \|x - (p + \pi_U(x - p))\|,$$

d.h. für jeden Punkt $y \in L$ gilt:

$$\|x - y\| \geq \|x - (p + \pi_U(x - p))\|.$$

Der Punkt $p + \pi_U(x - p)$ heißt auch hier *Lotfußpunkt* von x auf L , und der Abstand von x zu seinem Lotfußpunkt heißt der *Abstand des Punktes x von L* und wird mit $d(x, L)$ bezeichnet. Es gilt also:

$$d(x, L) = \|x - p - \pi_U(x - p)\|.$$

Besonders einfach ist die Situation, wenn es sich um eine Hyperebene H handelt: Sei $ax = b$ die Gleichung von H und p ein Punkt auf H , also $ap = b$. Dann ist $H = p + U$, wobei U die Gleichung $ax = 0$ hat.

Wenn $\|a\| = 1$ ist, dann gilt also:

$$\begin{aligned} d(x, H) &= \|x - p - \pi_U(x - p)\| = \|x - p - ((x - p) - ((x - p) \cdot a)a)\| = \\ &= \|((x - p) \cdot a)a\| = |(x - p) \cdot a| = |a(x - p)| = |ax - ap| = |ax - b|, \text{ also} \end{aligned}$$

$$d(x, H) = |a \cdot x - b|.$$

Wenn $ax = b$ die Gleichung einer Hyperebene mit $\|a\| = 1$ ist, so erhält man den Abstand eines Punktes von H also, indem man den Punkt in die Gleichung der Hyperebene einsetzt und dann den absoluten Betrag der Differenz zwischen linker und rechter Seite bildet.

Bemerkung: Eine Hyperebenengleichung $ax = b$ mit normiertem Normalvektor a nennt man *Hesse'sche Normalform* (nach Ludwig Otto Hesse 1811-1874; Königsberg, Heidelberg, München).

Beispiele:

a) Sei G die Gerade im \mathbb{R}^3 mit Parameterdarstellung $x = (1, 1, 1) + \lambda(2, 0, -1)$.

Wir wollen den Abstand des Punktes $q = (1, 2, 3)$ von G berechnen:

Hier ist $p = (1, 1, 1)$, $u = \frac{1}{\sqrt{5}}(2, 0, -1)$ und wir haben:

$$d(q, G) = \|q - p - ((q - p) \cdot u)u\| = \left\| (0, 1, 2) + \frac{2}{5}(2, 0, -1) \right\| = \frac{1}{5} \|(4, 5, 8)\| = \frac{1}{5}\sqrt{105}.$$

b) Sei H die Ebene im \mathbb{R}^3 mit Gleichung $x_1 + 2x_2 - x_3 = 4$. Den Abstand des Punktes $q = (1, 2, 3)$ von H erhalten wir dann so:

$$d(q, H) = \frac{1}{\sqrt{6}} |(1, 2, -1) \cdot (1, 2, 3) - 4| = \frac{2}{\sqrt{6}}.$$

7.2.7 Orthogonale Matrizen

Eine $n \times n$ -Matrix heißt *orthogonal*, wenn ihre Spaltenvektoren ein Orthonormalsystem bilden. Wegen der linearen Unabhängigkeit bilden Sie dann natürlich eine Orthonormalbasis des \mathbb{R}^n .

Wenn A eine orthogonale Matrix ist, so nennt man f_A eine *orthogonale Abbildung*.

Bemerkung: Für orthogonale Matrizen und Abbildungen sollte man eigentlich besser das Adjektiv "orthonormal" verwenden, allerdings hat sich das bis jetzt nicht durchgesetzt.

ACHTUNG: Die orthogonale Projektion π_U (aufgefasst als Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^n$) ist keine orthogonale Abbildung, abgesehen vom trivialen Fall $U = \mathbb{R}^n$: Nach dem folgenden Satz ist jede orthogonale Abbildung bijektiv, wogegen π_U für $U \neq \mathbb{R}^n$ weder surjektiv noch injektiv ist.

Satz 35 Eine quadratische Matrix A ist genau dann orthogonal, wenn sie regulär ist und $A^{-1} = A^T$ gilt.

(Die Inverse einer orthogonalen Matrix kann man also sehr leicht berechnen.)

Beweis:

- Seien a_1, \dots, a_n die Spaltenvektoren einer orthogonalen $n \times n$ -Matrix A , d.h.

$$a_i \cdot a_k = \delta_{ik}.$$

Da die Spaltenvektoren von A gleich den Zeilenvektoren von A^T sind, sehen wir schon: $A^T A = E$, und daraus folgt $A^T = A^{-1}$ (siehe Bemerkung 24).

- Sei umgekehrt $A^{-1} = A^T$, also $A^T A = E$. Das heißt aber nichts anderes als $a_i \cdot a_k = \delta_{ik}$, und das bedeutet die Orthogonalität von A . \square

Es folgen nun einige weitere Charakterisierungen von orthogonalen Matrizen:

Satz 36 *Eine quadratische Matrix ist genau dann orthogonal, wenn ihre Zeilenvektoren ein Orthonormalsystem bilden.*

Beweis: Seien a_1^*, \dots, a_n^* die Zeilenvektoren von A . Die Orthogonalität der Zeilenvektoren bedeutet $a_i^* \cdot a_k^* = \delta_{ik}$, das heißt $A A^T = E$, und das ist wiederum gleichbedeutend mit $A^T = A^{-1}$. \square

Satz 37 *Eine $n \times n$ -Matrix A ist genau dann orthogonal, wenn für alle $x, y \in \mathbb{R}^n$ gilt:*

$$Ax \cdot Ay = x \cdot y.$$

Beweis:

- Sei A orthogonal, also $A^T A = E$. Wenn wir die Vektoren x, y sowie Ax, Ay mit den entsprechenden einspaltigen Matrizen identifizieren, so gilt

$$Ax \cdot Ay = (Ax)^T (Ay) = (x^T A^T)(Ay) = x^T (A^T A)y = x^T E y = x^T y = x \cdot y.$$

- Sei umgekehrt $Ax \cdot Ay = x \cdot y$ für alle $x, y \in \mathbb{R}^n$. Daraus folgt mit $x = e_i$ und $y = e_k$: $a_i \cdot a_k = A e_i \cdot A e_k = e_i \cdot e_k = \delta_{ik}$, also ist A orthogonal. \square

Der letzte Satz besagt, dass eine Matrix genau dann orthogonal ist, wenn die zugehörige lineare Abbildung f_A das innere Produkt unverändert ("*invariant*") lässt. Insbesondere folgt daraus, dass eine orthogonale Abbildung Abstände und Winkel unverändert lässt. Es gilt sogar:

Satz 38 *Eine $n \times n$ -Matrix A ist genau dann orthogonal, wenn für alle $x \in \mathbb{R}^n$ gilt:*

$$\|Ax\| = \|x\|.$$

Wir haben uns nur mehr eine Richtung zu überlegen:

Voraussetzung: $\|Ax\| = \|x\|$ für alle $x \in \mathbb{R}^n$.

Behauptung: A ist orthogonal.

Zum *Beweis* zeigen wir, dass sich das innere Produkt folgendermaßen durch Normen ausdrücken lässt und daher aus der *Invarianz der Norm* die *Invarianz des inneren Produkts* folgt:

$$x \cdot y = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

Wir formen das Vierfache der rechten Seite folgenderweise um:

$$(x+y) \cdot (x+y) - (x-y) \cdot (x-y) = (x \cdot x + 2x \cdot y + y \cdot y) - (x \cdot x - 2x \cdot y + y \cdot y) = 4x \cdot y,$$

und damit sind wir schon fertig. \square

Beispiele:

1. Für jede reelle Zahl γ ist die Matrix

$$D_\gamma := \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix}$$

orthogonal. (Das folgt aus der bekannten Beziehung $\cos^2 \gamma + \sin^2 \gamma = 1$.)

Die zugehörige lineare Abbildung beschreibt im \mathbb{R}^2 eine *Drehung* (um den Ursprung) um den Winkel γ . Das sieht man z.B., wenn man die Bilder der Basisvektoren e_1, e_2 ansieht (das sind die beiden Spaltenvektoren der Matrix D_γ).

2. Ebenso sieht man leicht, dass die Matrix

$$D_\gamma^1 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{pmatrix}$$

orthogonal ist. Sie beschreibt im \mathbb{R}^3 eine Drehung um die 1. Koordinatenachse $\text{lin}(e_1)$ um den Winkel γ : Der Vektor e_1 bleibt unverändert, und die Einschränkung auf die Ebene $\text{lin}(e_2, e_3)$ ist eine ebene Drehung.

(Wie man Drehungen um beliebige Achsen darstellen kann, wird unter anderem im Abschnitt 8.2 erläutert.)

Die Zusammensetzung einer orthogonalen Abbildung mit einer Translation nennt man *Bewegung* (auch: *Kongruenztransformation*). Eine Bewegung des \mathbb{R}^n ist also eine Abbildung

$$\beta : \mathbb{R}^n \rightarrow \mathbb{R}^n : Ax + b$$

mit einer orthogonalen Matrix A . Damit können wir jetzt auch den Begriff "kongruent" im Rahmen unserer Theorie definieren: Zwei Teilmengen M und N des \mathbb{R}^n heißen *kongruent*, wenn es eine Bewegung β gibt, sodass $\beta(M) = N$ ist. Es ist leicht einzusehen, dass auf diese Weise eine Äquivalenzrelation auf der Menge aller Teilmengen des \mathbb{R}^n definiert wird.

ACHTUNG: Eine "kontinuierliche" Bewegung, wie sie z.B. in der Physik betrachtet wird, ist etwas anderes, hängt aber mit dem hier definierten Begriff folgenderweise zusammen: Bei einer kontinuierlichen Bewegung wird jedem Zeitpunkt t aus einem gewissen Intervall eine Bewegung $\beta(t)$ in obigem Sinn zugeordnet (für $n = 3$). Die Lage eines Körpers $K \subset \mathbb{R}^3$ zur Zeit t wird dann durch $\beta(t)(K)$ beschrieben.

Da sich Abstände und Winkel bei einer Translation nicht ändern, bleiben sie auch bei beliebigen Bewegungen unverändert.

8 Koordinatentransformationen

8.1 Umrechnung der Koordinaten eines Punktes oder Vektors

Sei $B = (b_1, \dots, b_n)$ eine (geordnete) Basis des \mathbb{R}^n . Wie berechnet man die Koordinaten eines Punktes $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ bezüglich dieser Basis?

Die Koordinaten von x bezüglich B sind die eindeutig bestimmten Zahlen x'_1, \dots, x'_n , sodass

$$x = x'_1 b_1 + \dots + x'_n b_n.$$

Die Koordinaten x'_1, \dots, x'_n bilden zusammen ein n -Tupel $x' = (x'_1, \dots, x'_n) \in \mathbb{R}^n$, das wir den *Koordinatenvektor* von x bezüglich B nennen. Fassen wir die b_k als Spaltenvektoren auf, so ist B eine Matrix, und es gilt (vgl. Abschnitt 5.2.2):

$$Bx' = x'_1 b_1 + \dots + x'_n b_n = x.$$

Damit könnten wir aus den Koordinaten bezüglich B die ursprünglichen Koordinaten (d.h. bezüglich der kanonischen Basis) berechnen. Wir haben uns

aber gerade die umgekehrte Frage gestellt. Diese ist aber jetzt ganz leicht zu beantworten: Da die n Spalten von B linear unabhängig sind, ist B regulär, und es folgt:

$$x' = B^{-1}x.$$

Beispiel:

Sei $b_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ und $b_2 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, also $B = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$.

$B^{-1} = \begin{pmatrix} -5 & 2 \\ 3 & -1 \end{pmatrix}$, und daher erhalten wir folgende Formeln für die Koordinaten x'_i bezüglich der Basis B :

$$\begin{aligned} x'_1 &= -5x_1 + 2x_2 \\ x'_2 &= 3x_1 - x_2 \end{aligned}$$

Betrachten wir konkret etwa den Vektor $x = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$. Wir erhalten

$$\begin{aligned} x'_1 &= -5 \cdot 3 + 2 \cdot 2 = -11 \\ x'_2 &= 3 \cdot 3 - 2 = 7 \end{aligned}$$

Zur Probe können wir $x'_1 b_1 + x'_2 b_2$ berechnen und erhalten wieder die ursprünglichen Koordinaten:

$$-11 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + 7 \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Wenn B eine ONB ist, ist die Umrechnung besonders einfach, weil ja $B^{-1} = B^T$ ist.

Beispiel:

Sei $B = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{2}} & \frac{\sqrt{6}}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{pmatrix}$. Die Orthogonalität von B kann man leicht (z.B.

auf Grund der Definition) überprüfen. Die Umrechnungsformel lautet $x' = B^T x$. Konkret für $x = (1, 2, 3)$ ergibt das zum Beispiel:

$$x' = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{2}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{6}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} \\ \frac{-3}{\sqrt{6}} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 4\sqrt{3} \\ -\sqrt{2} \\ -\sqrt{6} \end{pmatrix}.$$

1. *Bemerkung:* Wir können die Koordinaten von x bezüglich einer Orthonormalbasis B natürlich auch gemäß Satz 31 berechnen, d.h. $x'_i = x \cdot b_i$. Im Grunde genommen ist das dasselbe, denn die i -te Koordinate von $B^T x$ erhalten wir durch Multiplikation der i -ten Zeile von B^T mit x , und das ergibt $b_i \cdot x$.

2. *Bemerkung:* Wenn x' und y' die Koordinatenvektoren von x und y bezüglich einer Basis B sind, dann ist natürlich $\lambda x' + \mu y'$ der Koordinatenvektor von $\lambda x + \mu y$ bezüglich B (für $\lambda, \mu \in \mathbb{R}$), denn $B^{-1}(\lambda x + \mu y) = \lambda B^{-1}x + \mu B^{-1}y = \lambda x' + \mu y'$.

Wir überlegen uns jetzt noch, was passiert, wenn man den Ursprung des Koordinatensystems in einen beliebigen Punkt c verlegen will. Für den Ortsvektor x' eines Punktes x in Bezug auf den neuen Ursprung c gilt natürlich:

$$x' = \vec{cx} = x - c$$

oder:

$$x = x' + c.$$

Wenn man außerdem noch zu einer anderen Basis B übergeht, so ergibt sich die folgende Transformationsformel:

$$x = Bx' + c.$$

Die umgekehrte Transformation ist

$$x' = B^{-1}(x - c).$$

8.2 Umrechnung der Matrix einer linearen Abbildung

Wir behandeln hier der Einfachheit halber nur lineare Abbildungen $\mathbb{R}^n \rightarrow \mathbb{R}^n$. Solche Abbildungen nennt man auch *Endomorphismen* des \mathbb{R}^n .

8.2.1 Matrix eines Endomorphismus in Bezug auf eine Basis

Sei f ein Endomorphismus und $B = (b_1, \dots, b_n)$ irgendeine (geordnete) Basis des \mathbb{R}^n . Unter der *Matrix* von f bezüglich der Basis B verstehen wir die $n \times n$ -Matrix A' , in deren k -ter Spalte die Koordinaten von $f(b_k)$ bezüglich B stehen (für $k = 1, \dots, n$). (Auf Grund dieser Definition ist also jede $n \times n$ -Matrix A die Matrix der Abbildung f_A bezüglich der kanonischen Basis des \mathbb{R}^n .)

Wir wollen uns überlegen, wie man die Matrix A' von f_A bezüglich B berechnen kann.

Sei x ein beliebiges Element des \mathbb{R}^n und $x' = (x'_1, \dots, x'_n)$ der Koordinatenvektor von x bezüglich der Basis B . Dann gilt (mit $f := f_A$)

$$y := f(x) = f(x'_1 b_1 + \dots + x'_n b_n) = x'_1 f(b_1) + \dots + x'_n f(b_n),$$

und daher gilt für den Koordinatenvektor $y' := (y'_1, \dots, y'_n)$ von y bezüglich B : $y' = x'_1 a'_1 + \dots + x'_n a'_n$, d.h.

$$y' = A' x'.$$

Wegen $x' = B^{-1}x$ und $y' = B^{-1}y$ heißt das $B^{-1}y = A'B^{-1}x$, also $y = BA'B^{-1}x$. Daraus erkennen wir

$$A = BA'B^{-1},$$

und das ist gleichbedeutend mit

$$A' = B^{-1}AB.$$

Bemerkung: Wenn es zu zwei $n \times n$ -Matrizen A, A' eine reguläre Matrix B gibt, sodass $A' = B^{-1}AB$, so nennt man A und A' *ähnlich* (zueinander). Dadurch wird für jedes $n \in \mathbb{N}$ eine Äquivalenzrelation auf der Menge aller $n \times n$ -Matrizen definiert. Eine interessante Aufgabe besteht darin, in jeder der zugehörigen Äquivalenzklassen einen möglichst einfachen Repräsentanten zu finden. Im Allgemeinen ist das nicht ganz einfach, die Lösung ist aber bekannt und heißt *Jordan'sche Normalform* (nach Camille Jordan, 1838-1922, hauptsächlich in Paris).

Für symmetrische Matrizen ist die Situation einfacher: Man kann zeigen, dass jede symmetrische Matrix zu einer Diagonalmatrix ähnlich ist (siehe Satz 53).

Beispiel: Sei $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ und $B = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Dann ist $B^{-1} = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$ und daher

$$A' = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} -7 & -11 \\ 5 & 8 \end{pmatrix}.$$

Betrachten wir z.B. den Vektor $x = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$, also $y = Ax = \begin{pmatrix} 9 \\ 4 \end{pmatrix}$.

$$x' = B^{-1}x = \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \quad y' = B^{-1}y = \begin{pmatrix} -19 \\ 14 \end{pmatrix},$$

andererseits ist aber auch $A'x' = \begin{pmatrix} -7 & -11 \\ 5 & 8 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -19 \\ 14 \end{pmatrix} = y'$.

Bemerkung: Wenn B eine Orthonormalbasis ist, dann gilt: Ein Endomorphismus f ist genau dann orthogonal, wenn seine Matrix bezüglich B orthogonal ist.

Beweis: Sei A' die Matrix von f bezüglich B , $A = BA'B^{-1}$ die Matrix von f bezüglich der kanonischen Basis, und $B^{-1} = B^T$. Wenn A' orthogonal ist, dann gilt $A^{-1} = BA'^{-1}B^{-1} = BA'^T B^T = A^T$. Wenn umgekehrt A orthogonal ist, dann ist $A'^{-1} = (B^{-1}AB)^{-1} = B^{-1}A^{-1}B = B^T A^T B = A'^T$. \square

8.2.2 Drehung im \mathbb{R}^3 um eine gegebene Achse

Wie berechnet man eine Drehung, wenn die Drehachse und der Drehwinkel vorgegeben sind? (Das ist z.B. eine wesentliche Frage bei der Robotersteuerung oder der graphischen Darstellung von Bewegungen.)

Der Einfachheit halber nehmen wir zunächst an, dass die Drehachse durch den Ursprung geht. Die Drehachse ist dann durch einen Vektor $b \neq o$ gegeben, von dem wir annehmen können, dass er normiert ist.

Wir wissen bereits, wie die Matrix D_γ^1 einer Drehung um den Winkel γ aussieht, wenn die Drehachse mit der ersten Koordinatenachse übereinstimmt (siehe Beispiel 2 in Abschnitt 7.2.7). Genauso sieht natürlich die Matrix einer Drehung in Bezug auf irgendeine Orthonormalbasis des \mathbb{R}^3 aus, wenn die Drehachse mit der durch den ersten Basisvektor definierten Achse übereinstimmt. Ergänzen wir also $\{b\}$ zu einer Orthonormalbasis B des \mathbb{R}^3 , so ist die Matrix der Drehung bezüglich der Basis B gleich D_γ^1 . Nach dem letzten Abschnitt muss daher für die gesuchte Matrix A gelten:

$$A = BD_\gamma^1B^{-1}.$$

Da B orthogonal ist, können wir auch schreiben: $A = BD_\gamma^1B^T$.

ACHTUNG: Je nachdem, wie man $\{b\}$ zu einer ONB ergänzt, erhält man zwei verschiedene Drehungen, die zueinander invers sind. Eindeutigkeit erreicht man, wenn man $\{b\}$ zu einer positiv orientierten ONB ergänzt (siehe Abschnitt 10.2.1).

Beispiel:

Sei $b = b_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Dann ist jedenfalls $\{b_1, e_2, e_3\}$ eine Basis des \mathbb{R}^3 , und das Schmidt'sche Orthonormalisierungsverfahren liefert folgendes:

$$\begin{aligned} b'_2 &= e_2 - (e_2 \cdot b_1) b_1 = \frac{1}{3} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \quad b_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \\ b'_3 &= e_3 - (e_3 \cdot b_1) b_1 - (e_3 \cdot b_2) b_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{6} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} = \\ &= \frac{1}{2} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad b_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \\ \text{also } B &= \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & -1 & -\sqrt{3} \\ \sqrt{2} & 2 & 0 \\ \sqrt{2} & -1 & \sqrt{3} \end{pmatrix}. \end{aligned}$$

Die Matrix A einer Drehung um die durch b bestimmte Achse und den Winkel $30^\circ = \frac{\pi}{6}$ ergibt sich daher so:

$$\begin{aligned} A &= BD_{\pi/6}^1 B^T = \frac{1}{6} \begin{pmatrix} \sqrt{2} & -1 & -\sqrt{3} \\ \sqrt{2} & 2 & 0 \\ \sqrt{2} & -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ -1 & 2 & -1 \\ -\sqrt{3} & 0 & \sqrt{3} \end{pmatrix} = \\ &= \frac{1}{3} \begin{pmatrix} 1 + \sqrt{3} & 1 - \sqrt{3} & 1 \\ 1 & 1 + \sqrt{3} & 1 - \sqrt{3} \\ 1 - \sqrt{3} & 1 & 1 + \sqrt{3} \end{pmatrix}. \end{aligned}$$

Zur Probe können wir z.B. folgendes überprüfen:

1. $Ab_1 = b_1$,
2. der Winkel zwischen b_2 und Ab_2 ist gleich $\frac{\pi}{6}$,

denn $\|b_2\| = \|Ab_2\| = 1$ und $b_2 \cdot Ab_2 = \frac{1}{6} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} -\sqrt{3} \\ \sqrt{3} \\ 0 \end{pmatrix} = \frac{1}{2}\sqrt{3} = \cos \frac{\pi}{6}$.

Wenn die Drehachse nicht durch den Ursprung geht, dann gehen wir so vor: Sei p ein Punkt auf der Drehachse. Wir führen zuerst eine Translation durch, die p in den Ursprung überführt, wenden dann die entsprechende Drehung

um die verschobene Achse an, und machen dann die Translation wieder rückgängig. Die gesuchte Abbildung ist daher

$$\alpha = \tau_p \circ f_A \circ \tau_{-p},$$

d.h.

$$\alpha(x) = A(x - p) + p = Ax + (p - Ap),$$

wobei $A = BD_\gamma^1 B^T$ wie oben ist.

9 Determinanten

9.1 Definition und grundlegende Eigenschaften

Unter der *Determinante* einer quadratischen Matrix A versteht man eine Zahl, die mit $\det A$ bezeichnet wird und folgenderweise rekursiv definiert werden kann:

1. Wenn A eine 1×1 -Matrix ist, also $A = (a)$, dann ist $\det A := a$.
2. Sei A eine $n \times n$ -Matrix mit Elementen a_{ik} und $n > 1$. Bezeichnen wir mit A'_{i1} die Matrix, die aus A durch Weglassen der i -ten Zeile und 1. Spalte entsteht, so ist

$$\begin{aligned} \det A &:= a_{11} \det A'_{11} - a_{21} \det A'_{21} + - \dots + (-1)^{n+1} a_{n1} \det A'_{n1} = \\ &= \sum_{i=1}^n (-1)^{i+1} a_{i1} \det A'_{i1}. \end{aligned}$$

Es gibt verschiedene zueinander äquivalente Definitionen der Determinante einer Matrix. Diese hier nennt man *Entwicklung nach der 1. Spalte*.

Häufig schreibt man statt $\det A$ auch $|A|$ oder ausführlicher

$$\left| \begin{array}{ccc} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{array} \right|$$

Sehen wir uns die Fälle $n = 2$ und 3 explizit an:

$$\left| \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right| = a_{11}a_{22} - a_{21}a_{12},$$

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

Häufig wird die Determinante einer $n \times n$ -Matrix A folgenderweise definiert:

$$\det A = \sum_{\sigma \in S_n} (\text{sign } \sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}. \quad (**)$$

Die Summation erstreckt sich dabei über alle Permutationen von $\{1, \dots, n\}$. (Bezüglich des Signums einer Permutation siehe z.B. [6].)

Für $n = 2$ und 3 sehen wir, dass diese Definition mit der vorher angegebenen übereinstimmt, und allgemein kann man das mit Induktion zeigen. Im Folgenden werden wir diese Definition allerdings nicht verwenden. Sie wird vor allem für Beweise gebraucht, die hier aus Zeitmangel gar nicht ausgeführt werden.

Die Determinante einer $n \times n$ -Matrix $A = (a_1, \dots, a_n)$ kann als Funktion der Spaltenvektoren a_i angesehen werden. Wir sagen dann auch einfach: $\det(a_1, \dots, a_n)$ ist die Determinante der Vektoren a_1, \dots, a_n . (Achtung: Die Determinante von n Vektoren ist nur im \mathbb{R}^n definiert. Man kann also z.B. nicht von der Determinante von zwei Vektoren im \mathbb{R}^3 reden.)

Wir überlegen uns nun zunächst einmal folgende Rechenregeln:

- a) $\det(\lambda a_1, a_2, \dots, a_n) = \lambda \det(a_1, \dots, a_n)$ für $\lambda \in \mathbb{R}$,
- b) $\det(a_1 + b, a_2, \dots, a_n) = \det(a_1, a_2, \dots, a_n) + \det(b, a_2, \dots, a_n)$ für $b \in \mathbb{R}^n$.

Beweis:

- a) $\det(\lambda a_1, a_2, \dots, a_n) = \sum_{i=1}^n (-1)^{i+1} \lambda a_{i1} \det A'_{i1} = \lambda \sum_{i=1}^n (-1)^{i+1} a_{i1} \det A'_{i1} = \lambda \det A$.
- b) $\det(a_1 + b, a_2, \dots, a_n) = \sum_{i=1}^n (-1)^{i+1} (a_{i1} + b_i) \det A'_{i1} = \sum_{i=1}^n (-1)^{i+1} a_{i1} \det A'_{i1} + \sum_{i=1}^n (-1)^{i+1} b_i \det A'_{i1} = \det A + \det(b, a_2, \dots, a_n)$.
□

Nicht ganz so einfach ist die folgende Regel einzusehen:

Satz 39 Sei \tilde{A} eine Matrix, die aus A durch Vertauschung zweier Spalten entsteht. Dann ist $\det \tilde{A} = -\det A$.

Für $n = 2$ ist das leicht nachzurechnen:

$$\begin{vmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{vmatrix} = a_{12}a_{21} - a_{22}a_{11} = -(a_{11}a_{22} - a_{21}a_{12}).$$

Auch für $n = 3$ ist es eine relativ einfache Übungsaufgabe. Der allgemeine Beweis kann mit vollständiger Induktion geführt werden oder unter Benutzung der zweiten Definition (**) für die Determinante. Die etwas komplizierten Details lassen wir hier aber weg. \square

Auf Grund von Satz 39 ist leicht zu sehen, dass die obigen Rechenregeln für jede beliebige Spalte an Stelle der 1. Spalte gelten:

Satz 40 a) $\det(\dots, \lambda a_k, \dots) = \lambda \det(\dots, a_k, \dots)$,

b) $\det(\dots, a_k + b, \dots) = \det(\dots, a_k, \dots) + \det(\dots, b, \dots)$.

Beweis für $k > 1$: Wir vertauschen zuerst die k -te Spalte mit der 1. Spalte, wenden dann die entsprechende Rechenregel für die 1. Spalte an und vertauschen dann wieder die 1. Spalte mit der k -ten Spalte. Dabei ändert sich das Vorzeichen der Determinante zweimal, bleibt also insgesamt gleich. \square

Dieser Satz besagt, dass die Abbildung

$$(\mathbb{R}^n)^n \rightarrow \mathbb{R} : (a_1, \dots, a_n) \mapsto \det(a_1, \dots, a_n)$$

eine *Multilinearform* ist. Der vorhergehende Satz besagt dann, dass diese Multilinearform *alternierend* (auch: *antisymmetrisch*) ist.

1. *Bemerkung:* Wenn in einer Matrix zwei Spalten übereinstimmen, dann ist ihre Determinante gleich Null, d.h.:

$$\det(\dots, a, \dots, a, \dots) = 0.$$

Beweis: Da sich die Matrix bei Vertauschung der beiden gleichen Spalten überhaupt nicht ändert, aber andererseits die Determinante das Vorzeichen ändert, muss gelten:

$$\det(\dots, a, \dots, a, \dots) = -\det(\dots, a, \dots, a, \dots),$$

und daraus folgt schon die Behauptung. \square

2. *Bemerkung:* Wenn man die ganze Matrix mit λ multipliziert, wird jede der n Spalten mit λ multipliziert, und daher gilt: $\det(\lambda A) = \lambda^n \det A$.

Satz 41 Die Determinante einer Matrix bleibt bei der Addition eines skalaren Vielfachen einer Spalte zu einer anderen Spalte unverändert.

Beweis:

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

$$\begin{aligned} \det(\dots, a_r + \lambda a_s, \dots, a_s, \dots) &= \det(\dots, a_r, \dots, a_s, \dots) + \det(\dots, \lambda a_s, \dots, a_s, \dots) = \\ &= \det A + \lambda \det(\dots, a_s, \dots, a_s, \dots) = \det A \text{ auf Grund der vorhergehenden ersten Bemerkung. } \square \end{aligned}$$

Ohne Beweis notieren wir:

Satz 42 $\det A^T = \det A$.

(Für 2×2 -Matrizen ist das ganz leicht nachzuprüfen. Für den Beweis des allgemeinen Falls kann man die zweite Definition (**) der Determinante benützen.)

Aus diesem Satz ergibt sich, dass alles, was wir vorhin über Spalten gesagt haben, analog für Zeilen gilt:

Satz 43 Die Determinante einer $n \times n$ -Matrix ist eine alternierende Multilinearform der Zeilenvektoren. Sie bleibt bei der Addition eines skalaren Vielfachen einer Zeile zu einer anderen Zeile unverändert.

Wir sehen also, dass die elementaren Umformungen, die wir in Abschnitt 4.1 besprochen haben, die Determinante einer Matrix unverändert lassen, mit Ausnahme der Multiplikation einer Zeile mit einem Skalarfaktor λ und abgesehen von einer Vorzeichenänderung bei der Vertauschung zweier Spalten. Mit den Überlegungen von Abschnitt 4.1 ist nun leicht zu sehen, dass man jede quadratische Matrix durch elementare Umformungen, welche die Determinante bis auf das Vorzeichen unverändert lassen, auf folgende *Halbdiagonalf orm* bringen kann:

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \dots & \dots & \tilde{a}_{1r} & \dots & \tilde{a}_{1n} \\ 0 & \ddots & & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \dots & 0 & \tilde{a}_{rr} & \dots & \tilde{a}_{rn} \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \text{ mit } \tilde{a}_{11}, \dots, \tilde{a}_{rr} \neq 0.$$

Dabei ist natürlich wieder $r = \text{rank } A$. In der Halbdiagonalform lässt sich die Determinante sehr leicht berechnen:

$$\det \tilde{A} = \begin{cases} \tilde{a}_{11}\tilde{a}_{22} \cdots \tilde{a}_{nn} & \text{falls } r = n, \\ 0 & \text{falls } r < n, \end{cases}$$

wie man mit Entwicklung nach der 1. Spalte und vollständiger Induktion leicht sieht. Daraus erkennen wir nun unmittelbar den folgenden ganz zentralen Satz, der erst den Sinn der Determinanten verständlich macht:

Satz 44 *Sei A eine $n \times n$ -Matrix. Dann gilt:*

$$\det A = 0 \Leftrightarrow \text{rank } A < n.$$

Die Determinante einer Matrix ist also genau dann Null, wenn ihre Spaltenvektoren bzw. Zeilenvektoren linear abhängig sind.

Beispiel:

$$\left| \begin{array}{ccc} 2 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 0 & 5 \end{array} \right| = \left| \begin{array}{ccc} 2 & 2 & 3 \\ 0 & -1 & -4 \\ 0 & -3 & \frac{1}{2} \end{array} \right| = \left| \begin{array}{ccc} 2 & 2 & 3 \\ 0 & -1 & -4 \\ 0 & 0 & \frac{25}{2} \end{array} \right| = -25, \text{ und der Rang dieser Matrix ist gleich 3.}$$

Satz 45 *Ein lineares Gleichungssystem $Ax = b$ mit quadratischer Matrix A ist genau dann eindeutig lösbar, wenn $\det A \neq 0$ ist.*

Beweis: Das folgt aus dem vorigen Satz auf Grund von Satz 19. \square

Der folgende Satz ist wieder etwas mühsam zu beweisen. Wir lassen daher den Beweis hier weg:

Satz 46 (Determinantenmultiplikationssatz) *Seien A und B zwei $n \times n$ -Matrizen. Dann gilt:*

$$\det AB = (\det A)(\det B).$$

Insbesondere folgt daraus mit Satz 44:

Satz 47 *Eine quadratische Matrix A ist genau dann invertierbar, wenn $\det A \neq 0$ ist, und in diesem Fall gilt*

$$\det A^{-1} = \frac{1}{\det A}.$$

Beweis: A ist genau dann invertierbar, wenn $\text{rank } A = n$, d.h. $\det A \neq 0$, und aus dem Determinantenmultiplikationssatz folgt dann $(\det A)(\det A^{-1}) = \det(AA^{-1}) = \det E = 1$ und damit die Behauptung. \square

Beispiel: Wir können die Determinante von $\begin{pmatrix} 2 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 0 & 5 \end{pmatrix}^{-1}$ berechnen, ohne die Matrixinversion durchzuführen: Auf Grund des letzten Beispiels ist sie gleich $-\frac{1}{25}$.

Wir können jetzt auch leicht feststellen, wie sich die Determinante einer Matrix bei einer Koordinatentransformation verhält: Für jede reguläre Matrix B mit derselben Zeilenanzahl wie A gilt:

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

d.h. die Determinante einer Matrix ist invariant gegenüber Koordinatentransformationen. Es ist daher sinnvoll, von der *Determinante eines Endomorphismus* f zu sprechen: $\det f := \det A$, wenn A die Matrix von f ist (bezüglich einer beliebigen Basis des \mathbb{R}^n).

Analog zur Entwicklung einer Determinante nach der 1. Spalte gemäß Definition kann man eine Determinante auch nach einer beliebigen Zeile oder Spalte entwickeln.

Satz 48 (Determinantenentwicklungssatz) Sei A eine $n \times n$ -Matrix und $s \in \{1, \dots, n\}$. Dann gilt

$$\det A = \sum_{i=1}^n (-1)^{i+s} a_{is} \det A'_{is} \quad (\text{Entwicklung nach der } s\text{-ten Spalte}),$$

und

$$\det A = \sum_{k=1}^n (-1)^{s+k} a_{sk} \det A'_{sk} \quad (\text{Entwicklung nach der } s\text{-ten Zeile}).$$

(Dabei bedeutet A'_{ik} die Matrix, welche aus A durch Weglassen der i -ten Zeile und der k -ten Spalte entsteht.)

Der *Beweis* beruht auf den Sätzen 39 und 42. Die Details lassen wir wieder weg. \square

Beispiel: Wir entwickeln die folgende Determinante nach der 2. Zeile, um den Rechenaufwand möglichst gering zu halten. Die Determinanten der entstehenden 3×3 -Matrizen entwickeln wir jeweils nach der 3. Zeile:

$$\begin{aligned}
\begin{vmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 1 & 0 \\ -6 & 3 & 1 & 1 \\ 2 & 0 & -1 & 1 \end{vmatrix} &= -2 \begin{vmatrix} 2 & 3 & 4 \\ 3 & 1 & 1 \\ 0 & -1 & 1 \end{vmatrix} - \begin{vmatrix} 0 & 2 & 4 \\ -6 & 3 & 1 \\ 2 & 0 & 1 \end{vmatrix} = \\
&= -2 \left(-(-1) \begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix} \right) - 2 \begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} - \begin{vmatrix} 0 & 2 \\ -6 & 3 \end{vmatrix} = \\
&= -2(-10 - 7) + 2 \cdot 10 - 12 = 42.
\end{aligned}$$

Satz 49 Sei A eine reguläre Matrix. Dann gilt

$$A^{-1} = \frac{1}{\det A} (c_{ik})$$

mit

$$c_{ik} = (-1)^{i+k} \det A'_{ki}.$$

(Achtung auf die Reihenfolge der Indizes bei c_{ik} und A'_{ki} !)

Beweis: Sei $C = (c_{ik})$ mit c_{ik} wie im Satz angegeben. Wir haben zu zeigen: $AC = (\det A)E$, denn das heißt $A\left(\frac{1}{\det A}C\right) = E$.

$AC = (d_{is})$ mit

$$d_{is} = \sum_{k=1}^n a_{ik} c_{ks} = \sum_{k=1}^n (-1)^{k+s} a_{ik} \det A'_{sk}.$$

Für $i = s$ ist das die Formel für die Entwicklung nach der s -ten Zeile, und daher gilt $d_{ss} = \det A$.

Für $i \neq s$ ist es die Entwicklung nach der s -ten Zeile für diejenige Matrix, die aus A entsteht, wenn man die s -te Zeile durch die i -te Zeile ersetzt. Das ist aber eine Matrix, in der zwei Zeilen übereinstimmen. Ihre Determinante ist daher gleich Null, d.h. $d_{is} = 0$ für $i \neq s$. \square

Diese Methode zur Matrizeninversion eignet sich besonders für kleine Matrizen. Insbesondere erhalten wir folgende einfache Formel für die Inverse einer 2×2 -Matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Ein lineares Gleichungssystem $Ax = b$ mit einer regulären Matrix A kann man im Prinzip mit Hilfe der inversen Matrix lösen: $x = A^{-1}b$. Im Allgemeinen ist jedoch die Berechnung von A^{-1} aufwendiger als die Lösung mit

Hilfe des Gauß'schen Eliminationsverfahrens. Manchmal, insbesondere für kleine Matrizen, hat es jedoch einen Sinn, die Lösung mit Hilfe der oben angegebenen Inversionsformel zu berechnen. Dabei kann man Rechenarbeit sparen, wenn man nur einige der Unbekannten x_1, \dots, x_n berechnen will:

Aus $x = A^{-1}b$ folgt mit den obigen Bezeichnungen

$$x_i = \frac{1}{\det A} \sum_{k=1}^n c_{ik} b_k = \frac{1}{\det A} \sum_{k=1}^n (-1)^{i+k} b_k \det A'_{ki}.$$

Diese Summe kann man als Entwicklung von $\det(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n)$ nach der i -ten Spalte auffassen, und daher gilt

$$x_i = \frac{\det(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n)}{\det A}.$$

Diese Methode zur Lösung eines linearen Gleichungssystems mit regulärer Matrix heißt *Cramer'sche Regel*.

Beispiel:

Sei $A = \begin{pmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 1 & 0 \\ -6 & 3 & 1 & 1 \\ 2 & 0 & -1 & 1 \end{pmatrix}$ die Matrix des letzten Beispiels. Wir wollen das lineare Gleichungssystem $Ax = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}$ lösen. Mit der Cramer'schen Regel erhalten wir z.B.

$$x_3 = \frac{\begin{vmatrix} 0 & 2 & 1 & 4 \\ 2 & 0 & 2 & 0 \\ -6 & 3 & 0 & 1 \\ 2 & 0 & 0 & 1 \end{vmatrix}}{\det A} = \frac{-2 \begin{vmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \\ -6 & 3 & 0 \end{vmatrix}}{42} = \frac{-2(-20) + (-18)}{42} = \frac{11}{21}.$$

Satz 50 Sei A eine orthogonale Matrix. Dann gilt: $|\det A| = 1$.

Beweis: Wenn A orthogonal ist, dann ist $A^{-1} = A^T$ und daher

$$\det A = \frac{1}{\det A^{-1}} = \frac{1}{\det A^T} = \frac{1}{\det A}, \text{ also } (\det A)^2 = 1. \quad \square$$

Bemerkung: Von diesem Satz gilt nicht die Umkehrung, wie folgendes Beispiel zeigt:

Sei $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Dann ist $\det A = 1$, aber A ist nicht orthogonal: $A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.

Satz 51 *Die Spiegelung an einer Hyperebene durch den Ursprung ist eine orthogonale Abbildung mit Determinante -1 .*

Sei σ_H die Spiegelung an einer Hyperebene H des \mathbb{R}^n , und $o \in H$. Dann gibt es eine ONB (u_1, \dots, u_{n-1}) von H , und diese bildet zusammen mit einem normierten Normalvektor u_n von H eine ONB des ganzen \mathbb{R}^n . Wie sieht nun die Matrix S von σ_H bezüglich dieser Basis aus?

$$\sigma_H(u_i) = u_i - 2(u_i \cdot u_n)u_n = \begin{cases} u_i & \text{für } i \in \{1, \dots, n-1\}, \\ -u_n & \text{für } i = n, \end{cases}$$

also gilt

$$S = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix}.$$

σ_H ist also eine orthogonale Abbildung mit Determinante -1 . \square

Es gibt natürlich noch andere orthogonale Abbildungen mit Determinante -1 , z.B. die Spiegelung am Ursprung im \mathbb{R}^3 (oder allgemeiner im \mathbb{R}^n für ungerades n). Die Matrix einer solchen Punktspiegelung ist ja gleich $-E$, und $\det(-E) = (-1)^n$.

Historische Bemerkung:

Die Determinante wurde unabhängig voneinander von Gottfried Wilhelm Leibniz (1646-1716) in Deutschland und Takakazu Seki Kowa (1642-1708) in Japan entdeckt, allerdings zunächst nur für lineare Gleichungssysteme mit 2, 3 oder höchstens 4 Unbekannten. Beide Mathematiker waren allerdings mit Publikationen darüber sehr zurückhaltend, und so wurde sie später mehrmals wiederentdeckt, u.a. von Gabriel Cramer (1704-1752, Genf, Basel). (Vgl. [8].)

9.2 Geometrische Bedeutung der Determinante im \mathbb{R}^2

9.2.1 Orientierung zweier Vektoren im \mathbb{R}^2

Seien a und b zwei linear unabhängige Vektoren im \mathbb{R}^2 , also $\det(a, b) \neq 0$. Welche Bedeutung hat das Vorzeichen dieser Determinante? Betrachten wir dazu die entsprechenden normierten Vektoren $a' = \frac{1}{\|a\|}a$ und $b' = \frac{1}{\|b\|}b$. Dann ist $\det(a', b') = \frac{1}{\|a\|\|b\|} \det(a, b)$, also hat $\det(a', b')$ dasselbe Vorzeichen wie $\det(a, b)$.

Es gibt eindeutig bestimmte Zahlen $\alpha, \beta \in (-\pi, \pi]$, sodass

$$\begin{aligned} a' &= (\cos \alpha, \sin \alpha), \\ b' &= (\cos \beta, \sin \beta). \end{aligned}$$

Es gilt also:

$$\det(a', b') = \cos \alpha \sin \beta - \sin \alpha \cos \beta = \sin(\beta - \alpha) \neq 0.$$

Die Zahl $\beta - \alpha$ liegt irgendwo im Intervall $(-\pi, 2\pi)$, möglicherweise außerhalb von $(-\pi, \pi)$. Es gibt aber eine eindeutig bestimmte Zahl $\gamma \in (-\pi, \pi)$, sodass $\sin(\beta - \alpha) = \sin \gamma$ und $\cos(\beta - \alpha) = \cos \gamma$, d.h. $\beta - \alpha \equiv \gamma \pmod{2\pi}$, und wir sehen:

$$\text{sign}(\det(a, b)) = \text{sign} \gamma.$$

γ ist auch dadurch charakterisiert, dass die Drehung um den Winkel γ den Vektor a' in den Vektor b' überführt:

$$D_\gamma a' = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = \begin{pmatrix} \cos(\alpha + \gamma) \\ \sin(\alpha + \gamma) \end{pmatrix} = \begin{pmatrix} \cos \beta \\ \sin \beta \end{pmatrix} = b',$$

da $\alpha + \gamma \equiv \beta \pmod{2\pi}$.

$\det(a, b)$ ist also genau dann positiv, wenn man von a' nach b' durch eine Drehung um einen positiven Winkel $\gamma \in (0, \pi)$ kommt. Man sagt dann auch: die geordnete Basis (a, b) des \mathbb{R}^2 ist *positiv orientiert*.

Diese Zahl γ nennt man auch den *orientierten Winkel* zwischen a und b , wobei man zusätzlich definiert:

$$\gamma := 0, \text{ falls } b = \lambda a \text{ mit } \lambda > 0,$$

$$\gamma := \pi, \text{ falls } b = \lambda a \text{ mit } \lambda < 0.$$

Natürlich stimmt $|\gamma|$ mit dem früher definierten Winkel zwischen a und b überein, denn $a' \cdot b' = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \cos(\alpha - \beta) = \cos \gamma$.

Achtung: Im \mathbb{R}^n mit $n > 2$ hat es keinen Sinn, vom orientierten Winkel zweier Vektoren zu sprechen.

Wenden wir auf a und b einen Endomorphismus f_T an, so gilt $\det(f_T(a), f_T(b)) = \det(Ta, Tb) = \det(T(a, b)) = (\det T)(\det(a, b))$. Das heißt, die Orientierung von (a, b) bleibt genau dann erhalten, wenn $\det T > 0$. In diesem Fall nennen wir f_T einen *orientierungstreuen* Endomorphismus. Wir sehen z.B., dass jede Drehung D_γ im \mathbb{R}^2 orientierungstreu ist, denn $\det(D_\gamma) = \cos^2 \gamma + \sin^2 \gamma = 1$. Andererseits ist die Spiegelung an einer Geraden nicht orientierungstreu, da ihre Determinante $= -1$ ist.

9.2.2 Flächeninhalt eines Parallelogramms

Seien a und b zwei Vektoren $\neq o$ im \mathbb{R}^n und

$$P(a, b) := \{\lambda a + \mu b \in \mathbb{R}^n : 0 \leq \lambda, \mu \leq 1\}$$

das von a und b aufgespannte Parallelogramm. Der *Flächeninhalt* eines solchen Parallelogramms wird auf Grund elementargeometrischer Überlegungen durch

$$F(P(a, b)) := \|a\| \|b\| |\sin \gamma|$$

definiert, wobei γ der Winkel zwischen a und b ist. Man kann das so auffassen, dass etwa $\|a\|$ die (Länge der) Grundlinie des Parallelogramms ist, und $\|b\| |\sin \gamma|$ die Höhe bezüglich dieser Grundlinie. Dann ist der Flächeninhalt gleich Grundlinie mal Höhe.

Mit der bekannten Formel für den Cosinus des Winkels zweier Vektoren erhalten wir:

$$F(P(a, b))^2 = \|a\|^2 \|b\|^2 (1 - \cos^2 \gamma) = \|a\|^2 \|b\|^2 \left(1 - \frac{(a \cdot b)^2}{\|a\|^2 \|b\|^2}\right),$$

also

$$F(P(a, b))^2 = \|a\|^2 \|b\|^2 - (a \cdot b)^2.$$

Im \mathbb{R}^2 kann man den Flächeninhalt eines Parallelogramms $P(a, b)$ auch mit Hilfe der Determinante der Vektoren a, b berechnen:

$$\begin{aligned} \|a\|^2 \|b\|^2 - (a \cdot b)^2 &= (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2 = \\ &= a_1^2 b_2^2 + a_2^2 b_1^2 - 2a_1 b_1 a_2 b_2 = (a_1 b_2 - a_2 b_1)^2, \text{ also} \end{aligned}$$

$$F(P(a, b)) = |\det(a, b)|.$$

Wir können jetzt noch die Frage behandeln, wie sich der Flächeninhalt eines Parallelogramms ändert, wenn man eine lineare Abbildung darauf anwendet. Sei also etwa T eine 2×2 -Matrix und $f = f_T$ die zugehörige lineare Abbildung. Wegen der Linearität gilt offensichtlich

$$f(P(a, b)) = P(Ta, Tb),$$

und daher ist der Flächeninhalt davon gleich $|\det(Ta, Tb)|$. Nun gilt aber $(Ta, Tb) = T(a, b)$. Aus dem Determinantenmultiplikationssatz folgt daher

$$F(f(P(a, b))) = |\det T| |\det(a, b)|.$$

Wir sehen also: Das Bild eines Parallelogramms unter einem Endomorphismus f ist wieder ein Parallelogramm, und der Flächeninhalt wird bei der Abbildung mit dem Faktor $|\det f| = |\det T|$ multipliziert.

Der Betrag der Determinante einer 2×2 -Matrix T hat also zweierlei geometrische Bedeutung: Er ist einerseits der Flächeninhalt des von den Spaltenvektoren aufgespannten Parallelogramms, und andererseits der Faktor, mit dem der Inhalt eines beliebigen Parallelogramms bei Anwendung der zu T gehörigen linearen Abbildung multipliziert wird.

10 Äußeres Produkt im \mathbb{R}^3

10.1 Definition und grundlegende Eigenschaften

Seien $a, b \in \mathbb{R}^3$. Wegen der Multilinearität der Determinante ist insbesondere die Abbildung

$$\mathbb{R}^3 \rightarrow \mathbb{R} : x \mapsto \det(a, b, x)$$

linear. Die zugehörige Matrix (siehe Satz 3) hat nur eine Zeile, wir können sie daher als Vektor auffassen. Es gibt also einen eindeutig bestimmten Vektor $v \in \mathbb{R}^3$, sodass

$$\det(a, b, x) = v \cdot x \quad \text{für alle } x \in \mathbb{R}^3.$$

Dieser Vektor v heißt das *äußere Produkt*, *Vektorprodukt* (auch: *vektorielles Produkt*) oder *Kreuzprodukt* von a und b . Bezeichnung: $v = a \times b$. (In analoger Weise kann man im \mathbb{R}^n das Vektorprodukt von $n - 1$ Vektoren definieren. Der Fall $n = 3$ kommt aber in den Anwendungen bei weitem am häufigsten vor.)

Auf Grund der Definition gilt also für $a, b, c \in \mathbb{R}^3$:

$$(a \times b) \cdot c = \det(a, b, c).$$

Da allgemein $v \cdot e_k = v_k$ ist (vgl. Satz 31), gilt folgende Formel für die Koordinaten von $a \times b$:

$$(a \times b)_k = \det(a, b, e_k) \text{ für } k \in \{1, 2, 3\}.$$

Durch Entwicklung nach der letzten Spalte erhalten wir daraus:

$$a \times b = \begin{pmatrix} & \left| \begin{array}{cc} a_2 & b_2 \\ a_3 & b_3 \end{array} \right| \\ - & \left| \begin{array}{cc} a_1 & b_1 \\ a_3 & b_3 \end{array} \right| \\ & \left| \begin{array}{cc} a_1 & b_1 \\ a_2 & b_2 \end{array} \right| \end{pmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

Einige wichtige Eigenschaften des Vektorprodukts sind die folgenden:

- a) $a \times (b + c) = a \times b + a \times c$,
- b) $\lambda(a \times b) = \lambda a \times b$,
- c) $a \times b$ ist orthogonal zu a und b , d.h. $(a \times b) \cdot a = (a \times b) \cdot b = 0$,
- d) $b \times a = -a \times b$,
- e) $\|a \times b\| = F(P(a, b))$ (= Flächeninhalt des von a und b aufgespannten Parallelogramms),
- f) $a \times b = o$ genau dann, wenn a und b linear abhängig sind.

Beweis:

- a) $\det(a, b + c, x) = \det(a, b, x) + \det(a, c, x)$ (Satz 40).
- b) $\lambda \det(a, b, x) = \det(\lambda a, b, x)$ (Satz 40).
- c) $\det(a, b, a) = \det(a, b, b) = 0$ (siehe 1. Bemerkung nach Satz 40).
- d) $\det(b, a, x) = -\det(a, b, x)$ (Satz 39).
- e) $\|a \times b\|^2 = (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - a_1 b_3)^2 + (a_1 b_2 - a_2 b_1)^2 = a_2^2 b_3^2 - 2a_2 b_3 a_3 b_2 + a_3^2 b_2^2 + a_3^2 b_1^2 - 2a_3 b_1 a_1 b_3 + a_1^2 b_3^2 + a_1^2 b_2^2 - 2a_1 b_2 a_2 b_1 + a_2^2 b_1^2$,

andererseits ist

$$F(P(a, b))^2 = \|a\|^2 \|b\|^2 - (a \cdot b)^2 = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2, \text{ und das ergibt ausmultipliziert dasselbe.}$$

f) $a \times b = o \Leftrightarrow \|a \times b\| = 0$, und das ist nach e) genau dann der Fall, wenn $\|a\|^2 \|b\|^2 - (a \cdot b)^2 = 0$, d.h. wenn in der Cauchy-Schwarz'schen Ungleichung Gleichheit eintritt, also genau dann, wenn a und b linear abhängig sind. \square

Bemerkungen:

1. Es gilt auch $\lambda(a \times b) = a \times \lambda b$, denn nach d) und b) folgt:

$$a \times \lambda b = -\lambda b \times a = -\lambda(b \times a) = \lambda(a \times b).$$

2. $(a + b) \times c = a \times c + b \times c$, denn nach d) und a) folgt:

$$(a + b) \times c = -c \times (a + b) = -c \times a - c \times b = a \times c + b \times c.$$

Zusammen bedeutet das: die Abbildung $\mathbb{R}^3 \times \mathbb{R}^3 : (a, b) \mapsto a \times b$ ist *alternierend* und *bilinear*. (Man verwendet hier nicht die Worte "Bilinearform" oder "Multilinearform", da sich diese auf skalarwertige Abbildungen beziehen.)

3. *ACHTUNG:* Für das Vektorprodukt gilt kein Assoziativgesetz: $a \times (b \times c)$ und $(a \times b) \times c$ sind zwar beide sinnvoll, aber im Allgemeinen verschieden voneinander. Das kommt z.B. durch die folgende "Grassmann-Identität" zum Ausdruck :

$$a \times (b \times c) = (a \cdot c)b - (a \cdot b)c.$$

(Das kann man durch direktes Nachrechnen beweisen, was zwar etwas umständlich, aber völlig elementar ist.)

Auf Grund dieser Beziehung gilt nämlich

$$(a \times b) \times c = -c \times (a \times b) = -(c \cdot b)a + (c \cdot a)b = (a \cdot c)b - (b \cdot c)a,$$

und wir sehen: $a \times (b \times c) \in \text{lin}(b, c)$ und $(a \times b) \times c \in \text{lin}(a, b)$. Es gibt natürlich Spezialfälle, wo die beiden Vektorprodukte übereinstimmen, z.B. wenn b auf a und c senkrecht steht.

Beispiele:

1. Sei $a = (1, 2, 3)$ und $b = (2, 0, -1)$. Dann ist $c := a \times b = (-2, 7, -4)$, und wir können leicht überprüfen, dass $a \cdot c = b \cdot c = 0$ ist.

$$F(P(a, b))^2 = \|a\|^2 \|b\|^2 - (a \cdot b)^2 = 14 \cdot 5 - (-1)^2 = 69, \text{ und } \|c\|^2 = c \cdot c = 69.$$

2. Sei $a = (1, 0, 0)$, $b = (0, 1, 0)$ und $c = (1, 2, 3)$.

Dann ist $a \times b = (0, 0, 1)$ und daher $(a \times b) \times c = (-2, 1, 0)$,

andererseits ist $b \times c = (3, 0, -1)$ und somit $a \times (b \times c) = (0, 1, 0)$.

Beide Produkte können wir auch mit der Grassmann-Identität ausrechnen:

$$(a \times b) \times c = (a \cdot c)b - (b \cdot c)a = 1b - 2a = (-2, 1, 0),$$

$$a \times (b \times c) = (a \cdot c)b - (a \cdot b)c = 1b - 0c = b = (0, 1, 0).$$

3. $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$.

Da $a \times b$ auf a und b senkrecht steht, kann man das Vektorprodukt benutzen, um zu einer in Parameterform gegebenen Ebene einen Normalvektor und damit auch eine Gleichung der Ebene in einfacher Weise zu berechnen:

Sei H eine Ebene mit Parameterdarstellung $x = p + \lambda a + \mu b$, wobei natürlich a und b linear unabhängig sein müssen. Mit $u := a \times b$ ist daher $u \cdot x = u \cdot p$ eine Gleichung von H .

Historische Bemerkung:

Hermann Grassmann (1809-1877) ist einer der Begründer der modernen linearen Algebra. Er war allerdings fast sein ganzes Berufsleben lang als Gymnasiallehrer in Stettin tätig. Die Bedeutung seiner Entdeckungen wurde erst spät erkannt bzw. anerkannt. So geht z.B. der sogenannte "Austauschsatz von Steinitz" (das ist eine Variante des Basisergänzungssatzes (Satz 12)) auf ihn zurück (siehe [3]).

Ernst Steinitz (1871-1928) ist vor allem wegen seiner Beiträge zur Theorie der (algebraischen) Körper bekannt, hat sich aber z.B. auch mit Polyedern beschäftigt.

10.2 Geometrische Bedeutung von Vektorprodukt und Determinante im \mathbb{R}^3

10.2.1 Orientierung im \mathbb{R}^3

Mit den obigen Eigenschaften c) und e) kann man das Vektorprodukt schon fast rein geometrisch, also ohne Bezugnahme auf Koordinaten oder Determinanten, erklären: Wenn a und b linear unabhängig sind, dann ist $a \times b$ ein Normalvektor der durch a und b erzeugten Ebene, und die Länge dieses

Vektors gibt an, wie groß das von a und b aufgespannte Parallelogramm ist. Damit ist $a \times b$ bis auf einen Faktor ± 1 eindeutig bestimmt. Wie kann man nun anschaulich-geometrisch entscheiden, welcher der beiden in Frage kommenden Vektoren tatsächlich das Vektorprodukt ist?

Sehen wir uns dazu den Spezialfall an, dass a und b in der durch e_1 und e_2 erzeugten Ebene liegen, und betrachten wir die entsprechenden normierten Vektoren: $a' = \frac{1}{\|a\|}a$ und $b' = \frac{1}{\|b\|}b$. Wegen $a' \times b' = \frac{1}{\|a\|\|b\|}(a \times b)$ zeigt $a' \times b'$ in dieselbe Richtung wie $a \times b$.

Ähnlich wie bei der Behandlung der Orientierung im \mathbb{R}^2 gibt es $\alpha, \beta \in (-\pi, \pi]$, sodass $a' = (\cos \alpha, \sin \alpha, 0)$ und $b' = (\cos \beta, \sin \beta, 0)$, und es gilt

$$a' \times b' = (0, 0, \cos \alpha \sin \beta - \sin \alpha \cos \beta) = (0, 0, \sin(\beta - \alpha)).$$

Sei wieder $\gamma \in (-\pi, \pi)$, sodass $\beta - \alpha \equiv \gamma \pmod{2\pi}$. Dann gilt $D_\gamma^3 a' = b'$, wobei D_γ^3 die Matrix der Drehung um die dritte Koordinatenachse um den Winkel γ bedeutet:

$$D_\gamma^3 a' = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \alpha \\ \sin \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\alpha + \gamma) \\ \sin(\alpha + \gamma) \\ 0 \end{pmatrix},$$

und $a \times b$ zeigt in die Richtung $(0, 0, 1)$, wenn $\gamma > 0$, und in die Richtung $(0, 0, -1)$, wenn $\gamma < 0$.

Den allgemeinen Fall kann man auf diesen Spezialfall zurückführen, wenn man eine geeignete orthogonale Koordinatentransformation durchführt.

Die Situation kann man anschaulich folgendermaßen interpretieren: Betrachtet man die Drehung, welche a' in b' überführt, dann zeigt $a \times b$ in die Richtung, in die sich eine Rechtsschraube bei dieser Drehung bewegen würde. Eine andere anschauliche Interpretation lautet folgendermaßen: Wenn a und b Daumen und Zeigefinger der rechten Hand entsprechen, dann zeigt $a \times b$ in die Richtung des Mittelfingers der rechten Hand (bei einigermaßen natürlicher Spreizung der Finger).

Wegen $\det(a, b, c) = (a \times b) \cdot c$ sehen wir jetzt, dass das Vorzeichen der Determinante von drei Vektoren im \mathbb{R}^3 folgendes über die Orientierung dieser Vektoren aussagt: Wenn $\det(a, b, c)$ positiv ist, dann ist die orthogonale Projektion von c auf die Gerade durch $a \times b$ ein positives Vielfaches von $a \times b$. Wir können das auch so ausdrücken: c zeigt auf dieselbe Seite der von a und b erzeugten Ebene wie der Vektor $a \times b$. Das heißt anschaulich, a, b und c liegen im Prinzip so im Raum wie die ersten drei Finger der rechten Hand.

Allgemein nennen wir eine geordnete Basis (b_1, \dots, b_n) des \mathbb{R}^n *positiv orientiert*, wenn $\det(b_1, \dots, b_n) > 0$ ist. Wie im \mathbb{R}^2 sehen wir, dass ein Endomorphismus f genau dann die Orientierung einer Basis unverändert lässt, wenn seine Determinante positiv ist. So ein Endomorphismus heißt daher *orientierungstreu*. So sind z.B. Drehungen im \mathbb{R}^3 orientierungstreu (Determinante = 1), dagegen sind Spiegelungen an Hyperebenen niemals orientierungstreu (Determinante = -1 , siehe Satz 51). Bewegungen, deren zugehörige lineare Abbildung orientierungstreu ist, heißen auch *eigentliche Bewegungen*. Eine Spiegelung ist also eine uneigentliche Bewegung.

Vektorprodukt und Determinante sind wichtige Hilfsmittel bei der computergeometrischen Behandlung von Problemen, die mit Umlaufssinn, links-rechts, innen-aussen und dergleichen zusammenhängen.

10.2.2 Volumen eines Parallelepipeds

Ein *Parallelelepiped* ist das dreidimensionale Analogon zu einem Parallelogramm. Genauer versteht man unter dem von drei Vektoren $a, b, c \in \mathbb{R}^3$ aufgespannten Parallelelepiped die Menge

$$P(a, b, c) = \{\lambda a + \mu b + \nu c \in \mathbb{R}^3 : 0 \leq \lambda, \mu, \nu \leq 1\}.$$

Elementargeometrisch versteht man unter dem Volumen von $P(a, b, c)$ den Flächeninhalt der "Grundfläche" $P(a, b)$ multipliziert mit der "Höhe" h , das ist der Abstand des Punktes c von der durch a, b erzeugten Ebene.

Die Gleichung dieser Ebene lautet $(a \times b) \cdot x = 0$, daher ist $h = \left| \frac{a \times b}{\|a \times b\|} \cdot c \right|$ (siehe Abschnitt 7.2.6) und

$$F(P(a, b)) h = \|a \times b\| \left| \frac{a \times b}{\|a \times b\|} \cdot c \right| = |(a \times b) \cdot c| = |\det(a, b, c)|.$$

Daher definiert man:

$$V(P(a, b, c)) = |\det(a, b, c)|.$$

Wie im \mathbb{R}^2 sieht man auch hier, dass ein Parallelelepiped durch einen Endomorphismus f wieder auf ein Parallelelepiped abgebildet wird, und dabei wird das Volumen mit dem Faktor $|\det f|$ multipliziert.

Im \mathbb{R}^n kann man analog zum \mathbb{R}^2 und \mathbb{R}^3 das von n Vektoren aufgespannte *Parallelotop* betrachten, und man nennt dann den Absolutbetrag der Determinante dieser Vektoren das (n -dimensionale) Volumen des Parallelotops.

Das Bild eines Parallelotops P unter einer Translation τ_v nennt man ebenfalls Parallelotop, und definiert sein Volumen als gleich dem von P . Daher gilt dann allgemein: Das Bild eines Parallelotops P unter einer affinen Abbildung $x \mapsto Ax + v$ mit einer quadratischen Matrix A ist wieder ein Parallelotop, und das Volumen von P wird dabei mit dem Faktor $|\det A|$ multipliziert.

Bemerkung zur Behandlung von Determinante und Vektorprodukt in der Schule:

Determinanten werden in der Schule normalerweise, wenn überhaupt, nur für 2×2 -Matrizen behandelt. Im Realgymnasium kommt zwar das Vektorprodukt vor, aber das Volumen eines Parallelepipedes wird höchstens am Rande behandelt. Gerade die oben besprochene geometrische Bedeutung von Determinante und Vektorprodukt scheint aber für die Behandlung dieses Themas in der Schule zu sprechen.

Andererseits ist klar, dass Determinanten in erster Linie theoretische Bedeutung haben. Für die praktische Lösung von linearen Gleichungssystemen oder die Inversion von Matrizen mit mehr als 2 oder 3 Zeilen sind sie nicht besonders gut geeignet.

11 Eigenvektoren und Eigenwerte

Die in diesem Kapitel behandelten Begriffe "Eigenvektor" und "Eigenwert" spielen an und für sich in der Linearen Algebra eine ganz wesentliche Rolle. Da sie in der Schule (derzeit) kaum vorkommen, werden sie hier nur in sehr knapper Form besprochen.

Sei f ein Endomorphismus des \mathbb{R}^n . Unter einem *Eigenvektor* von f versteht man einen Vektor $v \neq o$, der durch f auf ein skalares Vielfaches von sich abgebildet wird, d.h. wenn es ein $\lambda \in \mathbb{R}$ gibt, sodass

$$f(v) = \lambda v.$$

Der Skalar λ heißt dann (der zugehörige) *Eigenwert* von f . Für $\lambda > 0$ bleibt die Richtung eines Eigenvektors also bei Anwendung von f unverändert, für $\lambda < 0$ wird sie umgedreht. Wenn $\lambda = 0$ ein Eigenwert ist, dann sind die zugehörigen Eigenvektoren genau die vom Nullvektor verschiedenen Elemente des Kerns von f .

Beispiel: Sei f eine Drehung im \mathbb{R}^3 . Dann ist jeder erzeugende Vektor der Drehachse ein Eigenvektor, und der zugehörige Eigenwert ist gleich 1.

Ein Vektor $v \neq o$ heißt *Eigenvektor* der (quadratischen) Matrix A , wenn er Eigenvektor des zugehörigen Endomorphismus f_A ist, d.h. wenn es ein $\lambda \in \mathbb{R}$ gibt, sodass

$$Av = \lambda v,$$

und dieses λ heißt dann natürlich (zugehöriger) *Eigenwert* von A .

λ ist also genau dann ein Eigenwert von A , wenn es einen Vektor $v \neq o$ gibt, sodass $Av = \lambda v$, d.h. wenn das homogene lineare Gleichungssystem $(A - \lambda E)x = o$ eine nicht-triviale Lösung hat. Wie wir wissen, ist das genau dann der Fall, wenn

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

Die Eigenwerte von A sind also genau die Lösungen dieser *charakteristischen Gleichung* (in der Unbekannten λ). Für jeden Eigenwert λ_i erhält man die zugehörigen Eigenvektoren durch Lösen des Gleichungssystems $(A - \lambda_i E)x = o$. Der gesamte Lösungsraum dieses Gleichungssystems heißt der *Eigenraum* von A zum Eigenwert λ_i . Dieser Eigenraum besteht also aus allen zu λ_i gehörigen Eigenvektoren und dem Nullvektor.

Die charakteristische Gleichung einer $n \times n$ -Matrix ist immer ein Polynom n -ten Grades (ohne Beweis).

Beispiel: Sei $A = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix}$. Dann ist $A - \lambda E = \begin{pmatrix} 2 - \lambda & 6 \\ 1 & 3 - \lambda \end{pmatrix}$, und die charakteristische Gleichung von A lautet $(2 - \lambda)(3 - \lambda) - 6 = 0$ oder, ausmultipliziert:

$$\lambda^2 - 5\lambda = 0.$$

Diese Gleichung hat zwei Lösungen: $\lambda_1 = 0$ und $\lambda_2 = 5$. Die zugehörigen Eigenvektoren sind die Lösungen der folgenden beiden Gleichungssysteme:

Zu $\lambda_1 = 0$:

$$\begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ also z.B. } v_1 = \begin{pmatrix} -3 \\ 1 \end{pmatrix},$$

und zu $\lambda_2 = 5$:

$$\begin{pmatrix} -3 & 6 \\ 1 & -2 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ also etwa } v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Zur Probe können wir nachrechnen:

$$Av_1 = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0v_1 = \lambda_1 v_1,$$

$$Av_2 = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 10 \\ 5 \end{pmatrix} = 5v_2 = \lambda_2 v_2.$$

Eigenvektoren und Eigenwerte können dazu dienen, die wesentlichen Eigenschaften eines Endomorphismus besser zu verstehen. Das gilt auf Grund des folgenden Satzes insbesondere für symmetrische Matrizen. (Eine quadratische Matrix A heißt *symmetrisch*, wenn $A^T = A$ ist.)

Satz 52 *Zu jeder symmetrischen $n \times n$ -Matrix A gibt es eine Orthogonalsbasis B des \mathbb{R}^n , welche aus Eigenvektoren von A besteht. Die Matrix von f_A bezüglich B ist dann eine Diagonalmatrix, in deren Diagonalen die zugehörigen Eigenwerte von A stehen.*

Auf den etwas längeren Beweis des ersten Teiles dieses Satzes verzichten wir hier. Der zweite Teil ist dann aber leicht einzusehen: Sei $B = (b_1, \dots, b_n)$ eine (geordnete) ONB des \mathbb{R}^n , welche aus Eigenvektoren von A besteht, und $\lambda_1, \dots, \lambda_n$ seien die zugehörigen Eigenwerte. Dann gilt $Ab_i = \lambda_i b_i$ für alle $i \in \{1, \dots, n\}$. Die Koordinaten von $f_A(b_i)$ bezüglich B sind daher gleich $0, \dots, 0, \lambda_i, 0, \dots, 0$ (mit λ_i an der i -ten Stelle). Die Matrix von f_A bezüglich B sieht daher so aus:

$$\begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

□

Satz 53 *Jede symmetrische Matrix ist zu einer Diagonalmatrix ähnlich.*

Beweis: Sei $B = (b_1, \dots, b_n)$ eine ONB aus Eigenvektoren der symmetrischen Matrix A , und D sei die Diagonalmatrix mit den zugehörigen Eigenwerten. Wir können B als orthogonale Matrix auffassen und sehen nach Obigem (vgl. Abschnitt 8.2): $D = B^{-1}AB = B^TAB$. □

(Es handelt sich hier um eine besondere Art von Ähnlichkeit, da die Transformationsmatrix orthogonal ist. Man spricht in diesem Fall auch von *orthogonaler Ähnlichkeit*.)

Beispiel: Sei $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$.

$$\det(A - \lambda E) = \begin{vmatrix} 2 - \lambda & 1 & 1 \\ 1 & 2 - \lambda & 1 \\ 1 & 1 & 2 - \lambda \end{vmatrix} = -\lambda^3 + 6\lambda^2 - 9\lambda + 4 = 0$$

ist eine Gleichung dritten Grades mit den Wurzeln $\lambda_1 = \lambda_2 = 1$ und $\lambda_3 = 4$. (Die Zahl 1 ist eine doppelt zu zählende Wurzel: Wenn man $-\lambda^3 + 6\lambda^2 - 9\lambda + 4$ durch $\lambda - 1$ dividiert, erhält man $-\lambda^2 + 5\lambda - 4 = -(\lambda - 1)(\lambda - 4)$.)

Der Eigenraum zu $\lambda_1 = \lambda_2 = 1$ ist die Lösungsmenge des homogenen Gleichungssystems mit der Matrix $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. Durch Subtraktion der 1. Zeile von den anderen beiden Zeilen erhalten wir $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, woraus wir die folgenden Basisvektoren des Eigenraums gewinnen:

$$b'_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad b'_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Diese sind allerdings nicht orthogonal zueinander. Wir können sie aber mit dem Schmidt'schen Verfahren orthonormalisieren:

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right) \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}, \text{ also}$$

$$b_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad b_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.$$

Jetzt brauchen wir noch einen Eigenvektor zum dritten Eigenwert $\lambda_3 = 4$: Die

Matrix des entsprechenden homogenen Gleichungssystems ist $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$.

Mit dem Gauß'schen Eliminationsverfahren erhalten wir zunächst $b'_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$,

und durch Normierung ergibt das

$$b_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Dieser Vektor ist orthogonal zu b_1 und b_2 , und daher erhalten wir:

$$B = (b_1, b_2, b_3) = \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Zur Probe könnten wir nachrechnen, dass tatsächlich gilt:

$$B^T AB = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

1. Bemerkung: Zwei zu verschiedenen Eigenwerten gehörige Eigenvektoren einer symmetrischen Matrix sind immer orthogonal zueinander.

Beweis: Sei $Av = \lambda v$ und $Aw = \mu w$ mit $\lambda \neq \mu$. Fassen wir v und w als einspaltige Matrizen auf, so gilt wegen $A^T = A$:

$$\begin{aligned} 0 &= v^T Aw - v^T Aw = v^T Aw - (w^T Av)^T = v^T \mu w - (w^T \lambda v)^T = \\ &= v^T \mu w - v^T \lambda w = (\mu - \lambda) v^T w = (\mu - \lambda) v \cdot w. \end{aligned}$$

Wegen $\lambda \neq \mu$ folgt daraus $v \cdot w = 0$. \square

2. Bemerkung: In der Diagonalmatrix gemäß Satz 52 steht jeder Eigenwert so oft, wie seiner Vielfachheit als Wurzel der charakteristischen Gleichung entspricht (ohne Beweis).

12 Quadratische Formen

12.1 Definition

Matrizen beschreiben nicht nur lineare Abbildungen, sondern auch "quadratische", wie im Folgenden beschrieben wird.

Sei A eine symmetrische $n \times n$ -Matrix. Dann nennt man die Abbildung

$$q_A : \mathbb{R}^n \rightarrow \mathbb{R} : x \mapsto x^T Ax$$

die zu A gehörige *quadratische Form*. (Dabei fassen wir die Vektoren $x \in \mathbb{R}^n$ wieder als einspaltige Matrizen auf. Statt $x^T Ax$ könnten wir aber auch $x \cdot Ax$ schreiben.)

Ausführlicher geschrieben heißt das:

$$\begin{aligned} q_A(x) &= (x_1, \dots, x_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \\ &= x_1 \sum_{k=1}^n a_{1k} x_k + \dots + x_n \sum_{k=1}^n a_{nk} x_k = \\ &= \sum_{i,k} a_{ik} x_i x_k. \end{aligned}$$

Wegen der Symmetrie der Matrix A können wir auch schreiben:

$$q_A(x) = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{1 \leq i < k \leq n} a_{ik} x_i x_k.$$

Beispiele:

1. Sei $A = \begin{pmatrix} 3 & 4 \\ 4 & 5 \end{pmatrix}$. Dann ist $q_A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} 3 & 4 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3x_1^2 + 8x_1 x_2 + 5x_2^2$.
2. Mit $A = E$ erhalten wir $q_E(x) = x^T Ex = x \cdot x$.

Bemerkung: Wenn A eine nicht-symmetrische Matrix ist, so ist die Abbildung $x \mapsto x^T Ax$ ebenfalls eine quadratische Form. Sie entspricht nämlich der folgenden symmetrischen Matrix:

$$\tilde{A} := \frac{1}{2}(A + A^T).$$

Beweis:

$$2x^T Ax = \sum a_{ik} x_i x_k + \sum a_{ki} x_k x_i = \sum (a_{ik} x_i x_k + a_{ki} x_k x_i) = \sum (a_{ik} + a_{ik}^T) x_i x_k = \sum (2\tilde{a}_{ik} x_i x_k) = \sum (\tilde{a}_{ik} x_i x_k + \tilde{a}_{ki} x_k x_i) = 2x^T \tilde{A} x,$$

wobei die Summation jedesmal über alle i und $k \in \{1, \dots, n\}$ geht. \square

12.2 Koordinatentransformation

Sei B eine geordnete Basis des \mathbb{R}^n bzw. die entsprechende Matrix. Dann gilt für den Koordinatenvektor x' von x bezüglich B bekanntlich: $x = Bx'$. Setzen wir das in die Definition von q_A ein, so erhalten wir:

$$q_A(x) = (Bx')^T A (Bx') = x'^T B^T A B x' = x'^T A' x'$$

mit

$$A' = B^T A B.$$

Diese Matrix A' heißt die Matrix der betrachteten quadratischen Form bezüglich der Basis B .

Da A symmetrisch ist, gibt es nach Satz 53 eine orthogonale Matrix B , sodass $B^T A B$ eine Diagonalmatrix ist. Es gilt also:

Satz 54 *Zu jeder quadratischen Form gibt es eine Orthonormalbasis, bezüglich der sie sich durch eine Diagonalmatrix beschreiben lässt.*

Beispiel:

Sei $A = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$, also $q_A(x) = 3x_1^2 + 3x_2^2 - 2x_1x_2$.

Die charakteristische Gleichung lautet: $\begin{vmatrix} 3-\lambda & -1 \\ -1 & 3-\lambda \end{vmatrix} = 8-6\lambda+\lambda^2 = 0$. Sie hat die Wurzeln $\lambda_1 = 2$ und $\lambda_2 = 4$. Wir können leicht zugehörige normierte Eigenvektoren berechnen:

$b_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $b_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Mit $B = (b_1, b_2)$ gilt dann $B^T A B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$. Bezeichnen wir die Koordinaten von x bezüglich B mit x'_1, x'_2 , so gilt also: $q_A(x) = 2x'^2_1 + 4x'^2_2$. Betrachten wir z.B. konkret den Vektor $x = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$, so ist $x' = B^T x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 11 \\ -1 \end{pmatrix}$, und einerseits

$$q_A(x) = 3x_1^2 + 3x_2^2 - 2x_1x_2 = 75 + 108 - 60 = 123,$$

andererseits

$$q_A(x) = 2x'^2_1 + 4x'^2_2 = 121 + 2 = 123.$$

Bemerkung: Der letzte Satz besagt also, dass man bei jeder quadratischen Form durch eine geeignete (sogar orthogonale) Koordinatentransformation die "gemischten" Glieder eliminieren kann.

Bezug zum Schulstoff und zu anderen Vorlesungen:

Quadratische Formen kommen in der Schule hauptsächlich in Zusammenhang mit Kurven zweiter Ordnung (Ellipsen, Hyperbeln, Parabeln) vor. Allgemein versteht man unter einer *Kurve zweiter Ordnung* die Lösungsmenge einer quadratischen Gleichung in zwei Unbekannten, d.h. einer Gleichung der Form

$$q(x_1, x_2) + f(x_1, x_2) + c = 0,$$

wobei q eine quadratische Form, f eine Linearform und c eine Konstante $\in \mathbb{R}$ ist. Wenn $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ die (symmetrische) Matrix von q und $b = (b_1, b_2)$ die (einzelige) Matrix von f ist, dann sieht so eine Gleichung ausgeschrieben folgendermaßen aus:

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + b_1x_1 + b_2x_2 + c = 0.$$

In der Schule wird meist nur der Fall behandelt, dass die Matrix A Diagonalfom hat. Für den Fall, dass außerdem $b_1 = b_2 = 0$ ist, erhält man also

$$a_{11}x_1^2 + a_{22}x_2^2 + c = 0,$$

das ist z.B. für $a_{11}, a_{22} > 0$ und $c < 0$ die Gleichung einer Ellipse. Die Achsen des entsprechenden Koordinatensystems nennt man dann *Hauptachsen*, die Koordinatentransformation gemäß Satz 54 daher auch *Hauptachsentransformation*.

In analoger Weise werden im \mathbb{R}^3 *Flächen zweiter Ordnung* definiert. Beispiele solche Flächen sind Kugeln, Zylinder und Kegel. Im \mathbb{R}^n spricht man dann von *Hyperflächen zweiter Ordnung* oder *Quadriken*. Diese werden normalerweise in der Vorlesung "Geometrie" eingehend behandelt.

12.3 Definite quadratische Formen

Eine quadratische Form q_A heißt *positiv definit*, wenn $q_A(x) > 0$ für alle $x \neq 0$. Wir wissen schon, dass die dem inneren Produkt entsprechende quadratische Form q_E positiv definit ist. Welche anderen quadratischen Formen sind auch positiv definit?

Sei q_A eine quadratische Form und B eine ONB des \mathbb{R}^n , sodass $B^T AB$ eine Diagonalmatrix ist, mit den Eigenwerten $\lambda_1, \dots, \lambda_n$ von A in der Diagonalen. Dann gilt mit den Koordinaten x'_i bezüglich der Basis B :

$$q_A(x) = \lambda_1 x'_1^2 + \dots + \lambda_n x'_n^2.$$

Wir erkennen daraus:

Satz 55 Eine quadratische Form q_A ist genau dann positiv definit, wenn alle Eigenwerte der zugehörigen Matrix positiv sind.

(Zum Beweis: Wäre etwa $\lambda_1 \leq 0$, dann hätte die quadratische Form für $x'_1 = 1$ und $x'_2 = \dots = x'_n = 0$ den nicht-positiven Wert λ_1 .)

Ohne Beweis erwähnen wir noch ein zweites nützliches Kriterium für die positive Definitheit. Dazu benötigen wir folgenden Begriff: Sei A eine $n \times n$ -Matrix. Dann versteht man unter den *Hauptminoren* von A die Determinanten der folgenden Untermatrizen von A :

$$A_k := \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix} \quad \text{für } k \in \{1, \dots, n\}.$$

Bezeichnung:

$$\Delta_k := \det A_k = \begin{vmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{vmatrix}$$

Es gilt nun:

Satz 56 Eine quadratische Form ist genau dann positiv definit, wenn alle Hauptminoren der zugehörigen Matrix positiv sind.

Die quadratische Form des obigen Beispiels ist positiv definit. Das sieht man einerseits daran, dass beide Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = 4$ positiv sind. Andererseits folgt es daraus, dass die beiden Hauptminoren $\Delta_1 = 3$ und $\Delta_2 = 8$ positiv sind.

Eine quadratische Form q heißt *negativ definit*, wenn $q(x) < 0$ für alle $x \neq 0$. Offensichtlich ist q genau dann negativ definit, wenn alle Eigenwerte negativ sind. Bezuglich der Hauptminoren gilt aber

Satz 57 Eine quadratische Form ist genau dann negativ definit, wenn alle Hauptminoren der zugehörigen Matrix abwechselndes Vorzeichen haben, beginnend mit einem negativen Vorzeichen, d.h. $\text{sign } \Delta_k = (-1)^k$.

Beweis: Dieser Satz folgt leicht aus dem vorhergehenden:

$x^T Ax < 0$ ist gleichbedeutend mit $x^T(-A)x > 0$. Daher ist A genau dann negativ definit, wenn $-A$ positiv definit ist. Wegen $\det(-A_k) = (-1)^k \det A_k > 0$ heißt das $\text{sign } \Delta_k = (-1)^k$. \square

Beispiel: Sei $A = \begin{pmatrix} -1 & 1 \\ 1 & -3 \end{pmatrix}$. Die Hauptminoren sind $\Delta_1 = -1$ und $\Delta_2 = 2$, also gilt $\text{sign } \Delta_k = (-1)^k$.

Die charakteristische Gleichung ist $\begin{vmatrix} -1 - \lambda & 1 \\ 1 & -3 - \lambda \end{vmatrix} = 2 + 4\lambda + \lambda^2 = 0$. Sie liefert die negativen Eigenwerte $\lambda_1 = -2 + \sqrt{2}$ und $\lambda_2 = -2 - \sqrt{2}$.

Ein quadratische Form q heißt *positiv semidefinit*, wenn $q(x) \geq 0$ für alle $x \in \mathbb{R}^n$. Das ist gleichbedeutend damit, dass alle Eigenwerte ≥ 0 sind. Dabei kann also $q(x) = 0$ sein für gewisse $x \neq 0$. Analog heißt q *negativ semidefinit*, wenn $q(x) \leq 0$ für alle x . Wenn q weder positiv noch negativ semidefinit ist, dann heißt q *indefinit*. Dann gibt es also $x, y \neq 0$, sodass $q(x) > 0$ und $q(y) < 0$.

Beispiele:

1. Sei $A = \begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix}$. Die Eigenwerte von A sind 0 und 10. Daher ist q_A positiv semidefinit. Es ist z.B. $q_A \begin{pmatrix} 3 \\ -1 \end{pmatrix} = 0$.

In diesem Beispiel kann man übrigens die positive Semidefinitheit leicht direkt einsehen: $q_A(x) = x_1^2 + 6x_1x_2 + 9x_2^2 = (x_1 + 3x_2)^2 \geq 0$ für alle x , und $q_A(x) = 0$ genau dann, wenn $x_1 = -3x_2$.

2. Sei $B = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$. Die Eigenwerte von B sind -2 und 4 . Daher ist q_B indefinit. Z.B. ist $q_B \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1$ und $q_B \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -4$. (Aus diesen beiden Werten von q_B folgt natürlich unmittelbar die Indefinitheit.)

Bemerkung: Wenn eine quadratische Form q_A positiv oder negativ (semi-)definit ist, dann nennen wir auch die zugehörige Matrix A positiv bzw. negativ (semi-)definit.

Bezug zum Schulstoff und zu anderen Vorlesungen:

Definite quadratische Formen treten insbesondere in Zusammenhang mit der Untersuchung von Extremwerten von Funktionen in mehreren Variablen auf. Zum Beispiel liegt ein lokales Minimum vor, wenn die durch die zweiten partiellen Ableitungen bestimmte quadratische Form positiv definit ist. (Im Falle einer Variablen heißt das, dass die zweite Ableitung positiv ist.) Diese Thematik wird normalerweise in der Vorlesung "Analysis II" genauer behandelt.

Da Extremwerte von Funktionen einer Variablen in der Schule ein wesentliches Thema darstellen, sollte der Lehrer wohl auch eine Vorstellung vom mehrdimensionalen Analagon haben. Dies wohl auch deshalb, weil die Bestimmung von Extremwerten ("Optimierung") in den Anwendungen der Mathematik eine bedeutende Rolle spielt (z.B. Wirtschaftsmathematik).

Literaturverzeichnis

- [1] Benz, Walter: Ebene Geometrie: Einführung in Theorie und Anwendung. Spektrum Akad. Verlag 1997.
- [2] Bürger, H. et al.: Mathematik Oberstufe, Bd. 1 - 4. Hölder-Pichler-Tempsky, Wien 1989-1992.
- [3] Gabriel, Peter: Matrizen, Geometrie, Lineare Algebra. Birkhäuser 1996.
- [4] Klingenberg, Wilhelm: Lineare Algebra und Geometrie. 3. Aufl., Springer 1992.
- [5] Koecher, Max: Lineare Algebra und analytische Geometrie. 4. Aufl., Springer 1997.
- [6] Linhart, Johann: Diskrete Mathematik. Abakus-Verlag, Salzburg 1996.
- [7] Padberg, F. und Kötting, H.: Lineare Algebra. Eine elementare Einführung. B.I.-Hochschultaschenbuch (Bd. 649) 1991.
- [8] The MacTutor History of Mathematics archive, University of St. Andrews, Scotland, <http://www-history.mcs.st-and.ac.uk/~history/>

Stichwortverzeichnis

- Ähnlichkeit
 - orthogonale, 105
 - von Matrizen, 83
- ähnliche Matrizen, 83
- äußeres Produkt, 97

- Abbildung
 - affine, 20
 - lineare, 15
- abhängig
 - linear, 39
- Abstand
 - eines Punktes von einem Teilraum, 76
 - eines Punktes von einer Hyperebene, 77
 - zweier Punkte, 65
- affine Abbildung, 20
- affiner Teilraum, 38
- allgemeine Lösung, 31
- alternierend, 88
- Anfangspunkt, 7
- antisymmetrisch, 88
- Array
 - zweidimensionales, 13
- Array, 10
- Assoziativgesetz, 11
- aufgespannter Teilraum, 38

- Basis, 38
- Bewegung, 80
 - eigentliche, 102
- bijektiv, 57
- Bild, 50
- Bildraum, 18, 50
- Bilinearform, 63
- Blockmatrizen, 52

- Cauchy-Schwarz'sche Ungleichung, 63
- Computergeometrie, 20
- Cramer'sche Regel, 93

- Defekt, 50
- Determinante, 86
- Determinantenentwicklungssatz, 91
- Determinantenmultiplikationssatz, 90
- Diagonalmatrix, 19
- Differenz
 - von Vektoren, 6
- Dimension, 42, 44
- Drehung, 79

- Ebene, 49
- eigentliche Bewegungen, 102
- eindeutige Lösung
 - eines linearen Gleichungssystems, 48
- Einheitsmatrix, 19
- Einheitsquadrat, 18
- elementare Umformungen, 25
- Eliminationsverfahren, 32
- Endomorphismus, 82
- Endpunkt, 7
- Entwicklung
 - einer Determinante, 86, 91
- erweiterte Matrix, 48
- erzeugter Teilraum, 38
- euklidischer Raum, 6

- Fixpunkt, 10
- Flächeninhalt
 - eines Parallelogramms, 96
- Flächen zweiter Ordnung, 110

- Gauß'sches Eliminationsverfahren, 32

- Gerade, 49
Gleichungssystem
 lineares, 13, 25
Gruppe, 11
- Halbdiagonalform, 27
Hauptachsen, 110
Hauptachsentransformation, 110
Hauptdiagonale, 19
Hauptminoren, 111
Hesse'sche Normalform, 77
homogen, 32
Homothetie, 8
Hülle
 lineare, 38
Hyperebene, 49
- identische Abbildung, 19
inhomogen, 32
inneres Produkt, 61
invariant, 78
Invarianz
 der Norm, 79
 des inneren Produkts, 79
inverse Matrix, 58
inverses Element, 11
invertierbare Matrix, 59
- Kern, 38
Komplement
 orthogonales, 73
kongruent, 80
Kongruenztransformation, 80
Koordinaten, 6, 39
Koordinatentransformation, 80
Koordinatenvektor, 80
Kreuzprodukt, 97
Kronecker-Symbol, 68
Kurve zweiter Ordnung, 110
- Länge
 eines Vektors, 63
- linear abhängig, 39
linear unabhängig, 39, 40
lineare Abbildung, 15
lineare Hülle, 38
linearer Teilraum, 37
lineares Gleichungssystem, 13, 25
Linearform, 17
Linearkombination, 8
Lösung
 allgemeine, 31
 eines linearen Gleichungssystems, 25
 partikuläre, 32
 triviale, 33
Lösungsmenge, 14, 25
Lotfußpunkt, 70, 76
- Matrix, 13
 beziiglich einer Basis, 82
 erweiterte, 48
Matrizenmultiplikation, 21
Mittelpunkt
 zweier Punkte, 10
Multilinearform, 88
Multiplikation
 von Matrizen, 21
- neutrales Element, 11
Norm
 eines Vektors, 63
normal, 67
Normalprojektion, 20, 70
Normalvektor, 74
normieren, 66
normiert, 66
Nullmatrix, 21
Nullraum, 38
Nullvektor, 6
numerische Fehler, 36
numerische Lösung, 36
- orientierter Winkel, 95

- Orientierung, 95, 101
- orientierungstreu, 96, 102
- orthogonal
 - Vektoren, 67
- Orthogonalbasis, 67
- orthogonale Ähnlichkeit, 105
- orthogonale Abbildung, 77
- orthogonale Matrix, 77
- orthogonale Projektion, 70
- orthogonales Komplement, 73
- Orthogonalsystem, 67
- Orthonormalbasis, 68
- Orthonormalisierungsverfahren, 71
- Orthonormalsystem, 68
- Ortsvektor, 7
- Parallelepiped, 102
- Parallelogramm, 7, 18, 96
- Parallelotop, 102
- Parallelverschiebung, 7
- Parameter, 32
- Parameterdarstellung, 32
- partikuläre Lösung, 32
- Pivot, 36
- Pivotsuche
 - vollständige, 36
 - positiv definit, 110
 - positiv orientiert, 95, 102
- Produkt
 - äußeres, 97
 - inneres, 61
 - skalares, 61
- Projektion
 - orthogonale, 70
- Punkt, 49
- Punkte, 6
- Punktspiegelung, 10
- Pythagoras, 67
- quadratische Form, 108
- Rang
- einer linearen Abbildung, 50
- einer Matrix, 27
- Raum
 - n-dimensional, 6
- Raum
 - euklidischer, 6
 - reguläre Matrix, 59
- Satz von Pythagoras, 67
- Schiebung, 7
- Schmidt'sches Orthonormalisierungsverfahren, 71
- Schnittpunkt, 53
- senkrecht, 67
- singuläre Matrix, 59
- Skalar, 8
- skalares Produkt, 61
- Spalten, 13
- Spaltenpivotsuche, 36
- Spaltenschreibweise, 7
- Spaltenumformungen, 57
- Spaltenvektoren, 13
- Spiegelung, 75
 - am Ursprung, 8
 - an einem Punkt, 10
- Stauchung, 8
- Streckung, 8
- Streckungsfaktor, 10
- Summe
 - von Vektoren, 6
- symmetrisch, 105
- Teilraum
 - affiner, 38
 - linearer, 37
- Translation, 7
- transponierte Matrix, 54
- triviale Lösung, 33
- Umformungen
 - elementare, 25
- Umkehrabbildung, 57

- unabhängig
 - linear, 39, 40
- Unbekannte, 13, 25
- Untervektorraum, 37
- Ursprung, 6
- Vektoren, 6
- Vektorprodukt, 97
- Vektorraum, 9
- Verknüpfung, 11
- Winkel
 - orientierter, 95
 - zwischen zwei Vektoren, 64
- Zeilen, 13
- Zeilenvektoren, 13
- Zentrum
 - einer Homothetie, 10