Lineare Algebra und Analytische Geometrie

Dr. V. Drumm

Prof. Dr. W. Weil

Institut für Algebra und Geometrie – Universität Karlsruhe(TH)

Vorwort

Die Vorlesung "Lineare Algebra und Analytische Geometrie" erstreckt sich über zwei Semester. Sie ist eine Pflichtvorlesung für alle, die Mathematik und Informatik studieren, wird aber auch den Studierenden der Physik empfohlen.

Der Stoff des ersten Teils der Vorlesung ist für alle Fachrichtungen im wesentlichen derselbe. Im zweiten Teil wird die Vorlesung auch inhaltlich getrennt, da dann der Studienplan für die Fachrichtung Informatik nur noch zwei Semesterwochenstunden vorsieht.

Das vorliegende Skriptum, das aus Vorlesungen hervorgegangen ist, die wir mehrfach in Karlsruhe gehalten haben, berücksichtigt diese Trennung. Es enthält im wesentlichen den Inhalt der Vorlesung für die Fachrichtung Mathematik und darüber hinaus noch einige Ergänzungen, jedoch können ohne Schwierigkeiten Teile davon für die verkürzte Vorlesung Mathematik für die Fachrichtung Informatik ausgewählt werden.

Das Skriptum soll allen Hörerinnen und Hörern das Mitverfolgen der Vorlesung erleichtern, ist aber kein Ersatz für die Vorlesung. Trotz seines Umfanges sind manche Beweise recht knapp gefaßt und einige Bemerkungen überhaupt nicht bewiesen. Weiterhin soll und kann das Skriptum auch kein Lehrbuch ersetzen. Wir haben deshalb am Ende eine umfangreiche Liste deutschsprachiger Lehrbücher aufgeführt.

Im Unterschied zu früheren Fassungen werden Vektoren, wie inzwischen allgemein üblich, jetzt mit kleinen lateinischen Buchstaben bezeichnet. Wir hoffen, daß dadurch die Lesbarkeit nicht erschwert wird. Im übrigen haben wir die alte Rechtschreibung beibehalten, aber auch das sollte das Verständnis nicht beeinträchtigen.

Karlsruhe, im Oktober 2007 V. Drumm, W. Weil

Inhaltsverzeichnis

Einleitung

Die lineare Algebra bildet neben der Analysis die Grundlage weiter Teile der Mathematik und damit auch der Mathematikausbildung. Sie behandelt die Theorie der Vektorräume und linearen Abbildungen sowie deren Anwendung auf lineare Gleichungssysteme und Eigenwertprobleme.

Systeme von linearen Gleichungen und Methoden zu deren Lösung waren schon den Chinesen vor über 1500 Jahren bekannt. Die Probleme bei der Auflösung linearer und nichtlinearer Gleichungen führten zur Entwicklung der Algebra (ca. $1000 - 1600$).

Die Beschreibung von Punkten durch Koordinaten und von Geraden und Ebenen durch lineare Gleichungen erwies sich als sehr nützlich für die Behandlung geometrischer Probleme. Diese "analytische Geometrie" hat zu der Einführung von Vektoren und der Entstehung der Vektorraumtheorie geführt. Bis zur Mitte des 20. Jahrhunderts dienten die Vektorräume in der Grundvorlesung hauptsächlich als formaler Rahmen, in dem geometrische Probleme behandelt wurden.

Das hat sich inzwischen grundlegend geändert. Zum einen verlangt der zunehmende Abstraktionsprozeß in der Mathematik, aber auch in Anwendungsgebieten wie Informatik, die Behandlung abstrakter Vektorräume, zum andern sind lineare Gleichungs– und Ungleichungssysteme sowie Eigenwertprobleme für die Anwendungen inzwischen von so grundlegender Bedeutung, daß ihrer Behandlung in der mathematischen Grundausbildung besonderes Gewicht zukommt.

Wir werden deshalb im folgenden zunächst hauptsächlich die lineare Algebra behandeln, geometrische Sprechweisen und Begriffe aber zur Veranschaulichung benutzen. Die affine und euklidische Geometrie wird dann in Kapitel 6 ausführlicher dargestellt.

Vorbemerkungen über Mengen, Abbildungen, Relationen

Die folgenden Bemerkungen umfassen eine Sammlung von Begriffen und Aussagen, die zur mathematischen Allgemeinbildung gehören und teilweise von der Schule her bekannt sein sollten. Diese Begriffe sind grundlegend für die weiteren Betrachtungen, sie gehören aber nicht zum eigentlichen Stoffgebiet der linearen Algebra.

1. Logische Symbole

Aufgabe der Mathematik ist es, abstrakte Srukturen zu entwickeln und Aussagen über diese Strukturen herzuleiten. Wird dazu die Umgangssprache benutzt, so kann das zu Unklarheiten und Mehrdeutigkeiten führen. Um mathematische Aussagen und Schlüsse einwandfrei formulieren zu können, wurde eine formalisierte Sprache entwickelt, die nur Symbole benutzt; ihre Untersuchung ist Gegenstand der *Mathematischen Logik.*

Da sie eine präzise und knappe Darstellung erlaubt, fand die formalisierte Schreibweise Eingang in viele Lehrbücher und zum Teil sogar in die Schulmathematik. Insbesondere für Informatiker ist eine formale Beschreibung mathematischer Vorgänge von Bedeutung.

Andererseits hat Mathematik auch viel mit Intuition, Fantasie und Vorstellungskraft zu tun, die nur durch Verwendung der Umgangssprache entstehen können. Die formalisierte Schreibweise erschwert das Verständnis eher. Sie ist deshalb aus den neueren Lehrbüchern wieder weitgehend verschwunden und wird auch im folgenden meistens nicht benutzt.

Damit man sich aber in der Literatur zurechtfindet, sei hier eine Liste der gebräuchlichsten Symbole (Abkürzungen) aufgeführt:

$\neg$ Negation (einer Aussage)

$\wedge$	Konjunktion (Verbindung zweier Aussagen durch *und*); häufig wird $\wedge$ durch ein Komma ersetzt
$\vee$	Disjunktion (Verbindung zweier Aussagen durch *oder*)
$\Longrightarrow$	Implikation (aus einer Aussage folgt eine andere); wenn ..., dann...; daraus folgt
$\Longleftrightarrow$	Äquivalenz (zweier Aussagen); ...genau dann, wenn...
$\forall$	Allquantor (für alle ...)
$\exists$	Existenzquantor (es gibt ein ...)
$\in$	Element von
$\notin$	nicht Element von

Die Symbole $:=$ und $:\Longleftrightarrow$ (bzw. $=:$ und $\Longleftrightarrow:$) werden benutzt, wenn Größen oder Begriffe, die auf der Seite des Doppelpunktes stehen, durch die andere Seite erklärt werden sollen.

Wer sich über die Mathematische Logik (und ihre Anwendungen in der Informatik) informieren möchte, dem sei das einführende Buch

Böhme, G. : Einstieg in die Mathematische Logik, Hanser, München u.a. 1981

empfohlen.

2. Mengen

Beim Aufbau mathematischer Strukturen werden aus bekannten Begriffen neue abgeleitet. Verfolgt man diesen Aufbau zurück, so stößt man zwangsläufig auf Grundbegriffe, die mathematisch nicht weiter erklärt werden können, wie z.B. Punkte in einem axiomatischen Aufbau der Geometrie. Man kann solche Begriffe nur dadurch festlegen, daß man den Umgang mit ihnen durch Gesetze (*Axiome*) regelt.

Ein für die gesamte Mathematik grundlegender Begriff ist der der Menge. Der Mengenbegriff wird hier nicht axiomatisch eingeführt, es genügt uns vielmehr der "naive Standpunkt" der Mengenlehre, der auf Cantor (1845 – 1918) zurückgeht:

Eine Menge ist eine Zusammenfassung von "Objekten", den Elementen der Menge.

Eine Menge ist also festgelegt, wenn ihre Elemente festgelegt sind; damit muß für jedes "Objekt" feststehen, ob es zu der Menge gehört oder nicht; ob man das auch entscheiden kann oder nicht, ist ein anderes Problem.

Zwei Mengen A und B sind somit *gleich*, wenn sie die gleichen Elemente besitzen.

Um Widersprüche zu vermeiden, wird festgelegt, daß eine Menge nicht sich selbst als Element enthalten darf. Mehr über den axiomatischen Aufbau der Mengenlehre und dabei mögliche Widersprüche findet man in dem elementaren Buch

Halmos, P.R. : Naive Mengenlehre, Vandenhoek & Ruprecht, Göttingen 1976.

Beispiele.

$\mathbb{N}$ Menge der natürlichen Zahlen.

$\mathbb{N}_0$ Menge der natürlichen Zahlen und der Null.

$\mathbb{Z}$ Menge der ganzen Zahlen.

$\mathbb{Q}$ Menge der rationalen Zahlen.

$\mathbb{R}$ Menge der reellen Zahlen.

Die Beschreibung von Mengen erfolgt dadurch, daß ihre Elemente aufgezählt oder durch eine charakteristische Eigenschaft festgelegt werden. Die Aufzählung kann dabei (besonders bei Mengen mit unendlich vielen Elementen) auch in symbolischer Form vorgenommen werden.

Beispiele.

$\mathbb{N} = \{1,2,3,...\}$,

$\mathbb{N}_0 = \{0,1,2,3,...\}$,

$\mathbb{Z} = \{0,1,-1,2,-2,...\}$ oder $\mathbb{Z} = \{...,-2,-1,0,1,2,...\}$,

$\mathbb{Q} = \{x \mid x = \frac{p}{q},\ p,q \in \mathbb{Z},\ q \neq 0\}$.

Wir verwenden immer geschweifte Klammern $\{...\}$ zur Angabe von Mengen.

Die drei Punkte "..." sollen bedeuten, daß die Folge der aufgeführten Elemente in offensichtlicher Weise fortgesetzt werden soll; das setzt voraus, daß die Fortsetzung wirklich offensichtlich ist. Wird eine Menge durch eine Eigenschaft ihrer Elemente angegeben, so werden die Elemente und die kennzeichnende Eigenschaft (bzw. die Eigenschaften) durch einen senkrechten Strich getrennt, wie wir es oben bei der Beschreibung von $\mathbb{Q}$ schon getan haben. Nicht alle Mengen können durch Aufzählen ihrer Elemente angegeben werden; das ist z.B. bei $\mathbb{R}$ nicht möglich.

Aus mathematischen Gründen benötigt man auch eine Menge, die keine Elemente besitzt, also "leer" ist; es gibt nur eine solche Menge.

Definition. $\emptyset$ sei die Menge, die keine Elemente hat. Sie heißt die *leere Menge.*

Bezeichnung. Für eine Menge A sei $|A|$ die Anzahl der Elemente von A. Ist A endlich, so ist $|A| \in \mathbb{N}_0$, ist A unendlich, so schreiben wir $|A| = \infty$.

Ein weiterer, viel benutzter Begriff ist der der Teilmenge.

Definition. Eine Menge B heißt *Teilmenge* einer Menge A, wenn alle Elemente von B auch Elemente von A sind. Schreibweise : $B \subset A$ oder $A \supset B$.

Wir wollen diese Definition zur Übung auch rein formal schreiben:

$$B \subset A \;:\Longleftrightarrow\; [x \in B \implies x \in A].$$

Beispiele. $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. Für jede Menge A gilt $A \subset A$ und $\emptyset \subset A$.

Die Gesamtheit aller Teilmengen einer Menge A bildet wieder eine Menge, die *Potenzmenge* $\mathcal{P}(A)$ von A :

$$\mathcal{P}(A) := \{B \mid B \subset A\}$$

$\mathcal{P}(A)$ und jede Teilmenge $\mathcal{M} \subset \mathcal{P}(A)$ ist eine Menge von Mengen. Um das besser ausdrücken zu können, spricht man auch von einem *Mengensystem* $\mathcal{M}$. Man be-

achte aber, daß aufgrund unserer Vereinbarung, daß eine Menge nicht sich selbst als Element enthalten darf, die "Menge aller Mengen" als System $\mathcal{M}$ nicht zulässig ist.

3. Mengenoperationen

Definition. Es seien A,B Mengen. Dann heißt

$A \cup B := \{x \mid x \in A \text{ oder } x \in B\}$ die *Vereinigung* von A und B,

$A \cap B := \{x \mid x \in A \text{ und } x \in B\}$ der *Durchschnitt* von A und B,

$A \setminus B := \{x \mid x \in A, x \notin B\}$ die *Differenz* von A und B,

$A \Delta B := (A \setminus B) \cup (B \setminus A)$ die *symmetrische Differenz* von A und B.

Ist $B \subset A$ und ist die Grundmenge A fest vorgegeben, so schreibt man für $A \setminus B$ auch B^c und nennt das das *Komplement von B in A*.

Bemerkung. Einige einfache Aussagen über Mengen lassen sich unmittelbar verifizieren. So gelten für alle Mengen A und B: **(a)** $A = B \iff A \subset B, B \subset A$.

(b) $B \subset A \iff A \cup B = A \iff A \cap B = B \iff A \Delta B = A \setminus B$.

(c) $A \cup A = A \cap A = A$, $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$.

Die etwas weniger trivialen Rechenregeln fassen wir in dem folgenden Satz zusammen.

Satz 1. *Es seien A,B,C Mengen. Dann gilt:*

(a) $(A \cup B) \cup C = A \cup (B \cup C)$,
$(A \cap B) \cap C = A \cap (B \cap C)$, *("Assoziativgesetze")*

(b) $A \cup B = B \cup A$,
$A \cap B = B \cap A$, *("Kommutativgesetze")*

(c) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, *("Distributivgesetze")*

(d) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$,
$A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$. *("de Morgan–Regeln")*

Beweis. Ein beliebtes Beweisprinzip für Mengengleichungen ist auf der vorangehenden Bemerkung (a) aufgebaut. Man zeigt zunächst, daß die Menge links Teilmenge der Menge rechts ist, dann umgekehrt, daß die Menge rechts Teilmenge der Menge links ist. Wir wollen das Prinzip jeweils an der ersten Gleichung von (c) und von (d) deutlich machen. Die restlichen Beweise werden als Übungsaufgabe empfohlen.

Beh. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

" $\subset$ ": Sei $x \in A \cup (B \cap C) \Rightarrow x \in A$ oder $x \in (B \cap C)$.

1. Fall: $x \in A \Rightarrow x \in A \cup B$ und $x \in A \cup C \Rightarrow x \in (A \cup B) \cap (A \cup C)$.

2. Fall: $x \in B \cap C \Rightarrow x \in B$ und $x \in C \Rightarrow x \in A \cup B$ und $x \in A \cup C \Rightarrow x \in (A \cup B) \cap (A \cup C)$.

" $\supset$ ": Sei $x \in (A \cup B) \cap (A \cup C) \Rightarrow x \in A \cup B$ und $x \in A \cup C \Rightarrow (x \in A$ oder $x \in B)$ und $(x \in A$ oder $x \in C) \Rightarrow x \in A$ oder $(x \in B$ und $x \in C) \Rightarrow x \in A$ oder $x \in B \cap C \Rightarrow x \in A \cup (B \cap C)$. ∎

Beh. $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.

" $\subset$ ": Sei $x \in A \setminus (B \cup C) \Rightarrow x \in A, x \notin B \cup C \Rightarrow x \in A, x \notin B, x \notin C \Rightarrow x \in A \setminus B$, $x \in A \setminus C \Rightarrow x \in (A \setminus B) \cap (A \setminus C)$.

" $\supset$ ": Sei $x \in (A \setminus B) \cap (A \setminus C) \Rightarrow x \in A \setminus B, x \in A \setminus C \Rightarrow x \in A, x \notin B, x \notin C \Rightarrow$ $x \in A, x \notin B \cup C \Rightarrow x \in A \setminus (B \cup C)$. ∎

Es ist offensichtlich, wie sich die Aussagen von Satz 1 auf endlich viele Mengen verallgemeinern lassen. Wir wollen nun aber zeigen, daß man sogar ganze Mengensysteme zulassen kann.

Definition. Es sei $\mathcal{M}$ ein nichtleeres Mengensystem. Dann sei

$$\bigcup_{B \in \mathcal{M}} B := \{x \mid \text{Es gibt ein } B \in \mathcal{M} \text{ mit } x \in B\},$$

$$\bigcap_{B \in \mathcal{M}} B := \{x \mid \text{Für alle } B \in \mathcal{M} \text{ gilt } x \in B\}.$$

Ist $\mathcal{M} = \{B_1,\ldots,B_n\}$, so schreibt man

$$\bigcup_{i=1}^{n} B_i \ , \quad \bigcap_{i=1}^{n} B_i .$$

Ist $\mathcal{M} = \{B_1,B_2,\ldots\}$, so schreibt man formal

$$\bigcup_{i=1}^{\infty} B_i \ , \quad \bigcap_{i=1}^{\infty} B_i .$$

Entsprechend vereinfachte Schreibweisen werden wir später häufig benutzen; z.B. bezeichnet im folgenden Satz

$$\bigcup_{B\in\mathcal{M}} (A \cup B)$$

die Vereinigung über das Mengensystem $\mathcal{M}' = \{A \cup B \mid B \in \mathcal{M}\}$, usw.

Satz 2. *Es seien $\mathcal{M}$ ein nichtleeres Mengensystem und A eine Menge. Dann gilt:*

(a) $$A \cup \Big(\bigcup_{B\in\mathcal{M}} B\Big) = \bigcup_{B\in\mathcal{M}} (A \cup B),$$

$$A \cap \Big(\bigcap_{B\in\mathcal{M}} B\Big) = \bigcap_{B\in\mathcal{M}} (A \cap B),$$

(b) $$A \cup \Big(\bigcap_{B\in\mathcal{M}} B\Big) = \bigcap_{B\in\mathcal{M}} (A \cup B),$$

$$A \cap \Big(\bigcup_{B\in\mathcal{M}} B\Big) = \bigcup_{B\in\mathcal{M}} (A \cap B),$$

(c) $$A \setminus \Big(\bigcup_{B\in\mathcal{M}} B\Big) = \bigcap_{B\in\mathcal{M}} (A \setminus B),$$

$$A \setminus \Big(\bigcap_{B\in\mathcal{M}} B\Big) = \bigcup_{B\in\mathcal{M}} (A \setminus B).$$

Beweis. Da alle Aussagen nach der gleichen Methode bewiesen werden können, wollen wir uns auf je eine Aussage beschränken. Wir werden hier allerdings ein etwas kürzeres Beweisprinzip anwenden, indem wir die Aussage, daß x Element der linken Menge ist, äquivalent umformen.

Beh. $A \cap \left(\bigcap_{B \in \mathcal{M}} B \right) = \bigcap_{B \in \mathcal{M}} (A \cap B)$.

Es gilt:

$$x \in A \cap \left(\bigcap_{B \in \mathcal{M}} B \right) \iff x \in A, x \in \bigcap_{B \in \mathcal{M}} B \iff x \in A, x \in B \text{ für alle } B \in \mathcal{M}$$

$$\iff x \in A \cap B \text{ für alle } B \in \mathcal{M} \iff x \in \bigcap_{B \in \mathcal{M}} (A \cap B). \qquad \blacksquare$$

Beh. $A \cap \left(\bigcup_{B \in \mathcal{M}} B \right) = \bigcup_{B \in \mathcal{M}} (A \cap B)$.

Es gilt:

$$x \in A \cap \left(\bigcup_{B \in \mathcal{M}} B \right) \iff x \in A, x \in \bigcup_{B \in \mathcal{M}} B \iff x \in A, x \in B \text{ für mindestens ein}$$

$$B \in \mathcal{M} \iff x \in A \cap B \text{ für mindestens ein } B \in \mathcal{M} \iff x \in \bigcup_{B \in \mathcal{M}} (A \cap B). \qquad \blacksquare$$

Beh. $A \setminus \left(\bigcap_{B \in \mathcal{M}} B \right) = \bigcup_{B \in \mathcal{M}} (A \setminus B)$.

Es gilt:

$$x \in A \setminus \left(\bigcap_{B \in \mathcal{M}} B \right) \iff x \in A, x \notin \bigcap_{B \in \mathcal{M}} B \iff x \in A, x \notin B \text{ für mindestens ein}$$

$$B \in \mathcal{M} \iff x \in A \setminus B \text{ für mindestens ein } B \in \mathcal{M} \iff x \in \bigcup_{B \in \mathcal{M}} (A \setminus B). \qquad \blacksquare$$

Für die Anzahl der Elemente einer Vereinigungsmenge gibt es eine nützliche Formel, die insbesondere in der Kombinatorik und der Stochastik benutzt wird. Wir benutzen sie in § 1.6 beim Beweis des Chinesischen Restsatzes.

Satz 3. *Es seien $A_1,\dots A_k$ endliche Mengen. Dann gilt :*

$$\Big|\bigcup_{i=1}^{k} A_i\Big| = \sum_{i=1}^{k} |A_i| - \sum_{1\le i<j\le k} |A_i\cap A_j|$$

$$+ - \cdots + (-1)^k \sum_{1\le i_1<\dots<i_{k-1}\le k} |A_{i_1}\cap\cdots\cap A_{i_{k-1}}| + (-1)^{k+1}\,|A_1\cap\cdots\cap A_k|\,.$$

Bemerkung. Das Summenzeichen Σ ist die Abkürzung für mehrfache Additionen. Im ersten Fall wird über alle $i \in \{1,\dots,k\}$ summiert, im zweiten Fall über alle $i \in \{1,\dots,k-1\}$ und für gewähltes i über alle $j \in \{i+1,\dots,k\}$, usw.

Beweis. Wir führen den Beweis durch vollständige Induktion nach k. Induktionsanfang: Für $k = 1$ steht links und rechts jeweils nur $|A_1|$ und die Behauptung ist trivial. Für $k = 2$ gilt :

$$|A_1 \cup A_2| = |A_1 \cup (A_2 \setminus (A_1 \cap A_2))| = |A_1| + |A_2 \setminus (A_1 \cap A_2)|$$
$$= |A_1| + |A_2| - |A_1 \cap A_2|.$$

Induktionsvoraussetzung : Die Behauptung sei richtig für k.

Induktionsschluß von k auf $k+1$: Es ist

$$\bigcup_{i=1}^{k+1} A_i = \Big(\bigcup_{i=1}^{k} A_i\Big) \cup A_{k+1}\,.$$

Nach dem eben bewiesenen Induktionsanfang, angewendet auf die zwei Mengen $A_1 \cup \cdots \cup A_k$ und A_{k+1}, erhalten wir

$$(*) \quad \Big|\bigcup_{i=1}^{k+1} A_i\Big| = \Big|\bigcup_{i=1}^{k} A_i\Big| + |A_{k+1}| - \Big|\Big(\bigcup_{i=1}^{k} A_i\Big) \cap A_{k+1}\Big|.$$

Nach Induktionsvoraussetzung gilt

$$\Big|\bigcup_{i=1}^{k} A_i\Big| = \sum_{i=1}^{k} |A_i| - \sum_{1 \le i < j \le k} |A_i \cap A_j| + - \cdots + (-1)^{k+1} |A_1 \cap \cdots \cap A_k|$$

und

$$\Big|\Big(\bigcup_{i=1}^{k} A_i\Big) \cap A_{k+1}\Big| = \Big|\bigcup_{i=1}^{k} (A_i \cap A_{k+1})\Big| = \sum_{i=1}^{k} |A_i \cap A_{k+1}| - + \cdots$$

$$+ (-1)^k \sum_{1 \le i_1 < \ldots < i_{k-1} \le k} |(A_{i_1} \cap A_{k+1}) \cap \cdots \cap (A_{i_{k-1}} \cap A_{k+1})|$$

$$+ (-1)^{k+1} |(A_1 \cap A_{k+1}) \cap \cdots \cap (A_k \cap A_{k+1})|$$

Dies in (*) eingesetzt, ergibt

$$\Big|\bigcup_{i=1}^{k+1} A_i\Big| = \sum_{i=1}^{k} |A_i| - \sum_{1 \le i < j \le k} |A_i \cap A_j| + - \cdots + (-1)^{k+1} |A_1 \cap \cdots \cap A_k|$$

$$+ |A_{k+1}| - \sum_{i=1}^{k} |A_i \cap A_{k+1}| + - \cdots$$

$$+ (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_{k-1} \le k} |(A_{i_1} \cap A_{k+1}) \cap \cdots \cap (A_{i_{k-1}} \cap A_{k+1})|$$

$$+ (-1)^{k+2} |(A_1 \cap A_{k+1}) \cap \cdots \cap (A_k \cap A_{k+1})|$$

$$= \sum_{i=1}^{k+1} |A_i| - \sum_{1 \le i < j \le k+1} |A_i \cap A_j| + - \cdots$$

$$+ (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le k+1} |A_{i_1} \cap \cdots \cap A_{i_k}| + (-1)^{k+2} |A_1 \cap \cdots \cap A_{k+1}|. \quad \blacksquare$$

Eine sehr wichtige Methode, aus Mengen neue Mengen zu konstruieren, ist das Bilden von Produktmengen. Dazu geben wir zunächst eine vorbereitende Erklärung.

Seien $A_1,\ldots,A_n$ Mengen. Aus je n Elementen $x_1 \in A_1,\ldots,x_n \in A_n$ kann ein neues Objekt gebildet werden, das *geordnete n–Tupel* $(x_1,\ldots,x_n)$. Dabei heißen zwei n–Tupel $(x_1,\ldots,x_n)$ und $(y_1,\ldots,y_n)$ *gleich*, wenn $x_i = y_i$ für alle $i = 1,\ldots,n$ gilt. Ein n–Tupel ist also eine Zusammenfassung von n Elementen (i.a. aus verschiedenen Mengen), wobei es auf die Reihenfolge ankommt.

Definition. Es seien $A_1,\ldots,A_n$ Mengen. Dann heißt die Menge

$$A_1 \times \cdots \times A_n := \{(x_1,\ldots,x_n) |\ x_1 \in A_1,\ldots,x_n \in A_n\}$$

das (*kartesische*) *Produkt* von $A_1,\ldots,A_n$. Sind alle A_i gleich A, so schreibt man für die Produktmenge auch kürzer A^n.

In der linearen Algebra spielt die Produktmenge $\mathbb{R}^n$ eine wichtige Rolle.

4. Abbildungen

Gegeben seien zwei Mengen A und B. Eine *Abbildung* f von A nach (oder in) B, Schreibweise: $f\colon A \longrightarrow B$, ist eine Zuordnungsvorschrift, die jedem $x \in A$ genau ein $y \in B$ zuordnet. Man schreibt für y auch $f(x)$ und für die Zuordnung : $x \longmapsto f(x)$. Kurzschreibweisen:

$$\begin{array}{l} f\colon A \longrightarrow B \\ \quad x \longmapsto y = f(x) \end{array} \qquad \text{oder} \qquad \begin{array}{l} A \xrightarrow{f} B \\ x \xmapsto{f} y \end{array}$$

Andere gebräuchliche Namen für Abbildung sind *Funktion* (besonders, wenn $B \subset \mathbb{R}$

ist), *Transformation* oder *Operator*. Die Menge aller Abbildungen von A nach B bezeichnen wir mit B^A.

Bemerkung. Der Begriff der Abbildung ist hier mit Hilfe des zwar anschaulichen, aber mathematisch nicht definierten Wortes "Zuordnungsvorschrift" umgangssprachlich erklärt worden; es wurde somit wieder der "naive" Standpunkt eingenommen. Prinzipiell läßt sich der Abbildungsbegriff aber auch rein mengentheoretisch erklären:

Da jedem $x \in A$ genau ein $y \in B$ zugeordnet werden soll, ist die Abbildung f gegeben durch die Menge $\{(x,f(x))|\ x \in A\} \subset A \times B$. Umgekehrt wird zu jeder Menge $C \subset A \times B$ mit den Eigenschaften

(a) für alle $x \in A$, $y,\overline{y} \in B$ gilt: Aus $(x,y) \in C$, $(x,\overline{y}) \in C$ folgt $y = \overline{y}$,

(b) zu jedem $x \in A$ gibt es ein $y \in B$ mit $(x,y) \in C$,

eine Zuordnungsvorschrift gegeben durch $x \longmapsto y$, wobei y das nach (a) und (b) eindeutig bestimmte Element aus B ist, für das $(x,y) \in C$ gilt.

Als mathematische Definition können wir deshalb wählen:

Definition. Eine *Abbildung* $f\colon A \longrightarrow B$ ist eine Teilmenge von $A \times B$, die (a) und (b) erfüllt.

Wir bevorzugen im folgenden die anschauliche Vorstellung der Abbildung und geben der Menge $\{(x,f(x))|\ x \in A\}$ gleich einen anderen Namen.

Definition. Es sei $f\colon A \longrightarrow B$ eine Abbildung. Dann heißt A *Definitionsbereich*, B *Wertebereich*, und $f(A) := \{f(x)|\ x \in A\}$ *Bild* von A (unter f).

Für $x \in A$ heißt $f(x)$ *Bild(-punkt)* von x und für $y \in f(A)$ heißt x (ein) *Urbild* von y, wenn $f(x) = y$ gilt.

Für $C \subset B$ heißt $f^{-1}(C) := \{x \in A|\ f(x) \in C\}$ das *Urbild der Menge* C.

Die Menge $\{(x,f(x))|\ x \in A\}$ heißt *Graph* von f.

Bemerkungen. **(a)** Es gilt $f(A) \subset B$, aber es muß nicht Gleichheit gelten. Jedoch ist

es üblich, Abbildungen $f: A \longrightarrow B$ und $g: A \longrightarrow f(A)$ als gleich anzusehen, wenn $f(x) = g(x)$ für alle $x \in A$ gilt.

Die Begriffe "Wertebereich", "Bild" oder ähnliche wie "Wertemenge", "Bildmenge" werden in der Literatur nicht einheitlich benutzt.

(b) Jedes $x \in A$ hat ein eindeutiges Bild $f(x) \in f(A)$, aber ein $y \in f(A)$ kann mehrere (sogar unendlich viele) Urbilder haben.

Definition. Es sei $f: A \longrightarrow B$ eine Abbildung.

(a) f heißt *surjektiv* oder *Abbildung auf*, falls $f(A) = B$ gilt.

(b) f heißt *injektiv* oder *eineindeutig*, wenn aus $x_1, x_2 \in A$, $x_1 \neq x_2$ folgt, daß $f(x_1) \neq f(x_2)$.

(c) f heißt *bijektiv*, wenn f surjektiv und injektiv ist.

Beispiele.

(a) $f: \mathbb{R} \longrightarrow \mathbb{R}$
$x \longmapsto (x-1)^2$

ist weder injektiv noch surjektiv.

(b) $f: \mathbb{R} \longrightarrow \mathbb{R}$
$x \longmapsto x\,(x-1)\,(x+1)$

ist surjektiv aber nicht injektiv.

(c) $f: \mathbb{N} \longrightarrow \mathbb{Z}$

$$n \longmapsto \begin{cases} \frac{n-1}{2} & \text{falls } n \text{ ungerade} \\ -\frac{n}{2} & \text{falls } n \text{ gerade} \end{cases}$$

ist bijektiv.

Mit Hilfe des Abbildungsbegriffes können wir nun die Aufzählbarkeit einer unendlichen Menge genauer definieren.

Definition. Eine unendliche Menge A heißt *abzählbar*, wenn eine bijektive Abbildung $f: \mathbb{N} \longrightarrow A$ existiert. Eine nichtabzählbare unendliche Menge heißt *überabzählbar*.

Beispiel. Die Mengen $\mathbb{Z}$ und $\mathbb{Q}$ sind abzählbar, $\mathbb{R}$ ist überabzählbar.

Eine triviale Möglichkeit, eine Abbildung $f\colon A \longrightarrow A$ zu definieren, ist durch die Vorschrift $x \longmapsto x$ gegeben. Da diese Abbildung öfter auftritt, soll sie einen Namen erhalten. Weiterhin wollen wir aus gegebenen Abbildungen neue Abbildungen konstruieren.

Definitionen. **(a)** Es sei A eine Menge. Die Abbildung $\mathrm{id}_A : \begin{array}{c} A \longrightarrow A \\ x \longmapsto x \end{array}$ heißt *Identität* (auf A) oder *identische Abbildung.*

(b) Es seien $f\colon \begin{array}{c} A \longrightarrow B \\ x \longmapsto f(x) \end{array}$ eine Abbildung und $C \subset A$. Dann heißt $f|_C\colon \begin{array}{c} C \longrightarrow B \\ x \longmapsto f(x) \end{array}$ *Einschränkung* oder *Restriktion* von f auf C. Ist $g : C \longrightarrow B$ eine Abbildung mit $g = f|_C$, so heißt f *Fortsetzung* von g auf A.

(c) Es sei $f : A \longrightarrow B$ eine bijektive Abbildung. Dann heißt die Abbildung $f^{-1}\colon \begin{array}{c} B \longrightarrow A \\ y \longmapsto x \end{array}$ mit $f(x) = y$ *Umkehrabbildung* oder *inverse Abbildung* von f.

(d) Es seien $f\colon \begin{array}{c} A \longrightarrow B \\ x \longmapsto f(x) \end{array}$ und $g : \begin{array}{c} B \longrightarrow C \\ y \longmapsto g(y) \end{array}$ Abbildungen. Dann heißt die Abbildung $g \circ f\colon \begin{array}{c} A \longrightarrow C \\ x \longmapsto g(f(x)) \end{array}$ *zusammengesetzte Abbildung* oder *Komposition* oder auch *Verkettung.* Sprechweise: "g Kreis f" oder "g nach f".

Bemerkungen. **(a)** Ist f bijektiv, so auch f^{-1} und es gilt $(f^{-1})^{-1} = f$.

(b) Ist f bijektiv, so gilt $f^{-1} \circ f = \mathrm{id}_A$ und $f \circ f^{-1} = \mathrm{id}_B$.

(c) Die Komposition ist nicht kommutativ (siehe (b)) aber assoziativ. Es gilt nämlich $h \circ (g \circ f) = (h \circ g) \circ f$.

(d) Man beachte folgendes: Ist $f\colon A \longrightarrow B$ eine Abbildung und ist $C \subset B$, so existiert das Urbild $f^{-1}(C)$ immer, auch wenn die Umkehrabbildung f^{-1} nicht existiert. Existiert aber f^{-1}, so ist $f^{-1}(C)$ gerade das Bild von C unter f^{-1}.

Beispiele. **(a)** Für die Funktion aus dem obigen Beispiel (c) existiert die Umkehrfunktion. Es ist

$$f^{-1}: \mathbb{Z} \longrightarrow \mathbb{N}$$

$$z \longmapsto \begin{cases} 2z+1 & \text{für } z \geq 0 \\ -2\ z & \text{für } z < 0 \end{cases}$$

(b) Für $f: \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R} \setminus \{0\}$ mit $f(x) = \frac{1}{x}$ gilt $f^{-1} = f$.

In den voranstehenden Beispielen haben wir die Umkehrabbildung f^{-1} einfach angegeben, d.h. wir haben eine Abbildung $f: A \longrightarrow B$ und eine weitere $f^{-1}: B \longrightarrow A$ mit $f^{-1} \circ f = \mathrm{id}_A$ und $f \circ f^{-1} = \mathrm{id}_B$. Reicht das nun aus, oder müssen wir noch nachweisen, daß f überhaupt bijektiv ist? Die Antwort gibt der folgende Satz.

Satz 4. *Es seien A, B nichtleere Mengen und $f: A \longrightarrow B$ eine Abbildung. Dann gilt:*

(a) *f ist genau dann injektiv, wenn eine Abbildung $g: B \longrightarrow A$ mit $g \circ f = id_A$ existiert.*

(b) *f ist genau dann surjektiv, wenn eine Abbildung $g: B \longrightarrow A$ mit $f \circ g = id_B$ existiert.*

(c) *f ist genau dann bijektiv, wenn eine Abbildung $g: B \longrightarrow A$ mit $g \circ f = id_A$ und $f \circ g = id_B$ existiert.*

Beweis. (a) Sei f injektiv, d.h. jedes $y \in f(A)$ hat genau ein Urbild $x \in A$. Wir erklären die Abbildung $g: B \longrightarrow A$ durch folgende Vorschrift

$$g: y \longmapsto \begin{cases} x & \text{, falls } y \in f(A) \text{ und } f(x) = y \\ x_0 & \text{, falls } y \notin f(A) \end{cases}$$

Hierbei ist $x_0 \in A$ ein festes Element, das wegen $A \neq \emptyset$ existiert. Dann gilt $g \circ f(x) = x$ für alle $x \in A$, also $g \circ f = \mathrm{id}_A$.

Umgekehrt sei eine Abbildung $g: B \longrightarrow A$ gegeben mit $g \circ f = \mathrm{id}_A$. Seien $x_1, x_2 \in A$ mit $f(x_1) = f(x_2)$. Dann gilt $x_1 = g \circ f(x_1) = g(f(x_1)) = g(f(x_2)) = g \circ f(x_2) = x_2$, d.h. f ist injektiv.

(b) Sei f surjektiv. Wir erklären $g: B \longrightarrow A$ dadurch, daß wir zu jedem $y \in B$ ein Urbild $x \in f^{-1}(\{y\})$ auswählen. Damit folgt für alle $y \in B$: $f \circ g(y) = f(g(y)) = f(x) = y$.

Umgekehrt sei $g : B \longrightarrow A$ mit $f \circ g = \mathrm{id}_B$ gegeben. Sei $y \in B$. Dann ist $y = f \circ g\,(y) = f(g\,(y))$, d.h. $y \in f(A)$ und f ist surjektiv.
(c) Ist f bijektiv, so erfüllt $g = f^{-1}$ die Bedingungen. Existiert umgekehrt eine Abbildung $g : B \longrightarrow A$ mit $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$, so ist f nach (a) injektiv und nach (b) surjektiv, also bijektiv. ∎

Bemerkung. Bei der Behauptung (b) haben wir g dadurch definiert, daß wir aus den Mengen $f^{-1}(\{y\})$, $y \in B$, jeweils ein Element x ausgewählt haben. Dieses plausible Vorgehen ist aber nur deswegen möglich, weil wir an dieser Stelle stillschweigend das sogenannte *Auswahlaxiom* der axiomatischen Mengenlehre verwendet haben.

Über die Verwendung dieses Axioms und seiner äquivalenten Formen herrschen unterschiedliche Auffassungen, zum Teil wegen der manchmal absurd erscheinenden Folgerungen daraus. Man versucht daher häufig, soweit wie möglich ohne Auswahlaxiom auszukommen, andererseits ist es aber für bestimmte mathematische Gebiete wie etwa die Funktionalanalysis unerläßlich. Abgesehen von der obigen, nicht wesentlichen Stelle werden wir das Auswahlaxiom, bzw. eine dazu äquivalente Aussage, nur noch einmal, beim Beweis des Basisergänzungssatzes für unendlich dimensionale Vektorräume, explizit verwenden.

5. Relationen

Eine Relation auf einer Menge A ist eine Beziehung, die zwischen den Elementen von A bestehen kann oder nicht. Es werden hier nur Beziehungen zwischen je zwei Elementen betrachtet. Man kann allgemeiner auch n–stellige Relationen untersuchen und Elemente aus verschiedenen Mengen zulassen; dann fällt z.B. der Abbildungsbegriff darunter.

Da eine Relation auf der Menge A vollständig durch die Paare $(x,y) \in A^2$ beschrieben wird, die in dieser Relation stehen, ist eine *Relation* R mathematisch nichts anderes als eine Teilmenge von A^2. Statt $(x,y) \in R$ schreiben wir aber $x\,R\,y$ und sagen, daß x *in Relation* R *zu* y *steht.*

Beispiele für Relationen sind etwa die Kleinerbeziehung, die Kleiner–Gleich–Beziehung und die Gleichheitsbeziehung bei reellen Zahlen oder die Teilerbeziehung bei natürlichen Zahlen. Besonders wichtig sind zwei spezielle Typen von Relationen, die Ordnungsrelationen und die Äquivalenzrelationen, die wir im folgenden näher untersuchen wollen. In der Informatik spielen allerdings auch andere Relationen eine Rolle.

Ordnungsrelationen

Hier versuchen wir die Beziehung $\leq$ von $\mathbb{R}$ auf allgemeinere Mengen zu übertragen, indem wir charakteristische Eigenschaften von $\leq$ als Axiome übernehmen.

Definition. Eine Relation R auf einer Menge A heißt *Ordnungsrelation* oder *Ordnung* (in älteren Büchern auch *partielle Ordnung* oder *Halbordnung*), wenn folgende Gesetze gelten. Statt $x\,R\,y$ schreiben wir $x \leq y$.

(a) Für alle $x \in A$ gilt: $x \leq x$. *("Reflexivität")*

(b) Für alle $x,y \in A$ gilt: Aus $x \leq y$ und $y \leq x$ folgt $x = y$. *("Antisymmetrie")*

(c) Für alle $x,y,z \in A$ gilt: Aus $x \leq y$ und $y \leq z$ folgt $x \leq z$. *("Transitivität")*

Ist $\leq$ Ordnung auf A, so heißt $(A,\leq)$ *geordnete Menge.*

Erfüllt eine Ordnung $\leq$ außerdem noch die Bedingung

(d) Für alle $x,y \in A$ gilt: $x \leq y$ oder $y \leq x$, *("Vergleichbarkeit")*

so heißt sie *Totalordnung* und $(A,\leq)$ heißt dann *total geordnete Menge.*

Beispiele. **(a)** Die übliche Ordnung $\leq$ auf $\mathbb{R}$ ist eine Totalordnung.

(b) Auf $\mathbb{R}^n$ wird durch

$$(x_1,\ldots,x_n) \leq (y_1,\ldots,y_n) \;:\Longleftrightarrow\; x_i \leq y_i\,, \quad i = 1,\ldots,n\,,$$

eine Ordnung erklärt, die keine Totalordnung ist.

(c) Die Inklusion $\subset$ von Mengen ist eine Ordnung, $(\mathcal{P}(A),\subset)$ ist geordnete Menge.

(d) Ist $(A,\leq)$ eine geordnete Menge und $B \subset A$, so ist $(B, \leq)$ mit der auf B eingeschränkten Ordnung ebenfalls geordnet. Man spricht dann von der *induzierten*

Ordnung auf B.

Wie bei den reellen Zahlen können wir auch bei geordneten Mengen von oberen und unteren Schranken, größten und kleinsten Elementen sowie von Supremum und Infimum sprechen.

Definition. Es seien $(A,\leq)$ eine geordnete Menge und $B \subset A$. Ein Element $z \in A$ heißt *obere Schranke* von B, falls $x \leq z$ für alle $x \in B$ gilt und *untere Schranke* von B, falls $z \leq x$ für alle $x \in B$ gilt. Gilt außerdem $z \in B$, so heißt z *größtes*, bzw. *kleinstes* Element von B.

$z \in A$ heißt *Supremum* von B oder *kleinste obere Schranke* von B, falls z obere Schranke von B ist und für jede obere Schranke z' von B gilt: $z \leq z'$. Schreibweise: $z = \sup B$ oder $z = \sup_{x \in B} x$.

$z \in A$ heißt *Infimum* von B oder *größte untere Schranke* von B, falls z untere Schranke von B ist und für jede untere Schranke z' von B gilt: $z' \leq z$. Schreibweise: $z = \inf B$ oder $z = \inf_{x \in B} x$.

Beispiele. **(a)** In $(\mathbb{R},\leq)$ besitzt jede nichtleere endliche Teilmenge $A \subset \mathbb{R}$ Supremum und Infimum. Für die abzählbare Teilmenge $\mathbb{N} \subset \mathbb{R}$ existiert zwar $\inf \mathbb{N} = 1$ aber nicht $\sup \mathbb{N}$.

(b) In $(\mathcal{P}(A),\subset)$ besitzt jede nichtleere Teilmenge $\mathcal{M} \subset \mathcal{P}(A)$ Infimum und Supremum, nämlich

$$\inf \mathcal{M} = \bigcap_{B \in \mathcal{M}} B, \quad \sup \mathcal{M} = \bigcup_{B \in \mathcal{M}} B.$$

Definition. Eine geordnete Menge $(A, \leq)$, in der je zwei Elemente immer ein Infimum und ein Supremum besitzen, heißt *Verband*. Besitzt jede nichtleere Teilmenge $B \subset A$ Infimum und Supremum, so heißt der Verband *vollständig*.

Beispiele. $(\mathbb{R},\leq)$ ist ein Verband, $(\mathcal{P}(A),\subset)$ ist ein vollständiger Verband.

Existiert das Supremum z einer geordneten Menge $(A, \leq)$, und ist $z \in A$, so ist

z das größte Element von A. Entsprechend ist inf A, falls es existiert und zu A gehört, das kleinste Element von A.

Nicht zu verwechseln mit diesen Begriffen sind die Begriffe maximales bzw. minimales Element von A.

Definition. Es sei $(A, \leq)$ eine geordnete Menge. Dann heißt $y \in A$ *maximales Element*, falls aus $y \leq x$, $x \in A$, stets $y = x$ folgt.

$y \in A$ heißt *minimales Element*, falls aus $x \leq y$, $x \in A$, stets $y = x$ folgt.

In einer totalgeordneten Menge $(A, \leq)$ ist jedes maximale Element auch größtes Element und umgekehrt ist ein größtes Element auch maximal. Entsprechendes gilt für minimale Elemente.

Die folgende Aussage ist ein hinreichendes Kriterium für die Existenz eines maximalen Elementes in einer geordneten Menge $(A, \leq)$. Sie ist zum Auswahlaxiom äquivalent und heißt Zornsches Lemma. Für einen Beweis mit Hilfe des Auswahlaxioms verweisen wir auf das schon zitierte Buch von Halmos.

Zornsches Lemma. *Es sei $(A,\leq)$ eine geordnete Menge und jede Menge $B \subset A$, die bezüglich der induzierten Ordnung totalgeordnet ist, besitze eine obere Schranke $z \in A$. Dann gibt es in A ein maximales Element.*

Äquivalenzrelationen

Eine wichtige Konstruktion in der Mathematik ist das Zusammenfassen von vergleichbaren (äquivalenten) Objekten, d.h. die Einteilung der Elemente einer Menge in Klassen. Jede Klasseneinteilung entspricht aber einer Relation auf dieser Menge, allerdings von einer anderen Art als die Ordnungsrelationen.

Definition. Eine Relation R auf einer Menge A heißt *Äquivalenzrelation*, wenn die folgenden Gesetze gelten. Statt $x\,R\,y$ schreiben wir $x \sim y$.

(a) Für alle $x \in A$ gilt: $x \sim x$. *("Reflexivität")*

(b) Für alle $x,y \in A$ gilt: Aus $x \sim y$ folgt $y \sim x$. *("Symmetrie")*

(c) Für alle $x,y,z \in A$ gilt: Aus $x \sim y$ und $y \sim z$ folgt $x \sim z$. *("Transitivität")*

Ist $\sim$ Äquivalenzrelation auf A und $x \in A$, so sei

$$[x]_\sim := \{y \in A \mid x \sim y\}.$$

$[x]_\sim$ heißt die *Äquivalenzklasse* von x, und x heißt *Repräsentant* der Äquivalenzklasse. Die Menge aller Äquivalenzklassen von Elementen $x \in A$ wird mit $A/_\sim$ bezeichnet und *Quotienten–* oder *Faktormenge* genannt.

Satz 5 und Definition. *Sei $\sim$ Äquivalenzrelation auf der Menge A. Dann gilt:*

(a) $[x]_\sim \neq \emptyset$ *für alle* $x \in A$,

(b) $[x]_\sim \neq [y]_\sim \Longrightarrow [x]_\sim \cap [y]_\sim = \emptyset$,

(c) $\bigcup_{x \in A} [x]_\sim = A$.

Man sagt, daß das System $\mathcal{M} = A/_\sim$ der Äquivalenzklassen eine *Partition* oder *Klasseneinteilung* von A bildet.

Ist umgekehrt $\mathcal{M} \subset \mathcal{P}(A)$ eine Partition von A, d.h. ein System nichtleerer, paarweise disjunkter Mengen, deren Vereinigung A ist, so gibt es eine Äquivalenzrelation $\sim$ auf A mit $\mathcal{M} = A/_\sim$.

Beweis. (a) Wegen $x \sim x$ ist $x \in [x]_\sim$, also $[x]_\sim \neq \emptyset$.

(b) Wir führen einen Widerspruchsbeweis und nehmen dazu an, daß $[x]_\sim \neq [y]_\sim$ und $[x]_\sim \cap [y]_\sim \neq \emptyset$ gilt. Dann existiert ein $z \in [x]_\sim \cap [y]_\sim$. Daraus folgt $z \sim x$ und $z \sim y$, also wegen der Eigenschaften (b),(c) einer Äquivalenzrelation auch $x \sim y$. Seien nun $x' \in [x]_\sim$, $y' \in [y]_\sim$. Aus $x' \sim x$, $x \sim y$ folgt $x' \sim y$, d.h. $x' \in [y]_\sim$ und aus $y' \sim y$, $y \sim x$ folgt analog $y' \in [x]_\sim$, also $[x]_\sim = [y]_\sim$. Dies ist ein Widerspruch. Es gilt daher $[x]_\sim \cap [y]_\sim = \emptyset$.

(c) ist trivial wegen $x \in [x]_\sim$.

Nun beweisen wir die Umkehrung. Da $\mathcal{M}$ eine Partition von A ist, liegt jedes Element $x \in A$ in genau einem $B_x \in \mathcal{M}$. Dadurch wird eine Abbildung $f: A \longrightarrow \mathcal{M}$ mit $f(x) = B_x$ definiert. Wir setzen nun $x \sim y :\Longleftrightarrow f(x) = f(y)$. Dann folgt sofort,

daß $\sim$ eine Äquivalenzrelation ist mit $B_x = [x]_\sim$. ∎

Bei dem Beweis von Satz 5 haben wir gleich auch die wichtigste Methode zur Entstehung von Äquivalenzrelationen kennengelernt, die *Faktorisierung nach einer Abbildung.*

Satz 6. *Es sei* $f: A \longrightarrow B$ *eine Abbildung. Dann wird durch* $x \sim y :\Longleftrightarrow f(x) = f(y)$ *eine Äquivalenzrelation auf* A *erklärt.*

Beweis. Einfache Übungsaufgabe. ∎

Schreibweise: Statt $A/_\sim$ schreibt man hier $A/_f$.

Bemerkung. Auf triviale Weise wird jede Äquivalenzrelation $\sim$ durch eine Abbildung erzeugt. Sei nämlich

$$\begin{aligned} k: A &\longrightarrow A/_\sim \\ x &\longmapsto [x]_\sim \end{aligned}$$

die sogenannte *kanonische Abbildung.* Diese ist surjektiv und es gilt $A/_\sim = A/_k$.

Nun können wir einen Satz beweisen, der den Ausgangspunkt einer Reihe später oft benutzter Sätze bildet. Dabei erklärt sich auch der Ausdruck *Faktorisieren* bzw. *Faktormenge*, da Abbildungen in "Faktoren" zerlegt werden (bezüglich der Komposition als "Multiplikation").

Satz 7 (Grundform des Homomorphie–Satzes). *Es seien* $f: A \longrightarrow B$ *eine Abbildung und* $k: A \longrightarrow A/_f$ *die kanonische Abbildung. Dann existiert eine injektive Abbildung* $\bar{f}: A/_f \longrightarrow B$ *mit* $f = \bar{f} \circ k$. *Ist* f *surjektiv, so ist* $\bar{f}$ *bijektiv.*

Beweis. Wir definieren $\bar{f}$ durch $\bar{f}([x]_\sim) = f(x)$ für alle $x \in A$. Diese Definition ist unabhängig von der speziellen Wahl des Repräsentanten $x \in [x]_\sim$, denn aus $[x]_\sim = [y]_\sim$ folgt $x \sim y$ und somit nach Definition von $\sim$ auch $f(x) = f(y)$. Offensichtlich ist $f = \bar{f} \circ k$. Schließlich folgt aus $f(x) = f(y)$ noch $x \sim y$ und $[x]_\sim = [y]_\sim$, also ist $\bar{f}$ injektiv. Ist f surjektiv, so ist auch $\bar{f}$ surjektiv, also insgesamt bijektiv. ∎

Kapitel 1 Grundbegriffe der Algebra

§ 1 Lineare Gleichungssysteme

Wie wir schon in der Einleitung gesagt haben, stand das Lösen von (linearen) Gleichungen bzw. Systemen von Gleichungen nicht nur am Anfang der historischen Entwicklung der linearen Algeba, sondern es stellt auch gegenwärtig eines der wichtigsten Anwendungsgebiete für diese mathematische Theorie dar. Wir wollen deshalb mit der Betrachtung linearer Gleichungssysteme beginnen und damit die nachfolgenden theoretischen Überlegungen motivieren. Wir betrachten zunächst ein Beispiel.

Beispiel. Ein Betrieb produziere n Nahrungsmittel $N_1,\ldots,N_n$, wozu m Rohstoffe $R_1,\ldots,R_m$ benötigt werden. Zur Herstellung einer Einheit des Nahrungsmittels N_j werden dabei a_{ij} Einheiten des Rohstoffs R_i gebraucht, $i = 1,\ldots,m$, $j = 1,\ldots,n$.

Gesucht ist ein optimaler Produktionsplan, d.h. ein Plan, wieviele Einheiten x_j von N_j produziert werden sollen, wenn insgesamt b_i Einheiten von R_i vorhanden sind und wenn möglichst keine Rohstoffe übrig bleiben sollen. Dies ist natürlich nur eine, auf unsere weiteren Überlegungen zugeschnittene Fragestellung; oft wird man in Kauf nehmen, daß Rohstoffe übrig bleiben, und dafür lieber die Produktionskosten minimieren oder den Gewinn maximieren.

Werden jeweils $x_1,\ldots,x_n$ Einheiten produziert, so benötigt man insgesamt

$$a_{i1}\,x_1 + \cdots + a_{in}\,x_n$$

Einheiten des Rohstoffs R_i. Ein Produktionsplan $(x_1,\ldots,x_n) \in \mathbb{R}^n$, der im obigen Sinne optimal ist, muß also das folgende System von Gleichungen erfüllen:

$$(*) \qquad \begin{array}{ccc} a_{11}\,x_1 + \cdots + a_{1n}\,x_n & = & b_1 \\ \vdots \qquad\qquad\quad \vdots & & \vdots \\ a_{m1}\,x_1 + \cdots + a_{mn}\,x_n & = & b_m \end{array}$$

Hierbei ist $a_{ij} \in \mathbb{R}$ und $b_i \in \mathbb{R}$. (*) ist die allgemeine Form eines *linearen Gleichungssystems* (Kurzschreibweise: LGS) und $x_1,\ldots,x_n$ sind die *Unbekannten* des LGS. Jedes n–Tupel $(x_1,\ldots,x_n) \in \mathbb{R}^n$, das (*) erfüllt, heißt *Lösung* des linearen Gleichungssystems. Alle Lösungen von (*) bilden die *Lösungsmenge.*

Wegen $a_{ij} \in \mathbb{R}$, $b_i \in \mathbb{R}$, $x_j \in \mathbb{R}$ sprechen wir auch von einem *reellen* LGS oder einem LGS über $\mathbb{R}$. In unserem Beispiel war sogar $a_{ij} \geq 0$, $b_i \geq 0$, und nur Lösungen mit $x_j \geq 0$ waren von Interesse. Wir wollen aber solche Vorzeichenbedingungen jetzt außer acht lassen, um die Theorie nicht zu speziell zu machen.

Welche Möglichkeiten ergeben sich für die Lösungsmenge eines linearen Gleichungssystems? Offensichtlich spielt der Fall $(b_1,\ldots,b_m) = (0,\ldots,0)$ eine besondere Rolle, da hier die triviale Lösung $(x_1,\ldots,x_n) = (0,\ldots,0)$ existiert. Ist $(b_1,\ldots,b_m) = (0,\ldots,0)$, so heißt das lineare Gleichungssystem (*) *homogen*, andernfalls *inhomogen* .

Wir geben nun einige Beispiele für des Lösungsverhalten linearer Gleichungs–systeme.

Beispiele. **(a)** Das reelle LGS

$$\begin{aligned} &(1) && x_1 + x_2 + x_3 = 3 \\ &(2) && x_1 - x_2 + 2\,x_3 = 2 \\ &(3) && 2\,x_1 \quad + 3\,x_3 = 1 \end{aligned}$$

ist unlösbar, denn durch Addition der ersten beiden Gleichungen erhalten wir mit $2\,x_1 + 3\,x_3 = 5$ einen Widerspruch zur 3. Gleichung.

(b) Wir betrachten das reelle LGS

$$\begin{aligned} &(1) && x_1 + x_2 + x_3 = 3 \\ &(2) && x_1 - x_2 + 2\,x_3 = 2 \\ &(3) && \quad x_2 + x_3 = 2 \end{aligned}$$

Aus (1), (3) folgt $x_1 = 1$. Aus (1) + (2) erhalten wir $2 + 3\,x_3 = 5$, also $x_3 = 1$. Aus (3) folgt schließlich $x_2 = 1$. Also kann höchstens $(1,1,1)$ Lösung sein. Daß dieses Tripel auch Lösung ist, folgt sofort durch Einsetzen. Hier liegt also eine eindeutige Lösung

vor.

(c) Als drittes Beispiel sei das reelle LGS

$$\begin{aligned} &(1) && x_1 + x_2 + x_3 = 3 \\ &(2) && x_1 - x_2 + 2\,x_3 = 2 \\ &(3) && 2\,x_1 + 3\,x_3 = 5 \end{aligned}$$

gegeben. Wegen (1) + (2) = (3) kann die Gleichung (3) weggelassen werden. Aus (1) und (2) folgen $2\,x_1 = 5 - 3\,x_3$, $2\,x_2 = 1 + x_3$. Also ist jedes Tripel

$$(\tfrac{5}{2} - \tfrac{3}{2}\,a, \tfrac{1}{2} + \tfrac{1}{2}\,a, a)\,, \qquad a \in \mathbb{R}\,,$$

eine Lösung.

Für ein reelles LGS gibt es somit folgende Möglichkeiten: Es ist unlösbar, oder es ist lösbar und besitzt genau eine Lösung, oder es ist lösbar und hat unendlich viele Lösungen.

Ein homogenes LGS ist immer lösbar; hier interessiert man sich dafür, ob nichttriviale Lösungen existieren.

Bei der Betrachtung eines allgemeinen linearen Gleichungssystems (*) ergeben sich nun die folgenden Fragen:

Ist es möglich, einem linearen Gleichungssystem die Lösbarkeit anzusehen? Dazu kann man zunächst eine übersichtliche Schreibweise für (*) einführen, etwa durch Einklammern zusammengehöriger Ausdrücke:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Für eine weitere Behandlung müssen dann Rechenoperationen und Rechenregeln für solche Klammern (Matrizen) gefunden werden.

Welche Struktur hat die Lösungsmenge eines linearen Gleichungssystems? Im homogenen Fall sind beispielsweise mit $(x_1,\ldots,x_n)$ und $(y_1,\ldots,y_n)$ auch die n–Tupel

$(x_1 + y_1,\ldots,x_n + y_n)$ sowie $(a\,x_1,\ldots,a\,x_n)$, $a \in \mathbb{R}$, Lösungen. Hier werden Rechenregeln für n–Tupel benötigt.

Es erweist sich nun als vorteilhaft, von der konkreten Situation des LGS (*) abzusehen und Mengen mit Rechenoperationen (Verknüpfungen) zunächst in allgemeinerem Rahmen zu untersuchen. Dabei wird man auch die Menge $\mathbb{R}$ der reellen Zahlen durch eine allgemeinere Menge mit entsprechenden Rechengesetzen (Körper) ersetzen. Dieses Vorgehen, das typisch ist für die Mathematik, erlaubt es, allgemeine Prinzipien und Gesetze zu erkennen, die sich im Spezialfall dann auf die Ausgangsfragestellung (hier: die Behandlung des LGS (*)) anwenden lassen. Der abstrakte Standpunkt der Algebra ist aber auch von den Anwendungen her gerechtfertigt; in der Physik (Schwingungen und Wellen) treten häufig Probleme auf, denen der Zahlkörper der komplexen Zahlen zugrundeliegt, in der Informatik (Codierungstheorie) arbeitet man oft mit endlichen Körpern.

Wir werden im folgenden also Mengen behandeln, auf denen eine oder mehrere Verknüpfungen gegeben sind. Man nennt solche Mengen auch *algebraische Strukturen*. Die wichtigsten algebraischen Strukturen mit einer Verknüpfung sind die Gruppen, sie werden in § 2 untersucht. In § 3 betrachten wir dann algebraische Strukturen mit zwei Verknüpfungen, nämlich Körper und Ringe. Eine umfassende Behandlung all dieser Strukturen ist Aufgabe der Algebra. Hier, im Rahmen der linearen Algebra, müssen wir uns im wesentlichen auf die Definitionen und einige für das spätere Verständnis wichtige Eigenschaften beschränken.

§ 2 Gruppen

Gruppen spielen nicht nur in der Mathematik selbst eine bedeutende Rolle, sondern auch in der Informatik sowie in vielen Naturwissenschaften wie etwa Physik, Chemie, Biologie oder Kristallographie. Dort werden insbesondere die Symmetriegruppen von Kristallen untersucht.

Wir präzisieren zunächst den Begriff Verknüpfung. Darunter verstehen wir eine Vorschrift, mit deren Hilfe wir aus je zwei Elementen einer Menge A ein weiteres Element dieser Menge erhalten, also mathematisch ausgedrückt:

Eine (*innere*) *Verknüpfung* auf der Menge A ist eine Abbildung $f\colon A \times A \longrightarrow A$.

Wir schreiben die Verknüpfungen üblicherweise nicht als Abbildungen f_1 , f_2 , f_3 ,... sondern verwenden Symbole wie $+$, $\cdot$, $\circ$ usw. Statt $f(x,y)$ heißt es dann $x+y$, $x\cdot y$ und $x \circ y$.

Beispiele. **(a)** Addition und Multiplikation auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$.

(b) Komposition von Abbildungen $g,h : B \longrightarrow B$ einer Menge B in sich, $g \circ h : B \longrightarrow B$.

(c) Auf der Potenzmenge $\mathcal{P}(B)$ einer Menge B sind $\cup$, $\cap$, $\setminus$, Δ Verknüpfungen.

Die meisten dieser Verknüpfungen haben die Eigenschaft assoziativ zu sein, einige sind auch kommutativ. Genauer bedeutet dies:

Eine Verknüpfung $\circ$ auf einer Menge A heißt *assoziativ*, falls für alle x, y, $z \in A$

$$(x \circ y) \circ z = x \circ (y \circ z) \qquad (\text{"Assoziativgesetz"})$$

gilt, und sie heißt *kommutativ*, falls für alle x, $y \in A$ gilt:

$$x \circ y = y \circ x\,. \qquad (\text{"Kommutativgesetz"})$$

Als erste einfache algebraische Strukturen betrachten wir nun Halbgruppen.

Definition. Es seien A eine nichtleere Menge und $\circ$ eine Verknüpfung auf A. Ist diese assoziativ, so heißt $(A,\circ)$ eine *Halbgruppe*. Ist die Verknüpfung außerdem kommutativ, so sprechen wir von einer *kommutativen Halbgruppe*.

In einer Halbgruppe sind Klammern entbehrlich, wir schreiben deshalb z.B. statt $(x \circ y) \circ z$ einfacher $x \circ y \circ z$. Außerdem läßt sich jedes Element x beliebig oft mit sich selbst verknüpfen. Wir verwenden hierfür die abkürzende Potenzschreibweise

$$x^k := \underbrace{x \circ x \circ \cdots \circ x}_{k\text{-mal}}, \quad k \in \mathbb{N}.$$

In vielen Halbgruppen $(A,\circ)$ gibt es ein Element, das vor allen anderen Elementen dadurch ausgezeichnet ist, daß es sich bezüglich der Verknüpfung $\circ$ neutral verhält.

Definition. $e \in A$ heißt *neutrales Element* oder *Neutralelement* der Halbgruppe $(A,\circ)$, falls für alle $x \in A$ gilt: $x \circ e = e \circ x = x$.

In einer Halbgruppe gibt es höchstens ein neutrales Element: Ist nämlich e' ebenfalls ein neutrales Element von $(A,\circ)$, so gilt nach Definition $e \circ e' = e' \circ e = e$ und $e' \circ e = e \circ e' = e'$, also $e = e'$.

In Halbgruppen, in denen ein neutrales Element existiert, können wir auch vom Invertieren eines Elementes reden.

Definition. Es sei $(A,\circ)$ eine Halbgruppe mit Neutralelement e. $x \in A$ heißt *invertierbar*, genau dann wenn ein $x^{-1} \in A$ existiert mit $x \circ x^{-1} = x^{-1} \circ x = e$. x^{-1} heißt dann *inverses Element* oder *Inverses* zu x.

In einer Halbgruppe $(A,\circ)$ mit Neutralelement e gibt es zu jedem $x \in A$ höchstens ein inverses Element: Ist nämlich x' ebenfalls ein inverses Element von x, so gilt nach Definition $x \circ x' = x' \circ x = e$ und $x \circ x^{-1} = x^{-1} \circ x = e$. Also ist $x \circ x^{-1} = x \circ x'$ und somit $x^{-1} \circ (x \circ x^{-1}) = x^{-1} \circ (x \circ x')$. Mit der Assoziativität folgt $e \circ x^{-1} = e \circ x'$, also $x^{-1} = x'$.

Beispiel. Ist $A \neq \emptyset$, so sind $(\mathcal{P}(A),\cup)$ und $(\mathcal{P}(A),\cap)$ kommutative Halbgruppen mit Neutralelement ($\emptyset$ im ersten Fall, A im zweiten), aber es gibt im allgemeinen keine Inversen. Dagegen ist $(\mathcal{P}(A),\Delta)$ eine kommutative Halbgruppe mit dem neutralen

Element $\emptyset$, in der zu jedem $B \in \mathcal{P}(A)$ ein inverses Element existiert, nämlich B selbst.

Besonders wichtige Beispiele von Halbgruppen sind die Gruppen, mit denen wir uns im folgenden eingehender beschäftigen wollen.

Definition. Es seien A eine Menge und $\circ$ eine Verknüpfung auf A. $(A,\circ)$ heißt *Gruppe*, wenn folgende Bedingungen erfüllt sind:

(a) $(A,\circ)$ ist eine Halbgruppe.

(b) $(A,\circ)$ besitzt ein neutrales Element e.

(c) Zu jedem $x \in A$ gibt es ein inverses Element.

Die Gruppe $(A,\circ)$ heißt *kommutativ* oder *abelsch* (nach Nils Henrik Abel, 1802 – 1829), wenn die Halbgruppe $(A,\circ)$ kommutativ ist.

Beim Rechnen in einer Gruppe $(A,\circ)$ schreibt man statt $x \circ y$ oft xy. In abelschen Gruppen wird als Verknüpfungssymbol meist + gewählt. Statt e schreibt man dann 0, statt x^{-1} auch $-x$ und statt $x + (-y)$ kurz $x - y$.

Ist klar, welche Verknüpfung gemeint ist, so redet man oft von der Gruppe A.

Bemerkung. Die Gruppenaxiome (b) und (c) können durch schwächere Axiome ersetzt werden:

(b') $(A,\circ)$ besitzt ein rechtsneutrales Element e, d.h. es gibt ein Element $e \in A$, so daß für alle $x \in A$ gilt: $x \circ e = x$.

(c') Zu jedem $x \in A$ gibt es ein rechtsinverses Element, d.h. es gibt ein $x^{-1} \in A$ mit $x \circ x^{-1} = e$.

Um den Umgang mit den Gruppenaxiomen einzuüben, beweisen wir das folgende Gruppenkriterium.

Satz 1. *Eine Halbgruppe $(A,\circ)$ ist genau dann eine Gruppe, wenn es zu jedem $x \in A$ und zu jedem $y \in A$ Elemente $z, \overline{z} \in A$ gibt mit $x \circ z = y$ und $\overline{z} \circ x = y$.*

Ist dies der Fall, so sind z und $\overline{z}$ eindeutig bestimmt.

Beweis. Sei zunächst $(A,\circ)$ eine Gruppe. Setzen wir $z := x^{-1} \circ y$, so gilt $x \circ z =$

$x \circ (x^{-1} \circ y) = (x \circ x^{-1}) \circ y = e \circ y = y$. Entsprechend ist $\overline{z} := y \circ x^{-1}$ eine Lösung der Gleichung $\overline{z} \circ x = y$.

Ist $(A,\circ)$ eine Halbgruppe, so folgern wir umgekehrt zunächst aus der Lösbarkeit der obigen Gleichungen die Existenz eines Neutralelementes in $(A,\circ)$: Wegen $A \neq \emptyset$ gibt es ein $x_0 \in A$. Nach Voraussetzung existieren dann in A Elemente e und e' mit $x_0 \circ e = x_0$ und $e' \circ x_0 = x_0$. Sei nun $x \in A$ beliebig. Dann gibt es wiederum nach Voraussetzung Elemente $z,\overline{z} \in A$ mit $x_0 \circ z = x$ und $\overline{z} \circ x_0 = x$. Daraus folgt

$$e' \circ x = e' \circ (x_0 \circ z) = (e' \circ x_0) \circ z = x_0 \circ z = x\,,$$

$$x \circ e = (\overline{z} \circ x_0) \circ e = \overline{z} \circ (x_0 \circ e) = \overline{z} \circ x_0 = x\,,$$

also speziell $e = e' \circ e = e'$. Damit ist e Neutralelement von $(A,\circ)$.

Nachweis von Axiom (c): Sei $x \in A$ beliebig. Zu x und e existieren $z \in A$ und $\overline{z} \in A$ mit $x \circ z = e$ und $\overline{z} \circ x = e$. Wegen $z = e \circ z = \overline{z} \circ x \circ z = \overline{z} \circ e = \overline{z}$ ist z inverses Element von x. Damit sind alle Gruppenaxiome nachgewiesen, also ist $(A,\circ)$ eine Gruppe.

Eindeutigkeit: Aus $x \circ z = y$ und $x \circ z' = y$ folgt $x \circ z = x \circ z'$ und daraus $x^{-1} \circ x \circ z = x^{-1} \circ x \circ z'$, also $z = z'$. Die Eindeutigkeit von $\overline{z}$ wird analog bewiesen. ∎

Beispiele. **(a)** $(\mathbb{N},+)$ und $(\mathbb{N}_0,+)$ sind kommutative Halbgruppen, aber keine Gruppen. In $\mathbb{N}$ gibt es kein neutrales Element. In $\mathbb{N}_0$ gibt es zwar ein neutrales Element, nämlich die 0, aber zu keinem $n \neq 0$ ein inverses Element. Dagegen sind $(\mathbb{Z},+)$, $(\mathbb{Q},+)$, $(\mathbb{R},+)$ abelsche Gruppen mit dem neutralen Element 0.

(b) $(\mathbb{Z},\cdot)$, $(\mathbb{Q},\cdot)$, $(\mathbb{R},\cdot)$ sind kommutative Halbgruppen mit Neutralelement 1, aber keine Gruppen. In $(\mathbb{Z},\cdot)$ gibt es zu keinem $z \neq \pm 1$ ein inverses Element, in $(\mathbb{Q},\cdot)$ und $(\mathbb{R},\cdot)$ hat 0 kein Inverses. Aber $(\mathbb{Q} \setminus \{0\},\cdot)$ und $(\mathbb{R} \setminus \{0\},\cdot)$ sind abelsche Gruppen mit Neutralelement 1.

(c) Die Produktmengen $\mathbb{R}^n$, $\mathbb{Q}^n$, $\mathbb{Z}^n$, $n \in \mathbb{N}$, sind abelsche Gruppen, wenn die Addition + komponentenweise erklärt wird:

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

Dann ist $e = (0,\ldots,0)$ und $(x_1,\ldots,x_n)^{-1} = (-x_1,\ldots,-x_n)$.

(d) Ist A endlich, $A = \{a_1,\ldots,a_n\}$, so wird eine Verknüpfung $\circ$ oft durch eine Tabelle angegeben. Wir sprechen dann von einer *Verknüpfungstafel*:

$\circ$	a_1	a_2	$\cdots$	a_n
a_1	$*$	$*$	$\cdots$	$*$
a_2	$*$	$*$	$\cdots$	$*$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_n	$*$	$*$	$\cdots$	$*$

Anhand dieser Tabelle läßt sich direkt nachprüfen, ob die Gruppenaxiome erfüllt sind. Ist dies der Fall, so spricht man von einer *Gruppentafel*. Wegen Satz 1 darf dann in jeder Zeile und in jeder Spalte jedes Element von A nur einmal auftreten.

Beispiel. $A = \{0,1\}$ mit der Verknüpfungstafel

$+$	0	1
0	0	1
1	1	0

ist eine Gruppe.

(e) Es sei B eine nichtleere Menge. Die Menge B^B aller Abbildungen $f: B \longrightarrow B$ ist mit der Komposition $\circ$ als Verknüpfung eine Halbgruppe mit neutralem Element, nämlich der identischen Abbildung, aber im allgemeinen keine Gruppe. Für $|B| \geq 2$ hat die "konstante" Abbildung

$$\begin{aligned} f: B &\longrightarrow B \\ x &\longmapsto x_0 \end{aligned} \quad , \quad x_0 \in B \text{ fest},$$

kein Inverses. Ist dagegen f bijektiv, so ist die Umkehrabbildung f^{-1} von f das inverse Element zu f. Also ist

$$S_B := \{f \mid f: B \longrightarrow B \text{ bijektiv}\}$$

eine Gruppe, die im allgemeinen nicht abelsch ist. Sie heißt die *symmetrische Gruppe* von B.

Das letzte Beispiel wollen wir uns noch etwas genauer ansehen, und zwar für endliche Mengen B. Wir beschränken uns auf $B = \{1,\ldots,m\}$ und schreiben für die

symmetrische Gruppe S_B kurz S_m. Die Elemente von S_m heißen *Permutationen* und werden im folgenden meistens mit dem griechischen Buchstaben π bezeichnet.

Es ist üblich, Permutationen in Gestalt einer "Wertetabelle" anzugeben. Wir schreiben deshalb eine Permutation π in der Form

$$\pi = \begin{bmatrix} 1 & \cdots\cdots & m \\ \pi(1) & \cdots\cdots & \pi(m) \end{bmatrix}.$$

Der Name Permutation (= Vertauschung) rührt daher, daß bei einer bijektiven Abbildung $f\colon \{1,...,m\} \longrightarrow \{1,...,m\}$ die Elemente $1,...,m$ umgeordnet werden.

Satz 2. *Für die Anzahl der Elemente der Permutationsgruppe S_m gilt $|S_m| = m!$ (mit $m! := 1 \cdot 2 \cdot 3 \cdots m$).*

Beweis. Wir zeigen sogar etwas mehr: Es seien $A = \{a_1,...,a_m\}$, $B = \{b_1,...,b_m\}$ zwei Mengen mit m Elementen. Dann gilt: $|\{f \mid f\colon A \longrightarrow B \text{ bijektiv}\}| = m!$.

Diese Behauptung wird durch vollständige Induktion über $m \in \mathbb{N}$ bewiesen:
Induktionsanfang: $m = 1$: $A = \{a_1\}$, $B = \{b_1\}$. Hier gibt es genau eine bijektive Abbildung $f\colon A \longrightarrow B$, nämlich $a_1 \longmapsto b_1$.
Induktionsannahme: Die Behauptung sei richtig für $m-1$.
Induktionsschluß von $m-1$ auf m : Wir betrachten alle bijektiven Abbildungen $f\colon A \longrightarrow B$. Es gibt m Möglichkeiten für $f(a_m)$. Für ein festes $b_k \in B$ gibt es nach Induktionsannahme $(m-1)!$ bijektive Abbildungen von $\{a_1,...,a_{m-1}\}$ auf $B \setminus \{b_k\}$, also $(m-1)!$ bijektive Abbildungen $f\colon A \longrightarrow B$ mit $f(a_m) = b_k$. Somit gibt es insgesamt $(m-1)! \cdot m = m!$ bijektive Abbildungen von A auf B. ∎

Beispiel. Die Gruppe S_3 hat $3! = 6$ Elemente:

$$\pi_1 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}.$$

π_1 ist die identische Abbildung id. Wir vertauschen nun die Elemente 1,2,3 zyklisch und erhalten

$$\pi_2 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \quad \pi_3 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Die restlichen drei Permutationen lassen jeweils ein Element fest:

$$\pi_4 = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \quad \pi_5 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \quad \pi_6 = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}.$$

Man erhält die folgende Gruppentafel für S_3:

$\circ$	π_1	π_2	π_3	π_4	π_5	π_6
π_1	π_1	π_2	π_3	π_4	π_5	π_6
π_2	π_2	π_3	π_1	π_6	π_4	π_5
π_3	π_3	π_1	π_2	π_5	π_6	π_4
π_4	π_4	π_5	π_6	π_1	π_2	π_3
π_5	π_5	π_6	π_4	π_3	π_1	π_2
π_6	π_6	π_4	π_5	π_2	π_3	π_1

Es ist $\pi_2 \circ \pi_4 = \pi_6$ aber $\pi_4 \circ \pi_2 = \pi_5$, also ist die Gruppe S_3 nicht abelsch.

Die Permutationen π_4, π_5, π_6 sind von besonders einfacher Art. Sie heißen Transpositionen und man sieht, daß sich alle anderen Permutationen aus S_3 durch Verkettung dieser Transpositionen darstellen lassen.

So ist z.B. $\pi_1 = \pi_4 \circ \pi_4$, $\pi_2 = \pi_5 \circ \pi_6$ und $\pi_3 = \pi_5 \circ \pi_4$.

Allgemein nennen wir jede Permutation von $\{1,\ldots,m\}$, welche zwei Elemente i,j mit $i < j$ vertauscht und die restlichen festläßt, eine *Transposition.*
Schreibweise:

$$\tau^{(i,j)} = \begin{bmatrix} 1 & 2 & \cdots & i\text{–}1 & i & i\text{+}1 & \cdots & j\text{–}1 & j & j\text{+}1 & \cdots & m \\ 1 & 2 & \cdots & i\text{–}1 & j & i\text{+}1 & \cdots & j\text{–}1 & i & j\text{+}1 & \cdots & m \end{bmatrix}$$

Bemerkung. Es gilt $\tau^{(i,j)} \circ \tau^{(i,j)} = \mathrm{id}$, d.h. Transpositionen sind selbst–invers.

Satz 3. *Jede Permutation $\pi \in S_m$ $(m \geq 2)$ läßt sich als Verkettung endlich vieler Transpositionen darstellen.*

Beweis. Wir führen den Beweis durch vollständige Induktion nach m. Für $m = 2$ gilt:

$$S_2 = \{\pi_1, \pi_2\} \text{ mit } \pi_1 = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \text{ und } \pi_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

$\pi_2 = \tau^{(1,2)}$ ist eine Transposition und π_1 hat die Darstellung $\pi_1 = \pi_2 \circ \pi_2$.
Schluß von $m-1$ auf m: Sei $\pi \in S_m$. Ist $\pi(m) = m$, so ist

$$\tilde{\pi} = \begin{bmatrix} 1 & \cdots\cdots & m-1 \\ \pi(1) & \cdots\cdots & \pi(m-1) \end{bmatrix} \in S_{m-1}.$$

Nach Induktionsannahme existieren Transpositionen $\tilde{\tau}_1,\ldots,\tilde{\tau}_k \in S_{m-1}$ mit $\tilde{\pi} = \tilde{\tau}_1 \circ \cdots \circ \tilde{\tau}_k$. Jede Transposition $\tilde{\tau}_j \in S_{m-1}$ geht aber in eine Transposition $\tau_j \in S_m$ über, wenn wir

$$\tau_j(i) = \begin{cases} \tilde{\tau}_j(i) & \text{für } i = 1,\ldots,m-1 \\ m & \text{für } i = m \end{cases}$$

setzen. Damit erhalten wir $\pi = \tau_1 \circ \cdots \circ \tau_k$.

Ist $\pi(m) = n \neq m$, so ist $\tau^{(n,m)} \circ \pi = \pi'$ eine Permutation aus S_m, die m festläßt. Wegen $\pi = \tau^{(n,m)} \circ \pi'$ folgt mit dem zuvor bewiesenen Teil auch in diesem Fall die Behauptung. ∎

Beispiel. Für

$$\pi = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 2 & 6 & 4 & 1 \end{bmatrix} \in S_6$$

gilt

$$\pi = \tau^{(1,4)} \circ \tau^{(4,6)} \circ \tau^{(2,3)} \circ \tau^{(1,5)}.$$

Bemerkung und Definition. Die Darstellung $\pi = \tau_1 \circ \cdots \circ \tau_k$ ist nicht eindeutig, so gilt etwa in dem vorangehenden Beispiel auch

$$\pi = \tau^{(1,6)} \circ \tau^{(4,5)} \circ \tau^{(1,4)} \circ \tau^{(2,3)}.$$

Wir werden aber später (S. 164) sehen, daß bei verschiedenen solchen Darstellungen die Anzahl der auftretenden Transpositionen entweder immer gerade oder immer ungerade ist.

Man nennt deshalb eine Permutation π *gerade*, wenn es eine Darstellung mit einer geraden Anzahl von Transpositionen gibt, andernfalls heißt π *ungerade*.

Die Verkettung gerader Permutationen ergibt wieder eine gerade Permu-

tation. Da das inverse Element einer Transposition $\tau^{(i,j)}$ wieder eine Transposition ist (nämlich $\tau^{(i,j)}$ selbst), ist auch die inverse Permutation einer geraden Permutation wieder gerade. Das führt uns auf den Begriff der Untergruppe.

Definition. Es seien $(A,\circ)$ eine Gruppe und $B \subset A$. Genau dann heißt B eine *Untergruppe* von A, falls die auf $B \times B$ eingeschränkte Abbildung $\circ$ eine Verknüpfung ist und $(B,\circ)$ eine Gruppe.

Ist B eine Untergruppe von A, so sind die Neutralelemente e_B von B und e_A von A gleich. Es gilt nämlich $e_B = e_B \circ e_B = e_A \circ e_B$, woraus nach Satz 1 $e_B = e_A$ folgt. Damit ist notwendigerweise für jedes $x \in B$ das inverse Element bezüglich $\circ$ in B gleich dem Inversen bezüglich $\circ$ in A.

Diese Eigenschaften genügen nun auch, um nachzuweisen, daß eine Teilmenge $B \subset A$ eine Untergruppe von A ist.

Satz 4. *Es seien $(A,\circ)$ eine Gruppe und $B \subset A$. Dann ist B genau dann Untergruppe von A, wenn B das Neutralelement von A enthält und mit x und y stets auch x^{-1} und $x \circ y$ zu B gehören.*

Beweis. Wir müssen nur noch die eine Richtung der Behauptung zeigen. Die Abbildung $\circ : (x,y) \mapsto x \circ y$, eingeschränkt auf $B \times B$, bildet nach Voraussetzung $B \times B$ in B ab, ist also eine Verknüpfung auf B. Sie ist in B assoziativ, weil sie schon in A assoziativ ist. Also ist $(B,\circ)$ eine Halbgruppe und wegen der übrigen Voraussetzungen sogar eine Gruppe. Somit ist B Untergruppe von A. ∎

Bemerkung. Das Untergruppenkriterium wird oft auch in der folgenden, leicht abgewandelten Fassung benutzt, die jedoch zu Satz 4 äquivalent ist:

Es seien $(A,\circ)$ eine Gruppe und $B \subset A$. Dann ist B genau dann Untergruppe von A, wenn $B \neq \emptyset$ ist und mit x und y stets auch $x \circ y^{-1}$ zu B gehört.

Beweis. Wegen $B \neq \emptyset$ gibt es zunächst ein $x_0 \in B$. Also folgt $e = x_0 \circ x_0^{-1} \in B$. Mit

jedem $x \in B$ ist dann auch $x^{-1} = e \circ x^{-1} \in B$. Somit folgt aus $x, y \in B$ auch $x, y^{-1} \in B$ und damit nach Voraussetzung $x \circ y = x \circ (y^{-1})^{-1} \in B$. Nach Satz 4 ist dann B eine Untergruppe von A. Die umgekehrte Richtung ist offensichtlich. ∎

Beispiel. Die geraden Permutationen aus der Gruppe S_m bilden eine Untergruppe von S_m, die ungeraden dagegen nicht.

Wie immer im folgenden, wenn wir Mengen mit Strukturen betrachten, sind auch die Abbildungen von Interesse, die die Struktur erhalten. Sie heißen Homomorphismen.

Definition. Es seien $(A,\circ)$ und $(A',*)$ Gruppen und $f: A \longrightarrow A'$ eine Abbildung. f heißt (*Gruppen–*) *Homomorphismus*, wenn für alle $x,y \in A$ gilt: $f(x \circ y) = f(x) * f(y)$. Ein bijektiver Homomorphismus heißt *Isomorphismus*. Gibt es einen Isomorphismus $f: A \longrightarrow A'$, so heißen die Gruppen $(A,\circ)$ und $(A',*)$ *isomorph*, und wir schreiben dann $(A,\circ) \cong (A',*)$. Ist außerdem $A = A'$ und $\circ = *$, so heißt f *Automorphismus* von A.

Bemerkung. Jede Gruppe A ist isomorph zu einer geeigneten Untergruppe von S_A.

Bemerkungen. **(a)** Ein Gruppenhomomorphismus $f: A \longrightarrow A'$ besitzt folgende einfache Eigenschaften:

$f(e) = e'$, e' neutrales Element von A'.

$(f(x))^{-1} = f(x^{-1})$, $x \in A$.

(b) Zu jedem Homomorphismus $f: A \longrightarrow A'$ gehören in natürlicher Weise zwei Untergruppen:

$f(A) = \{f(x) \mid x \in A\}$ ist Untergruppe von A'.

Kern $f := \{x \in A \mid f(x) = e'\}$ ist Untergruppe von A.

(c) Der Homomorphismus f ist genau dann surjektiv, wenn $f(A) = A'$ und genau dann injektiv, wenn Kern $f = \{e\}$.

Wir beweisen nur den zweiten Teil von (b) und (c) und überlassen die restlichen Behauptungen als Übungsaufgaben:

(b) Es ist Kern $f \neq \emptyset$, da $e \in$ Kern f. Aus $x,y \in$ Kern f folgt $f(x \circ y^{-1}) = f(x) * f(y^{-1})$ $= e' * (f(y))^{-1} = (f(y))^{-1} = e'^{-1} = e'$, d.h. $x \circ y^{-1} \in$ Kern f. ∎

(c) Es sei f injektiv. Aus $x \in$ Kern f folgt $f(x) = e'$. Da auch $f(e) = e'$ gilt, erhalten wir $f(x) = f(e)$ und daraus $x = e$.

Ist umgekehrt Kern $f = \{e\}$, so folgt aus $f(x) = f(y)$ zunächst $f(x) * f(y)^{-1} = e'$ und daraus $f(x \circ y^{-1}) = e'$. Also gilt $x \circ y^{-1} \in$ Kern f, $x \circ y^{-1} = e$ und somit $x = y$. ∎

Die Untergruppe Kern f hat nun eine wichtige Eigenschaft. Es gilt nämlich für alle $x \in A$ und alle $y \in$ Kern f

$$(*) \qquad x \circ y \circ x^{-1} \in \text{Kern } f,$$

denn es ist $f(x \circ y \circ x^{-1}) = f(x) * e' * f(x)^{-1} = e'$.

Also ist Kern f invariant unter allen Abbildungen der Form

$$\begin{aligned} A &\longrightarrow A \\ y &\longmapsto x \circ y \circ x^{-1} \end{aligned} \quad , \; x \in A.$$

Diese Abbildungen sind, wie man leicht nachprüft, Automorphismen von A. Sie heißen *innere Automorphismen* von A. Die Eigenschaft (*) bedeutet somit, daß jeder innere Automorphismus von A die Untergruppe Kern f in sich abbildet, also *invariant* läßt. Wir nennen dies die Normalteilereigenschaft von Kern f.

Allgemein definieren wir:

Definition. Eine Untergruppe B einer Gruppe $(A,\circ)$ heißt *Normalteiler* von A, wenn für alle $x \in A$ und alle $y \in B$ stets $x \circ y \circ x^{-1} \in B$ gilt.

Beispiele. **(a)** A und $\{e\}$ sind triviale Normalteiler von A.

(b) Der Kern eines Homomorphismus ist Normalteiler.

(c) Jede Untergruppe einer abelschen Gruppe ist Normalteiler.

Satz 5. *Es seien $(A,\circ)$ eine Gruppe und B eine Untergruppe von A. Dann sind folgende Aussagen äquivalent:*

(a) *B ist Normalteiler.*

(b) *Für alle* $x \in A$ *gilt* $x \circ B \circ x^{-1} = B$.

(c) *Für alle* $x \in A$ *gilt* $x \circ B = B \circ x$.

(d) *Für alle* $x,x' \in A$ *und alle* $y,y' \in A$ *gilt: Aus* $x' \circ x^{-1} \in B$ *und* $y' \circ y^{-1} \in B$ *folgt stets* $x' \circ y' \circ y^{-1} \circ x^{-1} \in B$.

Hierbei ist $x \circ B \circ x^{-1} := \{x \circ b \circ x^{-1} \mid b \in B\}$, $x \circ B := \{x \circ b \mid b \in B\}$ und $B \circ x := \{b \circ x \mid b \in B\}$.

Beweis. (a) $\Longrightarrow$ (b): Wegen der Normalteilereigenschaft von B gilt offensichtlich $x \circ B \circ x^{-1} \subset B$. Sei nun $y \in B$. Dann ist $x^{-1} \circ y \circ (x^{-1})^{-1} \in B$ und somit

$$y = x \circ (x^{-1} \circ y \circ x) \circ x^{-1} \in x \circ B \circ x^{-1},$$

also $B \subset x \circ B \circ x^{-1}$.

(b) $\Longrightarrow$ (c): trivial

(c) $\Longrightarrow$ (a): trivial

(a) $\Longrightarrow$ (d): Mit $x' \circ x^{-1} \in B$ ist wegen der Normalteilereigenschaft von B auch $x^{-1} \circ x' = x'^{-1} \circ (x' \circ x^{-1}) \circ x' \in B$. Damit folgt aus $x' \circ x^{-1} \in B$ und $y' \circ y^{-1} \in B$ auch $x^{-1} \circ x' \circ y' \circ y^{-1} \in B$, woraus sich wieder wegen der Voraussetzung, daß B Normalteiler ist, die Behauptung

$$x' \circ y' \circ y^{-1} \circ x^{-1} = x \circ (x^{-1} \circ x' \circ y' \circ y^{-1}) \circ x^{-1} \in B$$

ergibt.

(d) $\Longrightarrow$ (a): Aus $x \in A$, $y \in B$ folgt für $x' = x$ und $y' = y \circ y$ zunächst

$$x' \circ x^{-1} = e \in B \text{ und } y' \circ y^{-1} = y \in B.$$

Damit sind die Voraussetzungen von (d) erfüllt, wir erhalten $x' \circ y' \circ y^{-1} \circ x^{-1} \in B$, d.h. $x \circ y \circ x^{-1} \in B$. ∎

Jede Untergruppe B einer Gruppe $(A,\circ)$ definiert auf dieser eine Äquivalenzrelation:

$$x \sim y :\Longleftrightarrow x \circ y^{-1} \in B \ , \ x,y \in A.$$

Zum Nachweis prüfen wir die drei definierenden Eigenschaften einer Äquiva-

lenzrelation nach.

Reflexivität: Für alle $x \in A$ gilt $x \sim x$, da $x \circ x^{-1} = e \in B$.

Symmetrie: Für alle $x,y \in A$ folgt aus $x \sim y$ stets $y \sim x$, da mit $x \circ y^{-1} \in B$ auch $y \circ x^{-1} = (x \circ y^{-1})^{-1} \in B$ gilt.

Transitivität: Für alle $x,y,z \in A$ folgt aus $x \sim y$, $y \sim z$ stets $x \sim z$, da mit $x \circ y^{-1} \in B$, $y \circ z^{-1} \in B$ auch $x \circ z^{-1} = (x \circ y^{-1}) \circ (y \circ z^{-1}) \in B$ gilt. ∎

Für die Faktormenge $A/_{\sim}$ schreiben wir, um den Zusammenhang von $\sim$ mit B zu betonen, auch $A/_B$.

Wir versuchen nun, diese Menge durch Definition einer geeigneten Verknüpfung $\cdot$, die mit der Verknüpfung $\circ$ von A zusammenhängen soll, zu einer Gruppe zu machen. Naheliegend ist der Ansatz

$$[x]_{\sim} \cdot [y]_{\sim} = [x \circ y]_{\sim} \quad , \; x,y \in A.$$

Dies ist aber nur dann eine sinnvolle Definition, wenn sie repräsentantenunabhängig ist, d.h. wenn für $x' \in [x]_{\sim}$, $y' \in [y]_{\sim}$ auch $x' \circ y' \in [x \circ y]_{\sim}$ gilt. Nach der Definition von $\sim$ bedeutet dies, daß aus $x' \circ x^{-1} \in B$, $y' \circ y^{-1} \in B$ stets $x' \circ y' \circ y^{-1} \circ x^{-1} \in B$ folgt, also nach Satz 5, daß B ein Normalteiler von A ist.

In der Gruppentheorie sind tatsächlich nur solche Faktormengen $A/_B$ von Interesse, bei denen B ein Normalteiler ist. Für diese Mengen gilt

Satz 6 und Definition. *Es seien $(A,\circ)$ eine Gruppe und $B \subset A$ ein Normalteiler. Dann ist die Faktormenge $A/_B$ zusammen mit der induzierten Verknüpfung $\cdot$*

$$[x]_{\sim} \cdot [y]_{\sim} = [x \circ y]_{\sim} \quad , \; x,y \in A,$$

eine Gruppe. $(A/_B, \cdot)$ heißt Quotienten– oder Faktorgruppe.

Beweis. Wir prüfen die Gruppeneigenschaften nach:

(a) Für alle $[x]_{\sim}$, $[y]_{\sim}$, $[z]_{\sim} \in A/_B$ gilt $([x]_{\sim} \cdot [y]_{\sim}) \cdot [z]_{\sim} = [x \circ y]_{\sim} \cdot [z]_{\sim} = [(x \circ y) \circ z]_{\sim} = [x \circ (y \circ z)]_{\sim} = [x]_{\sim} \cdot [y \circ z]_{\sim} = [x]_{\sim} \cdot ([y]_{\sim} \cdot [z]_{\sim})$.

(b) Neutrales Element von $A/_B$ ist $[e]_\sim$, weil $[e]_\sim \cdot [x]_\sim = [e \circ x]_\sim = [x]_\sim = [x \circ e]_\sim = [x]_\sim \cdot [e]_\sim$ für alle $[x]_\sim \in A/_B$ gilt. Weiterhin ist $[x^{-1}]_\sim$ inverses Element zu $[x]_\sim$, da $[x]_\sim \cdot [x^{-1}]_\sim = [x \circ x^{-1}]_\sim = [e]_\sim$ und $[x^{-1}]_\sim \cdot [x]_\sim = [e]_\sim$ gelten. ∎

Spezielle Normalteiler sind die Kerne von Gruppenhomomorphismen. Für die Faktorgruppen nach solchen Kernen ist der folgende Satz von besonderer Wichtigkeit.

Satz 7 (Homomorphiesatz für Gruppen). *Es seien $(A,\circ)$ und $(A',*)$ Gruppen sowie $f: A \longrightarrow A'$ ein Homomorphismus. Dann gilt:*

(a) $A/_{\text{Kern } f}$ *ist eine Faktorgruppe und die kanonische Abbildung*

$$k : A \longrightarrow A/_{\text{Kern } f}$$
$$x \longmapsto [x]_\sim$$

ist ein Homomorphismus.

(b) *Es gibt einen injektiven Homomorphismus $\bar{f} : A/_{\text{Kern } f} \longrightarrow A'$ mit $f = \bar{f} \circ k$.*

(c) *Ist f surjektiv, so sind $A/_{\text{Kern } f}$ und A' isomorph.*

Beweis. Die Behauptung (a) ist schon bewiesen. Bei (b) und (c) müssen wir wegen Satz 7 aus den Vorbemerkungen nur noch zeigen, daß die durch $\bar{f}([x]_\sim) := f(x)$ definierte Abbildung $\bar{f}$ ein Homomorphismus ist, was aber wegen

$$\bar{f}([x]_\sim \cdot [y]_\sim) = \bar{f}([x \circ y]_\sim) = f(x \circ y) = f(x) * f(y) = \bar{f}([x]_\sim) * \bar{f}([y]_\sim)$$

der Fall ist. ∎

Korollar 8. *Es gilt* $A/_{\text{Kern } f} \cong f(A)$.

§ 3 Körper und Ringe

Während der Gruppenbegriff sich auf nur eine Verknüpfung bezog, werden wir nun Mengen mit zwei Verknüpfungen betrachten und diese in Anlehnung an das übliche Zahlenrechnen mit + und · bezeichnen und Addition bzw. Multiplikation nennen. Neutralelemente werden mit 0 (bezüglich +) und 1 (bezüglich ·) bezeichnet und Inverse mit $-x$ (bezüglich +) sowie $x^{-1} = \frac{1}{x}$ (bezüglich ·).

Definition. Eine Menge A mit zwei Verknüpfungen + und · heißt *Körper*, wenn folgende Bedingungen erfüllt sind:

(a) $(A,+)$ ist eine abelsche Gruppe,

(b) $(A \setminus \{0\},\cdot)$ ist eine abelsche Gruppe,

(c) für alle $x,y,z \in A$ gilt: $x\cdot(y+z) = x\cdot y + x\cdot z$, $(x+y)\cdot z = x\cdot z + y\cdot z$. (*"Distributivgesetze"*)

Bemerkungen. **(a)** Es ist eine gängige Verabredung, statt $x\cdot y$ auch xy und statt $(x\cdot y) + z$ auch $xy + z$ zu schreiben. Letzteres bedeutet, daß die Multiplikation stärker binden soll als die Addition. Diese Konvention wurde schon bei der Formulierung der Distributivgesetze (c) benutzt. Außerdem schreiben wir statt xy^{-1} auch $\frac{x}{y}$.

(b) Das Axiom (c) läßt sich nicht aus den ersten beiden Axiomen folgern. Dies zeigt folgendes Beispiel:

Auf der Menge $A = \{0,1\}$ definieren wir Addition + und Multiplikation · durch die Verknüpfungstafeln

+	0	1
0	0	1
1	1	0

·	0	1
0	1	0
1	0	1

Dann sind $(A,+)$ und $(A \setminus \{0\},\cdot)$ abelsche Gruppen, aber es gilt keines der beiden Distributivgesetze.

(c) Ein Körper besitzt mindestens zwei Elemente, nämlich 0 und 1.

(d) Für alle $x \in A$ gilt $x\cdot 0 = 0\cdot x = 0$.

Aus $x\cdot x = x(x+0) = x\cdot x + x\cdot 0$ folgt nämlich $x\cdot 0 = 0$ und aus $x\cdot x = (x+0)x = x\cdot x + 0\cdot x$ folgt $0\cdot x = 0$. ∎

(e) Für alle $x,y \in A$ gilt $x(-y) = (-x)y = -(xy)$.
Aus $xy + x(-y) = x(y-y) = x\cdot 0 = 0$ folgt nämlich $x(-y) = -(xy)$ und aus $xy + (-x)y = (x-x)y = 0\cdot y = 0$ folgt $(-x)y = -(xy)$. ∎

(f) Bezüglich der Multiplikation gelten das Assoziativgesetz, die Neutraleigenschaft der 1 und das Kommutativgesetz zunächst nur für Elemente, die von 0 verschieden sind. Wegen (d) gelten diese Eigenschaften aber uneingeschränkt.

(g) In einem Körper folgt aus $xy = 0$ stets $x = 0$ oder $y = 0$, denn $A \setminus \{0\}$ ist bezüglich der Multiplikation abgeschlossen, d.h. aus $x \neq 0$ und $y \neq 0$ folgt $xy \neq 0$. Man sagt auch, ein Körper sei *nullteilerfrei*.

Beispiele. **(a)** $(\mathbb{Z},+,\cdot)$ ist kein Körper, da es in $(\mathbb{Z} \setminus \{0\},\cdot)$ zu $z \neq \pm 1$ kein inverses Element gibt. Dagegen sind $\mathbb{Q}$ und $\mathbb{R}$ Körper.

(b) $A = \{x + \sqrt{2}\, y \mid x,y \in \mathbb{Q}\}$ ist bezüglich der Addition und Multiplikation reeller Zahlen ein Körper, der echt zwischen $\mathbb{Q}$ und $\mathbb{R}$ liegt (Übungsaufgabe).

(c) $A = \{0,1\}$ ist mit den durch die folgenden Verknüpfungstafeln gegebenen Verknüpfungen

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

ein Körper und zwar der "kleinste" Körper.

Weitere endliche Körper sind die sogenannten *Restklassenkörper*:

Es sei m eine feste natürliche Zahl. Wir definieren auf $\mathbb{Z}$ eine Relation $\sim$ durch

$$x \sim y :\Longleftrightarrow x - y \in m\cdot\mathbb{Z} := \{mz \mid z \in \mathbb{Z}\}.$$

Es ist leicht nachzuprüfen, daß $\sim$ eine Äquivalenzrelation ist und daß x und y genau dann äquivalent sind, wenn sie bei der Division durch m den gleichen Rest $r \in \{0,...,m-1\}$ besitzen. Die Äquivalenzklassen bezüglich $\sim$ heißen deshalb auch *Restklassen.*

Für die Menge $\mathbb{Z}/_{\sim}$ der Äquivalenzklassen schreiben wir $\mathbb{Z}_m$. Da es zu jedem $x \in \mathbb{Z}$ ein $z \in \mathbb{Z}$ und ein $r \in \{0,...,m-1\}$ gibt mit $x = mz + r$, ist $[x]_\sim = [r]_\sim$ und daher $\mathbb{Z}_m = \{[0]_\sim, ..., [m-1]_\sim\}$.

Wir wollen $\mathbb{Z}_m$ nun mit einer Addition und einer Multiplikation versehen. Dazu definieren wir

$$[x]_\sim + [y]_\sim := [x+y]_\sim$$

und

$$[x]_\sim \cdot [y]_\sim := [xy]_\sim .$$

Zunächst müssen wir die Wohldefiniertheit dieser Verknüpfungen nachprüfen, d.h. mit $x' \in [x]_\sim$, $y' \in [y]_\sim$ muß auch $x'+y' \in [x+y]_\sim$ und $x'y' \in [xy]_\sim$ gelten:
Aus $x' - x = mz$, $y' - y = mz'$ folgen $(x'+y') - (x+y) = m(z+z')$ und $x'y' - xy = m(zy + z'x + mzz')$. ∎

Die meisten Rechenregeln übertragen sich nun sofort von $\mathbb{Z}$ auf $\mathbb{Z}_m$. Es folgt, daß $(\mathbb{Z}_m, +)$ eine abelsche Gruppe ist mit dem neutralen Element $[0]_\sim$, daß in $(\mathbb{Z}_m, \cdot)$ Assoziativgesetz und Kommutativgesetz gelten, daß $[1]_\sim$ neutrales Element der Multiplikation ist und daß die Distributivgesetze gelten. Trotzdem ist für gewisse m $(\mathbb{Z}_m \setminus \{[0]_\sim\}, \cdot)$ keine abelsche Gruppe und damit $\mathbb{Z}_m$ kein Körper.

In $\mathbb{Z}_1$ zum Beispiel gilt wegen $1 - 0 \in 1 \cdot \mathbb{Z}$, daß $[0]_\sim = [1]_\sim$, also ist $\mathbb{Z}_1 \setminus \{[0]_\sim\} = \emptyset$. In $\mathbb{Z}_4$ gilt wegen $4 - 0 \in 4 \cdot \mathbb{Z}$, daß $[2]_\sim \cdot [2]_\sim = [0]_\sim$, also ist die Multiplikation auf $\mathbb{Z}_4 \setminus \{[0]_\sim\}$ keine Verknüpfung.

Das letzte Beispiel läßt sich verallgemeinern: Ist nämlich $m \geq 2$ keine Primzahl, so gibt es natürliche Zahlen $p,q \in \{2,...,m-1\}$ mit $m = p \cdot q$. Dann ist $[p]_\sim \neq [0]_\sim$, $[q]_\sim \neq [0]_\sim$, aber $[p]_\sim \cdot [q]_\sim = [p \cdot q]_\sim = [m]_\sim = [0]_\sim$, ein Wider-

spruch zur Nullteilerfreiheit in einem Körper, also kann $\mathbb{Z}_m$ dann kein Körper sein.

Bevor wir jedoch die Frage beantworten, für welche $m \in \mathbb{N}$ wir einen Körper erhalten, wollen wir einige Beispiele für das Rechnen in $\mathbb{Z}_m$ geben.

Beispiele. **(a)** $\mathbb{Z}_2$ hat genau zwei Elemente, nämlich $[0]_\sim = \{0 + 2z |\ z \in \mathbb{Z}\}$ und $[1]_\sim = \{1 + 2z |\ z \in \mathbb{Z}\}$.
$\mathbb{Z}_9 = \{[0]_\sim , [1]_\sim ,\ldots,[8]_\sim\}$ mit $[0]_\sim = \{9z |\ z \in \mathbb{Z}\}$, $[1]_\sim = \{1 + 9z |\ z \in \mathbb{Z}\}$, $[2]_\sim = \{2 + 9z |\ z \in \mathbb{Z}\}$, ... , $[8]_\sim = \{8 + 9z |\ z \in \mathbb{Z}\}$.

(b) Was ist $[137]_\sim$ in $\mathbb{Z}_{15}$? Wegen $137 = 9 \cdot 15 + 2$ ist $137 \sim 2$, d.h. $[137]_\sim = [2]_\sim$.
Was ist $[-79]_\sim$ in $\mathbb{Z}_8$? Wegen $-79 = (-10) \cdot 8 + 1$ ist $[-79]_\sim = [1]_\sim$.

Satz 9. *Genau dann ist $\mathbb{Z}_m$, $m \in \mathbb{N}$, ein Körper, wenn m eine Primzahl ist.*

Beweis. Ist $\mathbb{Z}_m$ ein Körper, so muß nach dem oben gesagten m eine Primzahl sein.

Sei jetzt m eine Primzahl. Wir müssen nur noch zeigen, daß die Multiplikation eine Verknüpfung auf $\mathbb{Z}_m \setminus \{[0]_\sim\}$ ist und daß jedes Element $[x]_\sim \neq [0]_\sim$ in $(\mathbb{Z}_m \setminus \{[0]_\sim\}, \cdot)$ ein inverses Element besitzt. Dazu beweisen wir zunächst, daß für alle $[x]_\sim , [y]_\sim , [z]_\sim \in \mathbb{Z}_m$ mit $[x]_\sim \neq [0]_\sim$ aus $[x]_\sim \cdot [y]_\sim = [x]_\sim \cdot [z]_\sim$ stets $[y]_\sim = [z]_\sim$ folgt:

$[x]_\sim \cdot [y]_\sim = [x]_\sim \cdot [z]_\sim$ ist äquivalent zu $[xy - xz]_\sim = [0]_\sim$. Somit gilt $x(y - z) = km$, $k \in \mathbb{Z}$. Wegen $[x]_\sim \neq [0]_\sim$ ist m kein Teiler von x. Da m jedoch das Produkt $x(y - z)$ teilt, muß m Teiler von $y - z$ sein, woraus sich $[y - z]_\sim = [0]_\sim$ und $[y]_\sim = [z]_\sim$ ergeben.

Aus $[x]_\sim \neq [0]_\sim$, $[y]_\sim \neq [0]_\sim$ folgt nun $[x]_\sim \cdot [y]_\sim \neq [0]_\sim$. Damit ist $\cdot$ eine Verknüpfung auf $\mathbb{Z}_m \setminus \{[0]_\sim\}$. Außerdem sind für $[x]_\sim \neq [0]_\sim$ die Produkte $[x]_\sim \cdot [1]_\sim$, $[x]_\sim \cdot [2]_\sim$,...,$[x]_\sim \cdot [m-1]_\sim$ paarweise verschieden; eines von ihnen muß daher gleich $[1]_\sim$ sein. Also besitzt jedes Element $[x]_\sim \in \mathbb{Z}_m \setminus \{[0]_\sim\}$ ein inverses Element bezüglich der Multiplikation. ∎

Bemerkung. Um die Abhängigkeit der Äquivalenzrelation $\sim$ von der festgewählten Zahl $m \in \mathbb{N}$ deutlich zu machen, schreibt man statt $x \sim y$ auch $x \equiv y \bmod m$ und

sagt, *x ist kongruent y modulo m* .

Beispiel. Wir suchen alle ganzen Zahlen x, für die $5 \equiv x \bmod 12$ gilt:

Das Rechnen mit Kongruenzen ergibt $x - 5 \equiv 0 \bmod 12$, $x - 5 = k \cdot 12$, $k \in \mathbb{Z}$, und somit $x = 5 + k \cdot 12$, $k \in \mathbb{Z}$.

Die Rechnung in $\mathbb{Z}_{12}$ ergibt $[5]_{\sim} = [x]_{\sim}$, $[x-5]_{\sim} = [0]_{\sim}$, also $x - 5 = k \cdot 12$, $k \in \mathbb{Z}$, und somit ebenfalls $x = 5 + k \cdot 12$, $k \in \mathbb{Z}$.

Das Rechnen mit Kongruenzen und insbesondere das Lösen von Systemen von Kongruenzgleichungen gehört in den Bereich der Zahlentheorie, spielt aber auch in der Informatik eine wichtige Rolle. Wir werden deshalb in Paragraph 1.6 einige Aussagen über Kongruenzen, wie etwa den Chinesischen Restsatz, herleiten.

Wie bei den Gruppen sollen nun auch bei den Körpern die strukturerhaltenden Abbildungen eingeführt werden.

Definition. Es seien $(A,+,\cdot)$ und $(A',+',\cdot')$ Körper und $f\colon A \longrightarrow A'$ eine Abbildung mit den Eigenschaften

$$f(x+y) = f(x) +' f(y) \ , \ x,y \in A \ ,$$

$$f(x \cdot y) = f(x) \cdot' f(y) \ , \ x,y \in A$$

und

$$f(1) = 1 \ .$$

Dann heißt f *(Körper–) Homomorphismus.* Ist f bijektiv, so heißt f *Isomorphismus* und die Körper A und A' heißen *isomorph* ; wir schreiben dann $A \cong A'$.

Bemerkungen und Bezeichnungen. **(a)** Es hat sich in der Algebra eingebürgert, für endliche Körper mit q Elementen, $q \in \mathbb{N}$, die Bezeichnungen $\mathbb{F}_q$ (für "field") oder $GF(q)$ (für "Galoisfeld" nach Evariste Galois, 1811 – 1832) zu verwenden. Es läßt sich zeigen, daß genau für die Primzahlpotenzen $q = p^n$, p Primzahl, $n \in \mathbb{N}$, solche endlichen Körper existieren und daß je zwei Körper mit q Elementen isomorph sind. Für $q = p$ sind daher alle Körper mit p Elementen zu $\mathbb{Z}_p$ isomorph, insbesondere auch der Körper $\mathbb{F}_p = \{0,\ldots,p-1\}$ mit den von $\mathbb{Z}_p$ induzierten Verknüpfungen. Der kleinste Körper ist $\mathbb{F}_2 = \{0,1\}$ aus Beispiel (c).

(b) In $\mathbb{F}_p$ gilt

$$\underbrace{1 + \cdots + 1}_{p \text{ Summanden}} = 0$$

und p ist die kleinste natürliche Zahl mit dieser Eigenschaft. Man nennt p die *Charakteristik* des Körpers $\mathbb{F}_p$.

Analog nennt man bei einem beliebigen Körper $\mathbb{K}$, sofern ein derartiges p existiert, diese Zahl die *Charakteristik* von $\mathbb{K}$. Gilt jedoch in $\mathbb{K}$ für alle $m \in \mathbb{N}$

$$\underbrace{1 + \cdots + 1}_{m \text{ Summanden}} \neq 0 ,$$

so sagt man, der Körper $\mathbb{K}$ habe die *Charakteristik* 0.

So sind zum Beispiel $\mathbb{Q}$ und $\mathbb{R}$ Körper der Charakteristik 0. Die Charakteristik eines Körpers ist entweder 0 oder eine Primzahl. (Übungsaufgabe).

(c) Da wir im folgenden meist einen beliebigen Körper zugrundelegen, wollen wir für Körper – so wie wir es schon in (b) getan haben – in Zukunft das Symbol $\mathbb{K}$ benutzen (analog zu $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{F}_p$). Für die Charakteristik schreiben wir dann char $\mathbb{K}$.

Wir wollen die Körpertheorie nicht weiter ausbauen, sondern uns hier damit begnügen, noch ein weiteres Beispiel für einen Körper kennenzulernen, nämlich den Körper $\mathbb{C}$ der komplexen Zahlen, der für die lineare Algebra besonders wichtig ist.

Komplexe Zahlen

Ausgangspunkt unserer Betrachtungen ist die abelsche Gruppe $(\mathbb{R}^2,+)$. Auf der Menge $\mathbb{R}^2$ werden wir eine Multiplikation "$\cdot$" so definieren, daß $(\mathbb{R}^2,+,\cdot)$ ein Körper ist.

Definition. Für alle $(a,b), (c,d) \in \mathbb{R}^2$ sei

$$(a,b)\cdot(c,d) := (ac - bd, ad + bc)$$

Diese Multiplikation ist offensichtlich kommutativ. Sie ist auch assoziativ:

$$\begin{aligned}[(a,b)\cdot(c,d)]\cdot(e,f) &= (ac - bd, ad + bc)\cdot(e,f) \\ &= (eac - ebd - adf - bcf,\ acf - bdf + ade + bce) ,\end{aligned}$$

$$(a,b)\cdot[(c,d)\cdot(e,f)] = (a,b)\cdot(ce - df, cf + de)$$
$$= (ace - adf - bcf - bde\ ,\ acf + ade + bce - bdf)\ .$$

Es gelten die Distributivgesetze. Wegen der Kommutativität der Multiplikation auf $\mathbb{R}^2$ genügt es, eines davon nachzurechnen:

$$(a,b)\cdot[(c,d) + (e,f)] = (a,b)\cdot(c + e, d + f)$$
$$= (ac + ae - bd - bf,\ ad + af + bc + be)\ ,$$

$$(a,b)\cdot(c,d) + (a,b)\cdot(e,f) = (ac - bd, ad + bc) + (ae - bf, af + be)$$
$$= (ac - bd + ae - bf\ ,\ ad + bc + af + be).$$

Die Multiplikation ist eine Verknüpfung auf $\mathbb{R}^2 \setminus \{(0,0)\}$:

Seien $(a,b) \neq (0,0)$ und $(c,d) \neq (0,0)$. Wäre $(a,b)\cdot(c,d) = (0,0)$, so folgte $ac - bd = 0$, $ad + bc = 0$ und somit $0 = (ac - bd)^2 + (ad + bc)^2 = (a^2 + b^2)\cdot(c^2 + d^2)$. Also wäre $a^2 + b^2 = 0$ und daher $(a,b) = (0,0)$ oder $c^2 + d^2 = 0$ und somit $(c,d) = (0,0)$. In beiden Fällen ergibt sich ein Widerspruch zur Voraussetzung.

Das neutrale Element bezüglich der Multiplikation ist $(1,0)$, denn es gilt für alle $(a,b) \in \mathbb{R}^2$, daß $(a,b)\cdot(1,0) = (a,b)$.

Das inverse Element zu $(a,b) \neq (0,0)$ ist

$$\left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right).$$

Insgesamt haben wir folgenden Satz bewiesen:

Satz 10. $(\mathbb{R}^2,+,\cdot)$ *ist ein Körper.*

$(\mathbb{R}^2,+,\cdot)$ heißt Körper der *komplexen Zahlen.* Wir verwenden in Zukunft die Bezeichnung $\mathbb{C}$. Die Elemente $z = (a,b) \in \mathbb{C}$ heißen *komplexe Zahlen* ; a heißt *Realteil* und b heißt *Imaginärteil* von z .

Die Teilmenge $\mathbb{R} \times \{0\} = \{(a,0) |\ a \in \mathbb{R}\}$ von $\mathbb{C}$ ist bezüglich der in $\mathbb{C}$ erklärten Addition und Multiplikation selbst ein Körper, und die Abbildung

$$f\colon \mathbb{R} \longrightarrow \mathbb{R} \times \{0\}$$
$$a \longmapsto (a,0)$$

ist ein Körperisomorphismus. Die reellen Zahlen sind also in den Körper der komplexen Zahlen eingebettet, und wir können sie als spezielle komplexe Zahlen auffassen.

Jede komplexe Zahl $z = (a,b)$ läßt sich nun folgendermaßen darstellen:

$$z = (a,b) = (a,0) + (0,b) = (a,0) + (0,1)\cdot(b,0).$$

Verwenden wir die Abkürzung

$$i := (0,1)$$

und beachten wir, daß a mit $(a,0)$ und b mit $(b,0)$ identifiziert wird, so gilt

$$z = a + ib.$$

Ist $z \neq 0$, so schreiben wir für z^{-1} auch $\frac{1}{z} = \frac{1}{a + ib}$.

In dieser Darstellung können wir mit den komplexen Zahlen wie im Körper $\mathbb{R}$ rechnen. Wir müssen nur beachten, daß $i^2 = -1$ gilt.

Beispiel. Sei $z = a + ib$, $w = c + id$. Dann ist

$$z\cdot w = (a + ib)\cdot(c + id) = ac + ibc + iad + i^2bd = (ac - bd) + i(bc + ad).$$

Ist $z \neq 0$, so gilt

$$\frac{1}{z} = \frac{1}{a + ib} = \frac{a - ib}{(a + ib)(a - ib)} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}.$$

Mit $z = a + ib \in \mathbb{C}$ ist auch $\overline{z} := a - ib \in \mathbb{C}$. $\overline{z}$ heißt die zu z *konjugiert komplexe Zahl*.
Es gelten folgende Rechenregeln:

(a) $\overline{z + w} = \overline{z} + \overline{w}$ und $\overline{z\cdot w} = \overline{z}\cdot\overline{w}$, $z,w \in \mathbb{C}$,

(b) $\overline{(\overline{z})} = z$, $z \in \mathbb{C}$,

(c) $z \in \mathbb{R} \iff z = \overline{z}$, $z \in \mathbb{C}$,

(d) $z\cdot\overline{z} \in \mathbb{R}$, $z\cdot\overline{z} \geq 0$ und $(z\cdot\overline{z} = 0 \iff z = 0)$, $z \in \mathbb{C}$.

Die reelle Zahl $|z| := (z\cdot\overline{z})^{1/2}$ heißt der *Betrag* von z.

Bemerkung. Die Abbildung $z \longmapsto \overline{z}$, $z \in \mathbb{C}$, ist wegen (b) offensichtlich bijektiv. Wegen (a) und (b) ist sie ein Isomorphismus von $\mathbb{C}$ in sich, also ein Automorphismus, und wegen (c) läßt sie den Körper $\mathbb{R}$ elementweise fest. Sie ist neben $\mathrm{id}_{\mathbb{C}}$ der einzige Automorphismus von $\mathbb{C}$ mit dieser Eigenschaft (Beweis als Übungsaufgabe).

Nun wenden wir uns algebraischen Strukturen mit zwei Verknüpfungen zu, die keine Körper sind. Beispiele dafür sind die Mengen $\mathbb{Z}_m$, wenn m keine Primzahl ist. Dann ist die Multiplikation auf $\mathbb{Z}_m$ zwar immer noch assoziativ und es gelten die Distributivgesetze, aber $\mathbb{Z}_m$ ist kein Körper mehr. Wir sprechen in diesem Fall von einem Ring.

Definition. Eine Menge A mit zwei Verknüpfungen $+$ und $\cdot$ heißt *Ring*, wenn folgende Bedingungen erfüllt sind:

(a) $(A,+)$ ist eine abelsche Gruppe,

(b) $(A,\cdot)$ ist eine Halbgruppe,

(c) für alle $x,y,z \in A$ gilt: $x\cdot(y + z) = x\cdot y + x\cdot z$, $(x + y)\cdot z = x\cdot z + y\cdot z$. (*"Distributivgesetze"*)

Ist $(A,\cdot)$ auch kommutativ, so sprechen wir von einem *kommutativen Ring*.

Besitzt $(A,\cdot)$ ein Neutralelement 1 (*Einselement*), so sprechen wir von einem *Ring mit* 1.

Beispiele. **(a)** Jeder Körper ist ein kommutativer Ring mit 1.

(b) $\mathbb{Z}_m$ ist für jedes $m \in \mathbb{N}$ ein kommutativer Ring mit 1. Falls m keine Primzahl ist, sprechen wir von dem *Restklassenring* $\mathbb{Z}_m$.

(c) $\mathbb{Z}$ ist ein kommutativer Ring mit 1.

(d) $(\mathcal{P}(A),\Delta,\cap)$ ist ein kommutativer Ring mit 1 (Übungsaufgabe).

Die nächsten Beispiele sind besonders wichtig. Wir wollen sie deshalb in einem eigenen Paragraphen behandeln.

§ 4 Matrizen und Polynome

Zunächst wollen wir uns mit den Matrizen beschäftigen.

Definition. Seien $m,n \in \mathbb{N}$. Eine (m,n)–*Matrix* A über einem Körper $\mathbb{K}$ ist ein $m \cdot n$–Tupel von Elementen von $\mathbb{K}$, das als rechteckiges Schema mit m Zeilen und n Spalten angeordnet ist:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1n} \\ a_{21} & a_{22} & \cdots\cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots\cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{K}.$$

Andere Schreibweisen:

$$A = (\!(a_{ij})\!)_{m \times n} \text{ oder } A = (\!(a_{ij})\!) \text{ oder auch } A = (a_{ij}).$$

$(1,n)$–Matrizen heißen *Zeilen*, $(m,1)$–Matrizen heißen *Spalten*. (n,n)–Matrizen heißen *n–reihige quadratische Matrizen* (über $\mathbb{K}$). Speziell heißt die (n,n)–Matrix

$$E_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots\cdots & 0 & 0 & 1 \end{bmatrix}$$

n–reihige Einheitsmatrix. Unter Verwendung des Kroneckersymbols

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

können wir auch kurz $E_n = (\!(\delta_{ij})\!)$ schreiben. Mit $\mathbb{K}^{m \times n}$ bezeichnen wir die Menge aller (m,n)–Matrizen über dem Körper $\mathbb{K}$.

Bemerkung. Da eine (m,n)–Matrix über $\mathbb{K}$ im Prinzip ein Element von $\mathbb{K}^{mn}$ in anderer Schreibweise ist, ergibt sich sofort, wann zwei Matrizen gleich sind.

Außerdem erhalten wir, daß $\mathbb{K}^{m\times n}$ mit der komponentenweisen Addition eine abelsche Gruppe ist.

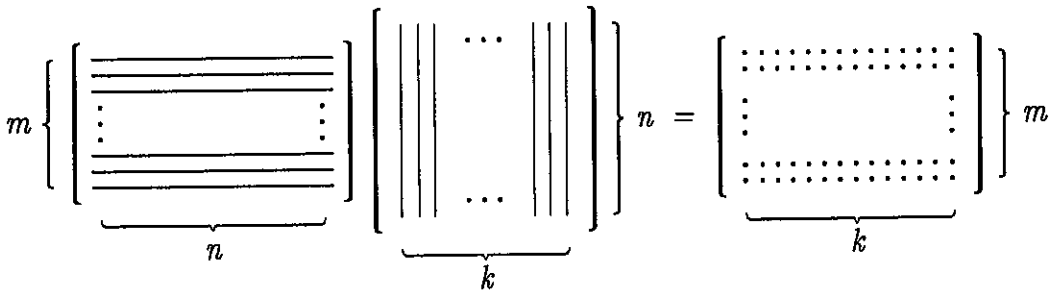
Nun wollen wir für Matrizen eine Multiplikation erklären.

Definition. Es seien $A \in \mathbb{K}^{m\times n}$ und $B \in \mathbb{K}^{n\times k}$. Dann heißt die Matrix $(\!(c_{ij})\!) \in \mathbb{K}^{m\times k}$, die durch

$$c_{ij} := \sum_{l=1}^{n} a_{il}\, b_{lj} \qquad i = 1,\ldots,m, \quad j = 1,\ldots,k$$

definiert ist, das *Produkt AB von A und B.*

Merkschema zur Matrizenmultiplikation ("i–te Zeile von A mal j–te Spalte von B ergibt das Element c_{ij} von C"):



Beispiel. Für $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$, $B = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}$ ist

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix},$$

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

Diese Art der Matrizenmultiplikation scheint zunächst etwas merkwürdig zu sein. Naheliegender wäre ja wohl, die Multiplikation, analog zur Addition, komponentenweise zu erklären. Wie wir später jedoch sehen werden, haben Matrizen etwas mit Abbildungen zu tun, und die Multiplikation von Matrizen ist deshalb so definiert,

daß sie mit der Verkettung der zugehörigen Abbildungen verträglich ist.

Satz 11. *Die folgenden Matrizenprodukte und Matrizensummen seien jeweils erklärt (über* $\mathbb{K}$*). Dann gilt:*

(a) $(AB)C = A(BC)$, *("Assoziativität")*

(b) $(A + B)C = AC + BC$, *("Distributivität")*

$A(B + C) = AB + AC$,

(c) $E_n A = A, \quad AE_n = A$.

Beispiel zur Interpretation. Das Produkt $E_n A$ ist erklärt für alle $A \in \mathbb{K}^{n\times k}$, das Produkt AE_n dagegen für alle $A \in \mathbb{K}^{m\times n}$. Also gilt $E_n A = A$ für alle $A \in \mathbb{K}^{n\times k}$ und $AE_n = A$ für alle $A \in \mathbb{K}^{m\times n}$.

Beweis. Zum Beweis wird man natürlich die entsprechenden Eigenschaften des Körpers $\mathbb{K}$ ausnutzen.

(a) In der i–ten Zeile von AB steht an der k–ten Stelle das Element $\sum_l a_{il}\, b_{lk}$, $k = 1,2,\ldots$; das (i,j)–te Element der Matrix $(AB)C$ ist daher $\sum_k (\sum_l a_{il}\, b_{lk}) c_{kj}$.

Nutzen wir in $\mathbb{K}$ die Kommutativität bezüglich $+$, die Assoziativität bezüglich $\cdot$ und die Distributivität aus, so können wir umformen:

$$\sum_k (\sum_l a_{il}\, b_{lk}) c_{kj} = \sum_k \sum_l (a_{il}\, b_{lk}\, c_{kj}) = \sum_l \sum_k (a_{il}\, b_{lk}\, c_{kj}) = \sum_l a_{il} (\sum_k b_{lk}\, c_{kj}).$$

Es ist $\sum_k b_{lk}\, c_{kj}$ das l–te Element in der j–ten Spalte von BC. Somit gibt das letzte Element das (i,j)–te Element von $A(BC)$ an.

(b) (i,j)–tes Element von $(A + B)C$: $\sum_k (a_{ik} + b_{ik}) c_{kj}$,

(i,j)–tes Element von AC: $\sum_k a_{ik}\, c_{kj}$,

(i,j)–tes Element von BC: $\sum_k b_{ik}\, c_{kj}$.

Das andere Distributivgesetz wird analog bewiesen.

(c) (i,j)–tes Element von $E_n A$: $\sum_k \delta_{ik}\, a_{kj} = a_{ij}$,

(i,j)–tes Element von AE_n : $\sum_k a_{ik}\, \delta_{kj} = a_{ij}$. ■

Bemerkung. Für beliebige Matrizen haben die Produkte AB und BA (wenn beide existieren) meist verschiedenes Format. Aber auch für quadratische Matrizen gilt im allgemeinen **nicht** $AB = BA$.

Beispiel. Für die (n,n)–Matrizen

$$A = \begin{bmatrix} 1\,0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & \cdots & 0\,1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix},$$

folgt

$$AB = \begin{bmatrix} 0 & \cdots & 0\,1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}.$$

Aus Satz 11 erhalten wir nun sofort das folgende Ergebnis:

Satz 12. *Die Menge* $\mathbb{K}^{n\times n}$ *aller* n*–reihigen quadratischen Matrizen über* $\mathbb{K}$ *bildet einen Ring mit* 1.

Bemerkungen und Definitionen. **(a)** Das obige Beispiel zeigt, daß $\mathbb{K}^{n\times n}$ für keinen Körper $\mathbb{K}$ und kein $n \geq 2$ ein Körper ist.

(b) Ist eine Matrix A in der Halbgruppe $(\mathbb{K}^{n\times n}, \cdot)$ invertierbar, so heißt A *regulär*. Ist A nicht regulär, so heißt A *singulär*.

Die Menge $GL(n,\mathbb{K})$ der regulären (n,n)–Matrizen ist eine Gruppe bezüglich der Matrizenmultiplikation. Sie heißt *allgemeine lineare Gruppe* (über $\mathbb{K}$).

Die obigen Matrizen A und B sind für $n \geq 2$ singulär, da andernfalls aus $BA = O$ folgte, daß A oder B die Nullmatrix wäre. Dagegen ist die (n,n)–Matrix

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots\cdots & 0 & 1 \end{bmatrix}$$

regulär, denn

$$A^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots\cdots & 0 & 1 \end{bmatrix}$$

ist die zu A inverse Matrix.

Man beachte, daß auch die Menge der regulären (n,n)–Matrizen für $n \geq 2$ keinen Körper bildet, da die Summe zweier regulärer Matrizen keine reguläre Matrix sein muß. So ist z.B. für die obige reguläre Matrix A die Matrix

$$A + (-E_n) = \begin{bmatrix} 0 & \cdots\cdots & 0 & 1 \\ 0 & \cdots\cdots & & 0 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 0 & \cdots\cdots & & 0 \end{bmatrix}$$

nicht regulär.

(c) Später werden wir ein Verfahren kennenlernen, mit dem man prüfen kann, ob eine Matrix regulär ist, und mit dem man im Fall der Regularität auch die inverse Matrix berechnen kann.

(d) Sei $A = (\!(a_{ij})\!) \in \mathbb{K}^{m\times n}$. Dann heißt die Matrix $(\!(b_{ij})\!) \in \mathbb{K}^{n\times m}$ mit $b_{ij} = a_{ji}$ die *transponierte Matrix* zu A ; Schreibweise: A^T.

A^T ist also die an der "Diagonalen a_{11}, a_{22},... gespiegelte" Matrix A.

Es gelten folgende Rechenregeln:

(i) $(A^T)^T = A$, $(A + B)^T = A^T + B^T$, $(AB)^T = B^T A^T$.

(ii) Ist A regulär, so auch A^T und es gilt $(A^T)^{-1} = (A^{-1})^T$.

(e) Eine Matrix A mit $A = A^\top$ heißt *symmetrisch*. Es können natürlich nur quadratische Matrizen symmetrisch sein. Die Summe symmetrischer Matrizen ist wieder symmetrisch, das Produkt im allgemeinen aber nicht. So ist beispielsweise das Produkt

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

nicht symmetrisch.

Eine weitere Matrizenoperation, die später noch eine wichtige Rolle spielen wird, ist die Multiplikation von Matrizen mit Körperelementen.

Definition. Es seien $A = (\!(a_{ij})\!) \in \mathbb{K}^{m\times n}$ und $c \in \mathbb{K}$. Dann sei cA die Matrix $(\!(b_{ij})\!) \in \mathbb{K}^{m\times n}$ mit $b_{ij} := c\,a_{ij}$ für $i = 1,\ldots,m$ und $j = 1,\ldots,n$.

Bemerkung. Es gelten die Distributivgesetze

$$(a+b)\,C = a\,C + b\,C\ , \quad a, b \in \mathbb{K},\ C \in \mathbb{K}^{m\times n}\ ,$$
$$a\,(B+C) = a\,B + a\,C\ , \quad a \in \mathbb{K},\ B, C \in \mathbb{K}^{m\times n}\ ,$$

sowie die Assoziativgesetze

$$(a\,b)\,C = a\,(b\,C)\ , \quad a, b \in \mathbb{K},\ C \in \mathbb{K}^{m\times n}\ ,$$
$$a\,(B\,C) = (a\,B)\,C = B\,(a\,C)\ , \quad a \in \mathbb{K},\ B \in \mathbb{K}^{m\times n},\ C \in \mathbb{K}^{n\times k}\ .$$

Ferner gilt $(c\,A)^\top = c\,A^\top$ für alle $c \in \mathbb{K}$ und alle $A \in \mathbb{K}^{m\times n}$.

Zum Abschluß wollen wir noch bemerken, daß die Definition einer Matrix $A = (\!(a_{ij})\!)$ sinnvoll bleibt, wenn die a_{ij} keine Körperelemente, sondern Elemente eines kommutativen Rings mit 1 sind. Auch die Ergebnisse dieses Paragraphen bleiben in diesem Fall richtig, weil sie nicht von der Division Gebrauch machen.

Polynome

In der Schule lernt man Polynome gewöhnlich als reelle Funktionen auf $\mathbb{R}$ kennen, die von der Form

$$(*) \qquad x \longmapsto a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \,,\ n \in \mathbb{N}_0,\ a_i \in \mathbb{R}\,,$$

sind.

Nun kann man aber ebensogut Funktionen mit der Abbildungsvorschrift (*) über einem beliebigen Körper $\mathbb{K}$ betrachten, oder, wenn wir noch einen Schritt weitergehen, kann man in (*) statt Körperelemente x auch Elemente eines Ringes mit 1 einsetzen, z.B. Matrizen $A \in \mathbb{K}^{n\times n}$. Dabei ist $A^k = \underbrace{A\cdot A\cdots A}_{k}$ für $k \in \mathbb{N}$ und a_0 wird ersetzt durch $a_0 E_n$. Man erhält dann eine Abbildung von $\mathbb{K}^{n\times n}$ in sich.

Es ist deshalb empfehlenswert, den Begriff des Polynoms abstrakter zu fassen und die Größe x als "unbestimmte Größe", d.h. einfach als Symbol aufzufassen, für das bei Bedarf andere Elemente eingesetzt werden können. Dann ist das Polynom also durch die Koeffizienten $a_0,a_1,\ldots,a_n \in \mathbb{K}$ gegeben, d.h. es ist einfach ein Element $(a_0,\ldots,a_n) \in \mathbb{K}^{n+1}$.

Da wir aber Verknüpfungen für Polynome ganz verschiedener "Längen" n erklären wollen, ist es günstiger, das $(n+1)$-Tupel $(a_0,\ldots,a_n)$ durch Nullen zu einer Folge von Elementen aus $\mathbb{K}$ zu ergänzen: $(a_0,a_1,\ldots,a_n,0,0,\ldots)$.

Definition. Ein *Polynom* p über dem Körper $\mathbb{K}$ ist eine Folge $(a_0,a_1,\ldots,a_n,a_{n+1},\ldots)$ von Körperelementen, also ein Element von $\mathbb{K}^{\mathbb{N}_0}$, derart daß ein $n \in \mathbb{N}$ existiert mit $a_k = 0$ für alle $k > n$: $p = (a_0,\ldots,a_n,0,0,\ldots)$.

Wir schreiben auch $p = (a_i)_{i \in \mathbb{N}_0}$ oder in symbolischer Form

$$p = \sum_{i=0}^{\infty} a_i X^i \quad \text{bzw.} \quad p = \sum_{i=0}^{n} a_i X^i \,.$$

$p = (0,0,\ldots)$ heißt das *Nullpolynom;* Kurzschreibweise: $p = o$.

Ist $p \neq o$, so heißt das größte $n \in \mathbb{N}_0$ mit $a_n \neq 0$ der *Grad* von p und a_n heißt der *Leitkoeffizient* von p. Wir verwenden die Schreibweise Grad $p = n$ und für $p = o$ setzen wir Grad $p = -1$. Ein Polynom vom Grad n, $n \geq 0$, heißt *normiert*, wenn $a_n = 1$ ist.

Die Menge aller Polynome über $\mathbb{K}$ bezeichnen wir mit $\mathbb{K}[X]$.

Auf der Menge $\mathbb{K}[X]$ wollen wir nun eine Addition und eine Multiplikation so einführen, daß wir formal wie mit Körperelementen rechnen können, z.B.:

$$(a_0 + a_1X)(b_0 + b_1X + b_2X^2) = a_0b_0 + (a_0b_1 + a_1b_0)X + (a_0b_2 + a_1b_1)X^2 + a_1b_2X^3 .$$

Definition. Für Polynome $p,q \in \mathbb{K}[X]$, $p = (a_i)_{i \in \mathbb{N}_0}$, $q = (b_i)_{i \in \mathbb{N}_0}$, sei

$$p + q := (a_i + b_i)_{i \in \mathbb{N}_0} ,$$

$$p \cdot q := (c_i)_{i \in \mathbb{N}_0} \text{ mit } c_i = \sum_{k=0}^{i} a_k b_{i-k} .$$

Da für endlich viele Polynome die Verknüpfungen praktisch in $\mathbb{K}^n$ mit geeignetem n ausgeführt werden (die verschwindenden Koeffizienten spielen ja keine Rolle), wissen wir sofort, daß $(\mathbb{K}[X],+)$ eine abelsche Gruppe ist. Das neutrale Element in dieser Gruppe ist das Nullpolynom. Inverses Element zu $p = (a_i)_{i \in \mathbb{N}_0}$ ist das Polynom $-p := (-a_i)_{i \in \mathbb{N}_0}$. Analog wie bei den Matrizen ergeben sich die Assoziativität von $\cdot$ und die Distributivgesetze. Wegen

$$p \cdot q = \Big(\sum_{l=0}^{i} a_l b_{i-l}\Big)_{i \in \mathbb{N}_0} \overset{i-l=k}{=} \Big(\sum_{k=0}^{i} a_{i-k} b_k\Big)_{i \in \mathbb{N}_0} = \Big(\sum_{k=0}^{i} b_k a_{i-k}\Big)_{i \in \mathbb{N}_0} = q \cdot p$$

ist die Multiplikation sogar kommutativ. Das konstante Polynom $1 = (1,0,0,\ldots)$ ist das Einselement. Also erhalten wir das folgende Ergebnis:

Satz 13. $(\mathbb{K}[X], +, \cdot)$ *ist ein kommutativer Ring mit* 1.

Bemerkungen. **(a)** Für alle Polynome $p,q \in \mathbb{K}[X]$ gilt:

$$\text{Grad } (p+q) \leq \max (\text{Grad } p, \text{Grad } q).$$

Für alle Polynome $p \neq o$, $q \neq o$ gilt: Grad $(p \cdot q)$ = Grad p + Grad q.

(b) Jedes Polynom $p = a_0 + a_1 X + \cdots + a_n X^n$ erzeugt eine Funktion auf $\mathbb{K}$, die Polynomfunktion $x \longmapsto a_0 + a_1 x + \cdots + a_n x^n$, $x \in \mathbb{K}$.

Die Menge der Polynomfunktionen $f: \mathbb{K} \longrightarrow \mathbb{K}$ bildet bezüglich der punktweise erklärten Addition und Multiplikation

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

einen kommutativen Ring mit Eins. Die Abbildung, die jedem Polynom $p \in \mathbb{K}[X]$ die zugehörige Polynomfunktion $x \longmapsto p(x)$, $x \in \mathbb{K}$, zuordnet, ist ein surjektiver Homomorphismus.

Bei endlichen Körpern ist diese Zuordnung nicht injektiv. Hier ist es wichtig, zwischen einem Polynom und der zugehörigen Polynomfunktion zu unterscheiden. Wir wollen dies an einem Beispiel verdeutlichen.

Beispiel. Wir betrachten die Polynome über dem Körper $\mathbb{F}_2$. Das Polynom $p = X + X^2 = (0,1,1,0,\ldots) \in \mathbb{F}_2[X]$ ist vom Nullpolynom $o = (0,0,\ldots)$ verschieden. Die Funktion $f : x \longmapsto x + x^2$, $x \in \mathbb{F}_2$, ist aber identisch 0, denn es gilt $f(0) = 0$ und $f(1) = 1 + 1 = 0$. Also ergeben die beiden Polynome p und o die gleiche Polynomfunktion f auf $\mathbb{F}_2$.

(c) $\mathbb{K}[X]$ ist kein Körper. Das Polynom $p = X$ besitzt z.B. kein inverses Element, denn für alle $p = (a_i)_{i \in \mathbb{N}_0} \in \mathbb{K}[X]$ gilt:

$$X \cdot (a_0 + a_1 X + \cdots + a_n X^n) = a_0 X + a_1 X^2 + \cdots + a_n X^{n+1} \neq 1.$$

Im Polynomring gelten bezüglich der Division ähnliche Aussagen, wie man sie vom Ring der ganzen Zahlen her kennt.

Satz 14. *Zu den Polynomen $p,q \in \mathbb{K}[X]$, $q \neq o$, gibt es eindeutig bestimmte Polynome $r,s \in \mathbb{K}[X]$ mit $p = s \cdot q + r$ und* Grad $r <$ Grad q.

Beweis. Wir zeigen zunächst die Existenz solcher Polynome r und s. Die Aussage ist trivial für Grad $p <$ Grad q. Dann setzen wir nämlich $s = o$ und $r = p$. Somit können wir Grad $p \geq$ Grad q annehmen. Der weitere Beweis erfolgt durch vollständige Induktion nach Grad p:

Grad $p = 0$: Dann ist $p = a_0$, $a_0 \neq 0$. Nach Voraussetzung ist dann auch q vom Grad 0, also $q = b_0$, $b_0 \neq 0$. In diesem Fall gilt $p = a_0 b_0^{-1} q$.

Induktionsschluß von Grad $p \leq k-1$ auf Grad $p = k$: Seien

$$p = a_0 + a_1 X + \cdots + a_k X^k \ , \ a_k \neq 0 \ ,$$

$$q = b_0 + b_1 X + \cdots + b_m X^m \ , \ b_m \neq 0 \ ,$$

und $m \leq k$. Wir betrachten das Polynom

$$p_1 = p - \frac{a_k}{b_m} X^{k-m} \cdot q$$

vom Grad $p_1 < k$. Für Grad $p_1 <$ Grad q folgt nach dem obigen, für Grad $p_1 \geq$ Grad q folgt nach Induktionsannahme, daß Polynome $r_1, s_1 \in \mathbb{K}[X]$ existieren mit $p_1 = s_1 q + r_1$ und Grad $r_1 <$ Grad q. Damit erhalten wir

$$p = (s_1 + \frac{a_k}{b_m} X^{k-m}) \cdot q + r_1$$

mit Grad $r_1 <$ Grad q.

Eindeutigkeit: Aus $p = s_1 q + r_1$, Grad $r_1 <$ Grad q und $p = s_2 q + r_2$, Grad $r_2 <$ Grad q, folgt durch Differenzbildung $r_2 - r_1 = (s_1 - s_2) \cdot q$. Wäre $s_1 - s_2 \neq o$, so folgte mit Grad $q \leq$ Grad $(r_2 - r_1) \leq \max \{\text{Grad } r_2, \text{Grad } r_1\}$ ein Widerspruch. Also gilt $s_1 = s_2$ und damit auch $r_1 = r_2$. ∎

Ein Element $x_0 \in \mathbb{K}$ heißt *Nullstelle* des Polynoms $p \in \mathbb{K}[X]$, wenn x_0 Nullstelle der zugehörigen Polynomfunktion $x \longmapsto p(x)$, $x \in \mathbb{K}$, ist, also $p(x_0) = 0$ gilt.

Ist x_0 Nullstelle von p, so erhalten wir als Spezialfall von Satz 14 für p die Darstellung $p = s \cdot (X - x_0)$ mit $s \in \mathbb{K}[X]$. Umgekehrt folgt aus einer solchen Darstellung

unmittelbar, daß x_0 Nullstelle des Polynoms p ist.

Korollar 15. *Genau dann ist x_0 Nullstelle eines Polynoms $p \in \mathbb{K}[X]$, wenn es eine Faktorisierung*

$$p = (X - x_0) \cdot s$$

mit $s \in \mathbb{K}[X]$ gibt.

Bemerkungen. **(a)** Ein Polynom vom Grad n , $n \geq 0$, besitzt höchstens n paarweise verschiedene Nullstellen.

(b) Hat der Körper $\mathbb{K}$ unendlich viele Elemente, so gehören zu verschiedenen Polynomen $p, q \in \mathbb{K}[X]$ auch verschiedene Polynomfunktionen. Wäre nämlich $p(x) = q(x)$ für alle $x \in \mathbb{K}$, so hätte das Polynom $p - q$ unendlich viele Nullstellen.

In diesem Fall sind $\mathbb{K}[X]$ und der Ring der Polynomfunktionen über $\mathbb{K}$ isomorph.

Nicht jedes Polynom besitzt eine Nullstelle, wie das Beispiel $X^2 + 1 \in \mathbb{R}[X]$ zeigt. Fassen wir dieses Polynom jedoch als Polynom über $\mathbb{C}$ auf, so besitzt es Nullstellen (nämlich i und $-i$). Dies ist eine Folge des Fundamentalsatzes der Algebra, den wir hier ohne Beweis zitieren wollen.

Fundamentalsatz der Algebra. *Jedes Polynom $p \in \mathbb{C}[X]$ mit* Grad $p \geq 1$ *besitzt eine Nullstelle.*

Als Folgerung ergibt sich wegen Korollar 15, daß über $\mathbb{C}$ jedes Polynom p mit Grad $p \geq 1$ vollständig in Linearfaktoren zerfällt, d.h. Produkt von Polynomen vom Grad 1 ist.

Das Polynom s heißt *Teiler* des Polynoms $p \in \mathbb{K}[X]$, wenn es ein $r \in \mathbb{K}[X]$ mit $p = s \cdot r$ gibt. Zwei Polynome heißen *teilerfremd*, wenn sie keinen gemeinsamen Teiler vom Grad ≥ 1 haben. Sind p und q teilerfremd, so muß eines der beiden Polynome vom Nullpolynom verschieden sein.

Zum Schluß dieses Paragraphen wollen wir noch einige wichtige Eigenschaften teilerfremder Polynome beweisen.

Satz 16. *Die Polynome $p,q \in \mathbb{K}[X]$ sind genau dann teilerfremd, wenn es Polynome $r,s \in \mathbb{K}[X]$ gibt mit*

$$r \cdot p + s \cdot q = 1.$$

Beweis. Seien p und q teilerfremde Polynome. Wir betrachten zunächst die Menge $I := \{rp + sq \mid r,s \in \mathbb{K}[X]\}$. Offensichtlich ist I eine Untergruppe der additiven Gruppe von $\mathbb{K}[X]$. Weiter gilt für jedes Polynom $rp + sq \in I$ und jedes $t \in \mathbb{K}[X]$, daß das Produkt $t(rp + sq) = trp + tsq$ Element von I ist. (I ist ein Ideal im Ring $\mathbb{K}[X]$).

Unter den normierten Polynomen in I gibt es ein Polynom $\tilde{t}$ vom kleinsten Grad. Sei nun $t \in I$ beliebig. Dann folgt aus Satz 14, daß es Polynome s_1, r_1 gibt mit $t = s_1\tilde{t} + r_1$ und Grad $r_1 <$ Grad $\tilde{t}$. Wegen $r_1 = t - s_1\tilde{t} \in I$ muß $r_1 = o$ sein. Somit folgt $t = s_1\tilde{t}$. Speziell können wir für t die Polynome p und q nehmen. Dann gilt $p = s_1\tilde{t}$ und $q = s_2\tilde{t}$. Da p und q teilerfremd sind, muß $\tilde{t}$ konstant sein, also $\tilde{t} = 1$. Wegen $\tilde{t} \in I$ folgt die Behauptung.

Umgekehrt gebe es nun Polynome $r,s \in \mathbb{K}[X]$ mit $rp + sq = 1$. Dann folgt für jeden gemeinsamen Teiler t von p und q aus $p = tp'$ und $q = tq'$ sofort $t(rp' + sq') = 1$, also Grad $t \leq$ Grad $1 = 0$. Somit sind p,q teilerfremd. ∎

Bemerkungen. **(a)** Es seien $k \in \mathbb{N}$, $p_1,\ldots,p_k, q \in \mathbb{K}[X]$, und jedes der Polynome $p_1,\ldots,p_k$ sei zu q teilerfremd. Dann sind auch $p_1 \cdots p_k$ und q teilerfremd.

Beweis. Vollständige Induktion nach k : Der Induktionsanfang $k = 1$ ist trivial. Schluß von $k-1$ auf k : Nach Induktionsvoraussetzung ist $p := p_1 \cdots p_{k-1}$ teilerfremd zu q. Daher gibt es Polynome $s,t \in \mathbb{K}[X]$ mit $sp + tq = 1$. Weil p_k und q teilerfremd sind, gibt es analog $s',t' \in \mathbb{K}[X]$ mit $s'p_k + t'q = 1$. Wir multiplizieren und erhalten $ss'pp_k + (spt' + ts'p_k + qtt')q = 1$. Also sind $p_1 \cdots p_k$ und q teilerfremd. ∎

(b) Die Polynome r und s aus Satz 16 lassen sich mit dem *Euklidischen Algorithmus* bestimmen.

Beispiel. Gegeben seien die Polynome $p = X^5 + 2X^3 - 3X^2 + 4$ und $q = X^2 + 1$ aus

$\mathbb{R}[X]$. Wir dividieren p durch q und erhalten

$$p = (X^3 + X - 3)q + (-X + 7) = s_1 q + r_1 .$$

Nun dividieren wir q durch r_1 und erhalten

$$q = (-X - 7)r_1 + 50 = s_2 r_1 + 50.$$

Damit ergibt sich durch Rückwärtseinsetzen

$$50 = q - s_2 r_1 = q - s_2(p - s_1 q) = -s_2 p + (1 + s_2 s_1)q$$

$$= (X + 7)p + (-X^4 - 7X^3 - X^2 - 4X + 22)q .$$

Für

$$r = \frac{1}{50}(x + 7) \quad \text{und} \quad s = \frac{1}{50}(-X^4 - 7X^3 - X^2 - 4X + 22)$$

folgt somit

$$rp + sq = 1.$$

§ 5 Der Gaußsche Algorithmus

Wir kommen nun wieder auf die Behandlung linearer Gleichungssysteme

$$(*) \qquad \begin{array}{c} a_{11}\, x_1 + \cdots + a_{1n}\, x_n = b_1 \\ \vdots \qquad\qquad\quad \vdots \quad \vdots \\ a_{m1}\, x_1 + \cdots + a_{mn}\, x_n = b_m \end{array}$$

zurück. Dabei können wir jetzt ein LGS (*) über einem beliebigen Körper $\mathbb{K}$ zulassen, d.h. ein LGS mit $a_{ij} \in \mathbb{K}$, $b_i \in \mathbb{K}$ und Variablen $x_j \in \mathbb{K}$.

Wir wollen ein konkretes Verfahren zur systematischen Lösung von (*) angeben. Grundlage dieses Verfahrens sind die *elementaren Umformungen* (genauer die *elementaren Zeilenumformungen* im Gegensatz zu den später benutzten *elementaren Spaltenumformungen*). Darunter verstehen wir die folgenden drei Umformungen eines linearen Gleichungssystems:

(a) *Vertauschen zweier Gleichungen.*

(b) *Multiplikation einer der Gleichungen mit* $c \in \mathbb{K}$, $c \neq 0$.

(c) *Addition einer, mit einem beliebigen* $c \in \mathbb{K}$ *multiplizierten, Gleichung zu einer anderen Gleichung.*

Satz 17. *Ändert man ein lineares Gleichungssystem durch elementare Zeilenumformungen, so ändert sich die Lösungsmenge des linearen Gleichungssystems nicht.*

Beweis. Offensichtlich ist jede Lösung des ursprünglichen LGS (*) auch Lösung des durch elementare Umformungen abgeänderten LGS (**). Da jede Zeilenumformung durch eine Zeilenumformung wieder rückgängig gemacht werden kann, ist umgekehrt auch jede Lösung von (**) eine Lösung von (*). ∎

Definition und Bemerkung. Zwei lineare Gleichungssysteme heißen *äquivalent*, wenn sie dieselbe Lösungsmenge besitzen. Satz 17 besagt also, daß elementare Zeilenumformungen ein lineares Gleichungssystem in ein dazu äquivalentes überführen.

Zur Vorbereitung des Algorithmus wollen wir ein Beispiel rechnen.

Beispiel. Gegeben sei das lineare Gleichungssystem (*) über den reellen Zahlen $\mathbb{R}$.

$$(*) \qquad \begin{array}{rrrrrrr} -2x_1 & + 4x_2 & - 2x_3 & - x_4 & + 4x_5 & = & -3 \\ 4x_1 & - 8x_2 & + 3x_3 & - 3x_4 & + x_5 & = & 2 \\ x_1 & - 2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ x_1 & - 2x_2 & & - 3x_4 & + 4x_5 & = & a \end{array} \quad , \quad a \in \mathbb{R}$$

Durch Vertauschen der 1. und 3. Gleichung ergibt sich folgende Form

$$\begin{array}{rrrrrrr} x_1 & - 2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ 4x_1 & - 8x_2 & + 3x_3 & - 3x_4 & + x_5 & = & 2 \\ -2x_1 & + 4x_2 & - 2x_3 & - x_4 & + 4x_5 & = & -3 \\ x_1 & - 2x_2 & & - 3x_4 & + 4x_5 & = & a \end{array} \begin{array}{l} \\ \left.\begin{array}{l} \\ \\ \end{array}\right. \\ \\ \end{array} \begin{array}{l} \\ -4 \\ \\ \\ \end{array} \quad \left.\begin{array}{l} \\ 2 \\ \\ \end{array}\right| \begin{array}{l} \\ \\ -1 \end{array}$$

Indem wir die angedeuteten Elementarumformungen des Typs (c) anwenden, erhalten wir daraus zunächst

$$\begin{array}{rrrrrrr} x_1 & - 2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ & & - x_3 & + x_4 & - 3x_5 & = & 2 \\ & & & - 3x_4 & + 6x_5 & = & -3 \\ & & - x_3 & - 2x_4 & + 3x_5 & = & a \end{array} \begin{array}{l} \\ \\ \left.\begin{array}{l} \\ \\ \end{array}\right. \\ \\ \end{array} \begin{array}{l} \\ \\ -1 \\ \\ \end{array}$$

dann

$$\begin{array}{rrrrrrr} x_1 & - 2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ & & - x_3 & + x_4 & - 3x_5 & = & 2 \\ & & & - 3x_4 & + 6x_5 & = & -3 \\ & & & - 3x_4 & + 6x_5 & = & a-2 \end{array} \begin{array}{l} \\ \\ \\ \left.\begin{array}{l} \\ \end{array}\right. \\ \end{array} \begin{array}{l} \\ \\ \\ -1 \\ \end{array}$$

und schließlich

$$\begin{array}{rrrrrrr} x_1 & - 2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ & & - x_3 & + x_4 & - 3x_5 & = & 2 \\ & & & - 3x_4 & + 6x_5 & = & -3 \\ & & & & 0 & = & a+1 \end{array}$$

Multiplizieren wir jetzt die 2. Gleichung mit -1 und die 3. Gleichung mit $-\frac{1}{3}$, so

ergibt sich die folgende *Treppennormalform* des linearen Gleichungssystems (*):

$$(**) \qquad \begin{array}{rrrrrrr} x_1 & -2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ & & x_3 & - x_4 & + 3x_5 & = & -2 \\ & & & x_4 & - 2x_5 & = & 1 \\ & & & & 0 & = & a{+}1 \end{array}$$

Offensichtlich ist (**) und damit auch (*) unlösbar, wenn $a+1 \neq 0$, also $a \neq -1$ ist.

Was läßt sich nun im Fall $a = -1$ über das Lösungsverhalten des LGS (**) aussagen ?

Die Unbekannte x_5 kann eine beliebige reelle Zahl sein, $x_5 = s$. Dann ist x_4 eindeutig festgelegt, nämlich $x_4 = 1 + 2s$. Setzen wir x_5 und x_4 in die 2. Gleichung ein, so ergibt sich für x_3 der Wert $x_3 = -1 - s$. Nun setzen wir x_5, x_4, x_3 in die 1. Gleichung ein und erhalten $x_1 - 2x_2 - 2s - 2 = 0$. In dieser Gleichung kann die Unbekannte x_2 beliebig gewählt werden, $x_2 = t$. Danach kann x_1 ausgerechnet werden, $x_1 = 2s + 2t + 2$.

Will man die Lösung direkt an dem linearen Gleichungssystem ablesen, so muß man dieses noch weiter umformen. Bei unserem Beispiel erhalten wir für $a = -1$ aus der Treppennormalform (**)

$$\begin{array}{rrrrrrr} x_1 & -2x_2 & + x_3 & - x_4 & + x_5 & = & 0 \\ & & x_3 & - x_4 & + 3x_5 & = & -2 \\ & & & x_4 & - 2x_5 & = & 1 \\ & & & & 0 & = & 0 \end{array} \quad \begin{array}{l} \\ \\ \Big] +1 \\ \Big] +1 \end{array}$$

durch die angedeuteten Elementarumformungen zunächst

$$\begin{array}{rrrrrrr} x_1 & -2x_2 & + x_3 & & - x_5 & = & 1 \\ & & x_3 & & + x_5 & = & -1 \\ & & & x_4 & - 2x_5 & = & 1 \\ & & & & 0 & = & 0 \end{array} \quad \Big] -1$$

und daraus schließlich die *Gaußsche Normalform* des linearen Gleichungssystems, wobei wir die Gleichung $0 = 0$ natürlich weglassen können:

$$(***) \qquad \begin{array}{rcl} x_1 - 2x_2 - 2x_5 &=& 2 \\ x_3 + x_5 &=& -1 \\ x_4 - 2x_5 &=& 1 \,. \end{array}$$

Aus dieser Normalform läßt sich nun die allgemeine Lösung von (***) und damit auch von (*) im Fall $a = -1$ direkt ablesen. Die Unbekannten x_5 und x_2 können beliebig gewählt werden:

$$\begin{array}{rcl} x_5 &=& s \\ x_2 &=& t \end{array} \quad , \; s,t \in \mathbb{R}.$$

Für die restlichen Unbekannten folgt aus (***) sofort

$$\begin{array}{rcl} x_1 &=& 2 + 2t + 2s \\ x_3 &=& -1 \quad - s \\ x_4 &=& 1 \quad + 2s \,. \end{array}$$

Bemerkung und Bezeichnung. Jedes lineare Gleichungssystem über dem Körper $\mathbb{K}$

$$\begin{array}{rcl} a_{11}\, x_1 + \cdots + a_{1n}\, x_n &=& b_1 \\ \vdots \qquad\qquad\quad \vdots & & \vdots \\ a_{m1}\, x_1 + \cdots + a_{mn}\, x_n &=& b_m \end{array}$$

läßt sich als Matrizengleichung in der Kurzform $A\, x = b$ schreiben mit

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{K}^{m\times n}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^{n\times 1}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{K}^{m\times 1}$$

A heißt die zu dem LGS *gehörige Matrix* und ist ebenso wie die $(m,1)$–Matrix b gegeben, und x ist eine unbekannte $(n,1)$–Matrix.

Bei den Elementarumformungen ändern sich nur die Koeffizienten a_{ij} und die b_i. Es genügt daher, anstatt des LGS nur die zugehörige Matrix A bzw. die *erweiterte Matrix*

$$(A|\,b) := \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix} \in \mathbb{K}^{m\times(n+1)}$$

zu betrachten und die Elementarumformungen als Zeilenumformungen von $(A|b)$ aufzufassen. Die ersten Schritte im vorigen Beispiel sind dann

$$(A|\,b) = \left[\begin{array}{rrrrr|r} -2 & 4 & -2 & -1 & 4 & -3 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ 1 & -2 & 1 & -1 & 1 & 0 \\ 1 & -2 & 0 & -3 & 4 & a \end{array}\right] \begin{matrix} \\ \\ \downarrow \\ \downarrow \end{matrix} \quad \text{und} \quad \left[\begin{array}{rrrrr|r} 1 & -2 & 1 & -1 & 1 & 0 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ -2 & 4 & -2 & -1 & 4 & -3 \\ 1 & -2 & 0 & -3 & 4 & a \end{array}\right] \begin{matrix} \\ \downarrow -4 \\ \downarrow 2 \\ \downarrow -1 \end{matrix}$$

und für $a = -1$ erhalten wir schließlich

$$\left[\begin{array}{rrrrr|r} 1 & -2 & 0 & 0 & -2 & 2 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Gaußscher Algorithmus (Grundform unter ausschließlicher Verwendung von Zeilenumformungen)

Wir beschreiben den Algorithmus ganz allgemein für Matrizen $A \in \mathbb{K}^{m\times n}$. Die durch Zeilenumformungen veränderte Matrix werden wir, wie bei Algorithmen üblich, wieder mit A bezeichnen. Danach wenden wir den Algorithmus auf lineare Gleichungssysteme an, sowie auf n–reihige quadratische Matrizen um gegebenenfalls die Inverse zu bestimmen.

1. Schritt : Ist in der 1. Spalte von A mindestens ein Element von Null verschieden, so kann man durch eventuelles Vertauschen der Zeilen erreichen, daß

$a_{11} \neq 0$ ist. Dann gehe man zum 2. Schritt.

Besteht die 1. Spalte nur aus Nullen, so gehe man zum 3. Schritt.

2. Schritt : Wir multiplizieren die erste Zeile mit $(a_{11})^{-1}$ und addieren dann das $(-a_{i1})$–fache der neuen 1. Zeile zur i–ten Zeile, $i = 2,...,m$. Danach hat die 1. Spalte die Form

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

und wir gehen zum 3. Schritt.

3. Schritt : Wir ersetzen die Matrix A durch die Restmatrix

$$\begin{bmatrix} a_{12} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{K}^{m\times(n-1)}, \text{ falls die 1. Spalte die Form } \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \text{ hat,}$$

bzw. durch

$$\begin{bmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{K}^{(m-1)\times(n-1)}, \text{ falls die 1. Spalte die Form } \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ hat.}$$

Dann gehen wir zurück zum 1. Schritt, der nun auf die Restmatrix angewendet wird, u.s.w.

Das Verfahren endet, wenn im 3. Schritt keine Restmatrix mehr gebildet werden kann. Da in diesem Schritt jeweils entweder eine Spalte oder eine Spalte und eine Zeile gestrichen wird, muß das Verfahren spätestens nach n–maligem Anwenden der 3. Schritte enden. Setzen wir dann alle Spalten und Zeilen, die im 3. Schritt

gestrichen, also beim Weiterlaufen des Algorithmus unverändert geblieben sind, zusammen, so ergibt sich für A die folgende *Treppennormalform*

$$A = \begin{bmatrix} 0 \cdots 0 & \boxed{1} * \cdots & & & & & \\ 0 \cdots\cdots 0 & \boxed{1} * \cdots & & & * & & \\ 0 \cdots\cdots\cdots 0 & \boxed{1} * \cdots & & & & & \\ \vdots & & \ddots & & & & \\ 0 \cdots\cdots\cdots\cdots\cdots 0 & \boxed{1} * \cdots\cdots & * & & & & \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots 0 & \boxed{1} * \cdots & * & & & & \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots & & & & & & 0 \\ \vdots & & & & & & \vdots \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots & & & & & & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \} k \\ \\ \\ \\ \\ \\ \} m-k \\ \\ \end{matrix}$$

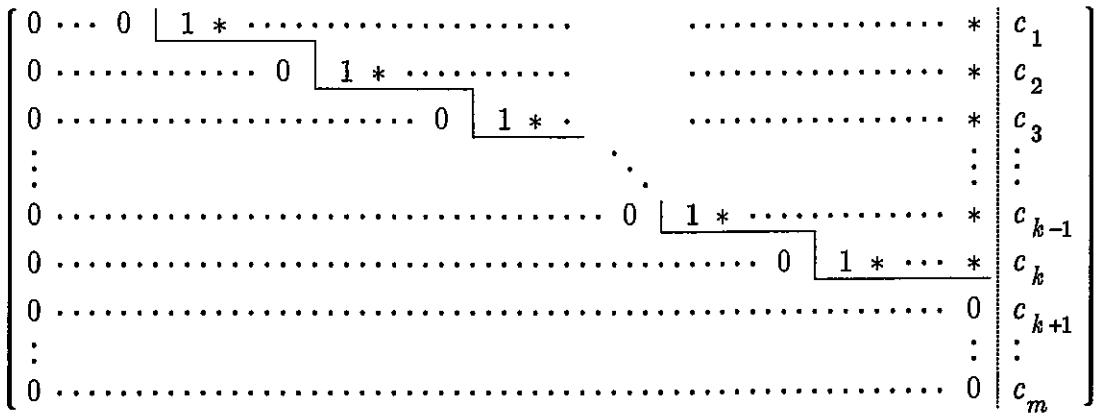
Dabei stehen unterhalb der "Treppe" nur Nullen, oberhalb an den "Stufen" Einsen und sonst irgendwelche Elemente von $\mathbb{K}$, die durch $*$ angedeutet sind.

Die Treppennormalform kann sukzessive durch weitere Zeilenumformungen (vgl. Beispiel) auf die folgende *Gaußsche Normalform* gebracht werden:

$$A = \begin{bmatrix} 0 \cdots 0 & \boxed{1} * \cdots * & 0 * \cdots * & 0 * \cdots & 0 * \cdots * \\ 0 \cdots\cdots 0 & \boxed{1} * \cdots * & 0 * \cdots & 0 * \cdots & 0 * \cdots * \\ 0 \cdots\cdots\cdots 0 & \boxed{1} * \cdots & & \vdots\ \vdots & \vdots\ \vdots\ \vdots\ \vdots \\ \vdots & & \ddots & 0 * \cdots & 0 * \cdots * \\ 0 \cdots\cdots\cdots\cdots\cdots 0 & \boxed{1} * \cdots * & 0 * \cdots & 0 * \cdots & 0 * \cdots * \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots 0 & \boxed{1} * \cdots & & & * \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots & & & & 0 \\ \vdots & & & & \vdots \\ 0 \cdots\cdots\cdots\cdots\cdots\cdots\cdots & & & & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \} k \\ \\ \\ \\ \\ \\ \} m-k \\ \\ \end{matrix}$$

Anwendung auf lineare Gleichungssysteme

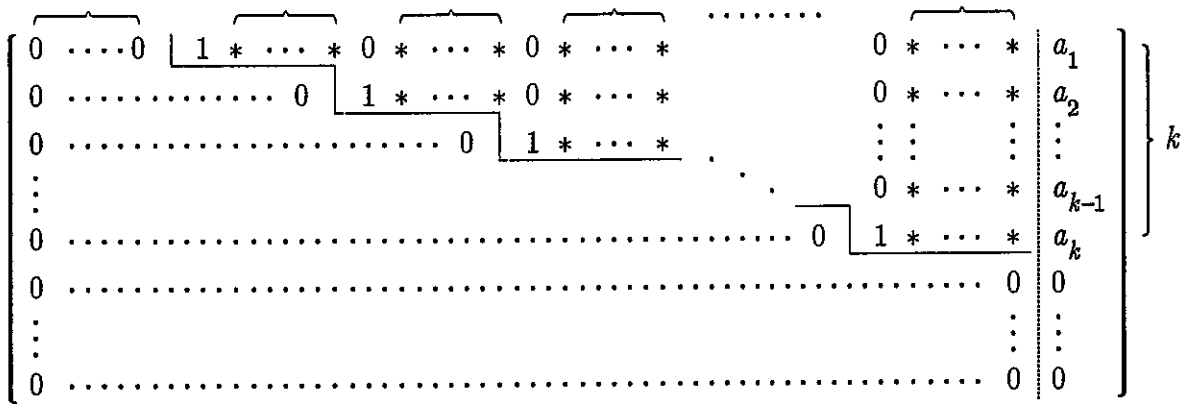
Wir wenden den Gaußschen Algorithmus auf ein lineares Gleichungssystem (*) mit der zugehörigen erweiterten Matrix $(A|b)$ an, und zwar zunächst so lange, bis die Matrix A Treppennormalform annimmt. Dann gehört zu dem umgeformten LGS (**) die erweiterte Matrix



Hieraus liest man ab:

Das LGS (**) und damit auch das LGS (*) ist unlösbar, falls einer der Koeffizienten $c_{k+1},\ldots,c_m$ von Null verschieden ist.

Ist $c_{k+1} = \cdots = c_m = 0$, so können wir das LGS (**) weiter umformen, bis die zugehörige Matrix Gaußsche Normalform hat. Die erweiterte Matrix ist dann von der Gestalt



Hieraus liest man wieder ab:

Das LGS (*) ist lösbar. Die durch geschweifte Klammern gekennzeichneten $(n-k)$ Spalten entsprechen den $(n-k)$ Unbekannten des linearen Gleichungssystems, die frei wählbar sind. Die restlichen k Unbekannten lassen sich dann direkt ausrechnen, und wir erhalten so die allgemeine Lösung des LGS (*). Insbesondere gilt:

Satz 18. *Ein homogenes LGS, das mehr Unbekannte als Gleichungen besitzt* $(m < n)$, *ist immer nichttrivial lösbar.*

Beweis. Es ist $k \leq m$. Also gibt es $n-k \geq n-m > 0$ frei wählbare Variablen. ∎

Beispiel. Wir betrachten das folgende lineare Gleichungssystem über $\mathbb{R}$

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 - 8x_4 &= 16 \\ x_1 \quad\quad - x_3 + 2x_4 &= 2 \\ 3x_4 &= -6 \\ 2x_1 - 2x_2 - 2x_3 &= 18 \end{aligned}$$

und wenden den Gaußschen Algorithmus auf die zugehörige erweiterte Matrix an:

$$\left[\begin{array}{cccc|c} 2 & 4 & -2 & -8 & 16 \\ 1 & 0 & -1 & 2 & 2 \\ 0 & 0 & 0 & 3 & -6 \\ 2 & -2 & -2 & 0 & 18 \end{array}\right] \begin{array}{l} \cdot\frac{1}{2} \\ \\ \\ \end{array} \begin{array}{l} \rfloor \\ \rfloor -1 \\ \\ \rfloor -2 \end{array} \longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & -4 & 8 \\ 0 & -2 & 0 & 6 & -6 \\ 0 & 0 & 0 & 3 & -6 \\ 0 & -6 & 0 & 8 & 2 \end{array}\right] \begin{array}{l} \\ \cdot(-\frac{1}{2}) \\ \\ \rfloor 6 \end{array}$$

Treppennormalform:

$$\longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & -4 & 8 \\ 0 & 1 & 0 & -3 & 3 \\ 0 & 0 & 0 & 3 & -6 \\ 0 & 0 & 0 & -10 & 20 \end{array}\right] \begin{array}{l} \\ \\ \cdot\frac{1}{3} \\ \rfloor 10 \end{array} \longrightarrow \left[\begin{array}{cccc|c} \underline{1} & 2 & -1 & -4 & 8 \\ 0 & \underline{1} & 0 & -3 & 3 \\ 0 & 0 & 0 & \underline{1} & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \\ \\ \rfloor 3 \\ \rfloor 4 \end{array}$$

Normalform:

$$\longrightarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -3 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \rfloor \\ \\ \\ \rfloor -2 \end{array} \longrightarrow \left[\begin{array}{cccc|c} \underline{1} & 0 & -1 & 0 & 6 \\ 0 & \underline{1} & 0 & 0 & -3 \\ 0 & 0 & 0 & \underline{1} & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Daraus liest man als allgemeine Lösung des linearen Gleichungssystems ab:

$$\begin{aligned} x_4 &= -2 \\ x_3 &= s \\ x_2 &= -3 \\ x_1 &= 6 + s \end{aligned} \quad , s \in \mathbb{R}.$$

Anwendung auf n–reihige quadratische Matrizen

Wir wollen mit Hilfe des Gaußschen Algorithmus prüfen, ob eine Matrix regulär ist und gegebenenfalls ihre Inverse berechnen. Dazu zeigen wir

Satz 19. *Es sei $A \in \mathbb{K}^{n\times n}$. Hat die Matrix $(A|E_n)$ die Gaußsche Normalform $(E_n|A')$, so ist A regulär und A' ist die zu A inverse Matrix.*

Beweis. Wir bezeichnen die k–te Spalte von A' bzw. E_n mit a_k' bzw. e_k $(k = 1,...,n)$. Geht nun $(A|E_n)$ bei den elementaren Zeilenumformungen in die Normalform $(E_n|A')$ über, so bedeutet dies für die n linearen Gleichungssysteme mit den zugehörigen erweiterten Matrizen $(A|e_1),...,(A|e_n)$, daß sie eindeutig lösbar sind und weiter, daß $a_1',...,a_n'$ die entsprechenden Lösungen sind. Also gilt $AA' = E_n$.

Machen wir die obigen Elementarumformungen wieder rückgängig, so geht die Matrix $(E_n|A')$ wieder über in $(A|E_n)$. Da es bei den elementaren Zeilenumformungen nicht auf die Reihenfolge der Spalten ankommt, bedeutet dies, daß die Matrix $(A'|E_n)$ übergeht in die Matrix $(E_n|A)$. Also gilt nach den obigen Überlegungen $A'A = E_n$. Somit ist A regulär und die Inverse ist A'. ∎

Es gilt auch die Umkehrung dieses Satzes:

Ist $A \in \mathbb{K}^{n\times n}$ regulär mit $A^{-1} = A'$, so hat die Matrix $(A|E_n)$ die Gaußsche Normalform $(E_n|A')$.

Den Beweis überlassen wir als Übungsaufgabe.

Beispiel. Gegeben sei die Matrix

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 2 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

Wir wenden den Gaußschen Algorithmus auf die Matrix $(A|E_4)$ an und erhalten

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 2 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1/2 & -1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 0 & 1/2 & 1/2 & -1/2 \\ 0 & 0 & 0 & 1 & -1 & 1/2 & 1/2 & 1/2 \end{array}\right]$$

Also ist A regulär, und es gilt

$$A^{-1} = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 1/2 & -1/2 & 1/2 \\ 0 & 1/2 & 1/2 & -1/2 \\ -1 & 1/2 & 1/2 & 1/2 \end{bmatrix}$$

Wir wenden nun Satz 19 und dessen Umkehrung auf *obere* bzw. *untere Dreiecksmatrizen* an. Darunter verstehen wir quadratische Matrizen, bei denen unterhalb bzw. oberhalb der Diagonalen nur Nullen stehen. Die Einzelheiten überlassen wir wieder als Übungsaufgabe:

Eine obere (untere) Dreiecksmatrix A ist genau dann regulär, wenn die Diagonalelemente von A alle von Null verschieden sind. Die inverse Matrix A^{-1} ist dann ebenfalls eine obere (untere) Dreiecksmatrix.

Da das Produkt zweier oberer (unterer) Dreiecksmatrizen wieder eine obere (untere) Dreiecksmatrix ist, bilden die regulären oberen (unteren) Dreiecksmatrizen eine Untergruppe der allgemeinen linearen Gruppe $GL(n,\mathbb{K})$.

Zum Abschluß dieses Paragraphen machen wir noch zwei Bemerkungen über lineare Gleichungssysteme.

Bemerkungen. **(a)** Ist A eine reguläre quadratische Matrix, so ist das lineare Gleichungssystem $Ax = b$ eindeutig lösbar, denn es gilt in diesem Fall $x = A^{-1}b$.

(b) Es seien L_{inh} die Lösungsmenge des inhomogenen LGS $Ax = b$ und $\tilde{x}$ eine beliebige Lösung. Ist L_h die Lösungsmenge des zugehörigen homogenen LGS $Ax = o$, so gilt

$$L_{inh} = \tilde{x} + L_h := \{ \tilde{x} + y \mid y \in L_h \}.$$

Damit lassen sich die Lösungsmengen der Beispiele von S.70 bzw. von S.77 nun folgendermaßen darstellen:

$$L_{inh} = \{\begin{bmatrix} 2+2t+2s \\ t \\ -1-s \\ 1+2s \\ s \end{bmatrix} \mid s,t \in \mathbb{R}\} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \{s \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mid s,t \in \mathbb{R}\}$$

bzw.

$$L_{inh} = \{\begin{bmatrix} 6+s \\ -3 \\ s \\ -2 \end{bmatrix} \mid s \in \mathbb{R}\} = \begin{bmatrix} 6 \\ -3 \\ 0 \\ -2 \end{bmatrix} + \{s \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \mid s \in \mathbb{R}\}.$$

Die weitere Behandlung linearer Gleichungssysteme wird im Rahmen der Vektorraumtheorie in den nächsten Kapiteln erfolgen.

§ 6 Anwendungen der Kongruenzrechnung

In § 3 haben wir die Restklassenringe $\mathbb{Z}_m$ eingeführt und bewiesen, daß sie genau dann Körper sind, wenn $m \in \mathbb{N}$ eine Primzahl ist. Sie spielen in vielen Anwendungen eine wichtige Rolle. So führt z.B. die Frage, auf welchen Wochentag ein bestimmtes Datum fällt, auf den Ring $\mathbb{Z}_7$, und wenn in der Datenverarbeitung nur eine begrenzte Anzahl von Symbolen zur Verfügung steht, werden Rechnungen in einem der Ringe $\mathbb{Z}_m$ erforderlich.

Wir wollen im folgenden zwei Problemstellungen behandeln und verwenden dazu wieder die klassische Kongruenzschreibweise

$$x \equiv y \bmod m$$

für $x - y \in m\,\mathbb{Z}$, d.h. für $[x]_\sim = [y]_\sim$ in $\mathbb{Z}_m$.

(1) Zu gegebenen $a \in \mathbb{Z}$, $n \in \mathbb{N}$ und $m \in \mathbb{N}$ wird ein $x \in \{0,1,\ldots,m-1\}$ gesucht mit

$$a^n \equiv x \bmod m.$$

Während man für Zahlen $a \in \mathbb{Z}$, auch wenn sie sehr groß sind, leicht ausrechnen kann, welcher Rest bei Division durch m bleibt, ist dies bei Potenzen a^n wesentlich schwieriger. Wie sieht man z.B. der Zahl 4^{10259} an, welchen Rest sie bei Division durch 18 hat?

(2) Es soll ein System von k Kongruenzen gelöst werden. Zu gegebenen Zahlen $k \in \mathbb{N}$, $a_1,\ldots,a_k \in \mathbb{N}_0$, $m_1,\ldots,m_k \in \mathbb{N}$ wird ein $x \in \mathbb{N}_0$ gesucht mit

$$\begin{aligned} x &\equiv a_1 \bmod m_1 \\ x &\equiv a_2 \bmod m_2 \\ &\vdots \\ x &\equiv a_k \bmod m_k \end{aligned}$$

Offensichtlich spielen bei der Behandlung solcher Fragen die Teilbarkeitseigenschaften ganzer Zahlen eine wesentliche Rolle. Einige dieser Eigenschaften haben wir schon in § 1.4 bei den Polynomen benutzt. Zum besseren Verständnis wol-

len wir hier nochmals einige grundlegende Tatsachen zusammenstellen (zum Teil ohne Beweis), soweit wir sie zur Behandlung der Probleme (1) und (2) benötigen.

Division mit Rest. *Für $a,b \in \mathbb{Z}$, $b \neq 0$, gibt es eindeutig bestimmte Zahlen $k \in \mathbb{Z}$ und $r \in \{0,1,\ldots,|b|-1\}$ mit*

$$(*) \qquad a = k\,b + r\,.$$

Ist $r = 0$, so ist b *Teiler* von a und wir schreiben dann $b \mid a$ (b teilt a). Die größte Zahl $d \in \mathbb{N}$, die sowohl a als auch b teilt, heißt *größter gemeinsamer Teiler* von a und b. Schreibweise: $d = \mathrm{ggT}\,(a,b)$. Ist $\mathrm{ggT}\,(a,b) = 1$, so heißen a und b *teilerfremd*.

Zur Ermittlung des ggT von a und b dient der *Euklidische Algorithmus* :

Durch mehrfache Anwendung von (*) erhält man:

$$\begin{aligned}
a &= kb + r && , && r \in \{0,\ldots,|b|-1\}\\
b &= k_1 r + r_1 && , && r_1 \in \{0,\ldots,r-1\}\\
r &= k_2 r_1 + r_2 && , && r_2 \in \{0,\ldots,r_1-1\}\\
&\vdots && && \vdots\\
r_{j-2} &= k_j\, r_{j-1} + r_j\,, && && r_j \in \{0,\ldots,r_{j-1}-1\}\\
r_{j-1} &= k_{j+1} r_j\,. && &&
\end{aligned}$$

Das Verfahren bricht ab, weil $r, r_1, r_2, \ldots$ eine monoton fallende Folge von Zahlen aus $\mathbb{N}_0$ ist.

Satz 20. Es gilt: $r_j = \mathrm{ggT}\,(a,b)$.

Beweis. Wir zeigen zunächst, daß r_j ein Teiler von a und von b ist, indem wir den Euklidischen Algorithmus von unten nach oben durchlaufen:

$$\begin{aligned}
&r_j \text{ teilt } r_{j-1} \Rightarrow r_j \text{ teilt } r_{j-2} \Longrightarrow r_j \text{ teilt } r_{j-3}\\
\Longrightarrow \cdots \Longrightarrow\ & r_j \text{ teilt } r \Longrightarrow r_j \text{ teilt } b \Longrightarrow r_j \text{ teilt } a\,.
\end{aligned}$$

Sei nun $c \in \mathbb{N}$ ein Teiler von a und b. Dann folgt analog, wenn wir das Verfahren von oben nach unten durchlaufen:

$$c \text{ teilt } r \Rightarrow c \text{ teilt } r_1 \Rightarrow \cdots \Rightarrow c \text{ teilt } r_{j-2} \Rightarrow c \text{ teilt } r_{j-1} \Rightarrow c \text{ teilt } r_j .$$

Also ist $c \leq r_j$ und somit $r_j = \text{ggT}\,(a,b)$. ∎

Beispiel. Wir wollen den ggT von 24 und 614 bestimmen:

$$\begin{aligned} 614 &= 25 \cdot 24 + 14 \\ 24 &= 1 \cdot 14 + 10 \\ 14 &= 1 \cdot 10 + 4 \\ 10 &= 2 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 \end{aligned}$$

Also ist ggT (24,614) = 2.

Mit Hilfe des Euklidischen Algorithmus erhalten wir folgendes Kriterium für die Teilerfremdheit ganzer Zahlen.

Satz 21. *Die Zahlen $a,b \in \mathbb{Z}$ sind genau dann teilerfremd, wenn es Zahlen $x,y \in \mathbb{Z}$ gibt mit $a\,x + b\,y = 1$.*

Beweis. Sei ggT $(a,b) = 1$. Lösen wir im Euklidischen Algorithmus die Gleichungen nacheinander nach $r, r_1, r_2, \ldots, r_{j-1}$ und $r_j = 1$ auf, so erhalten wir

$$r = a - k\,b \;,\; r_1 = x_1\,a + y_1\,b \;,\cdots,\; 1 = r_j = x_j\,a + y_j\,b$$

mit $x_i\,,\; y_i \in \mathbb{Z}$.

Gilt umgekehrt $a\,x + b\,y = 1$, und ist ggT $(a,b) = m \in \mathbb{N}$, so folgt $a = u \cdot m$, $b = v \cdot m$ mit $u,v \in \mathbb{N}$ und somit $(x\,u + y\,v)\,m = 1$. Also ist $m = 1$. ∎

Wir wollen einige Folgerungen aus Satz 21 ziehen.

Folgerungen. **(a)** *Für $a,b,c \in \mathbb{Z}$ mit* ggT $(a,b) = 1$, ggT $(a,c) = 1$ *folgt* ggT $(a,bc) = 1$.

Beweis. Nach Satz 21 gilt $a\,u + b\,v = 1$ und $a\,x + c\,y = 1$ mit $u,v,x,y \in \mathbb{Z}$. Daraus folgt $a\,(a\,u\,x + c\,u\,y + b\,v\,x) + b\,c\,v\,y = 1$ und somit nach Satz 21 die Behauptung. ∎

(b) *Für $a,b,c \in \mathbb{Z}$ mit $a \mid bc$ und* ggT $(a,b) = 1$ *folgt* $a \mid c$.

Beweis. Nach Voraussetzung und nach Satz 21 gilt $a\,z = b\,c$ sowie $a\,x + b\,y = 1$ mit

geeigneten $x,y,z \in \mathbb{Z}$. Daraus folgt $a\,z\,y - b\,c\,y = 0$ und $a\,x\,c + b\,y\,c = c$. Wir addieren beide Gleichungen und erhalten $a\,(z\,y + x\,c) = c$, also $a \mid c$. ∎

(c) *Für $m \in \mathbb{N}$, $a,x,y \in \mathbb{Z}$ mit* ggT $(a,m) = 1$ *und* $a\,x \equiv a\,y \bmod m$ *folgt* $x \equiv y \bmod m$.

Beweis. Aus $a\,x \equiv a\,y \bmod m$ folgt $a\,(x - y) = z\,m$ mit $z \in \mathbb{Z}$. Wegen $a \mid z\,m$ und ggT $(a,m) = 1$ erhalten wir unter Verwendung von (b), daß $z = z'\,a$ mit $z' \in \mathbb{Z}$ gilt. Somit folgt $x - y = z'\,m$ und $x \equiv y \bmod m$. ∎

(d) *Für $x,y \in \mathbb{Z}$, $m \in \mathbb{N}$ mit* ggT $(x,m) = 1$ *und* $x \equiv y \bmod m$ *folgt* ggT $(y,m) = 1$.

Beweis. Nach Satz 21 gibt es ganze Zahlen u,v mit $u\,x + v\,m = 1$, und nach Voraussetzung ist $x - y = z\,m$ mit $z \in \mathbb{Z}$. Damit erhalten wir $u\,(y + z\,m) + v\,m = 1$ und weiter $u\,y + (u\,z + v)\,m = 1$. Also folgt wieder mit Satz 21, daß ggT $(y,m) = 1$. ∎

(e) *Für $x,y \in \mathbb{Z}$, $m_1,m_2 \in \mathbb{N}$ mit* ggT $(m_1,m_2) = 1$ *und* $x \equiv y \bmod m_1$, $x \equiv y \bmod m_2$ *folgt* $x \equiv y \bmod m_1 m_2$.

Beweis. Es gilt nach Voraussetzung $x - y = z_1 m_1$ und $x - y = z_2 m_2$ mit $z_1,z_2 \in \mathbb{Z}$. Daraus folgt $z_1 m_1 = z_2 m_2$ und wegen (b) $m_1 \mid z_2$, $m_2 \mid z_1$. Somit gilt $z_2 = z_2'\,m_1$, $z_2' \in \mathbb{Z}$ und $x - y = z_2'\,m_1\,m_2$. Also ist $x \equiv y \bmod m_1 m_2$. ∎

Definition. Für $m \in \mathbb{N}$ sei

$$\varphi\,(m) := |\{k \in \{1,\ldots,m\} \mid \text{ggT}\,(k,m) = 1\}|.$$

Die Abbildung $\varphi : m \longmapsto \varphi\,(m)$ heißt *Eulersche φ–Funktion.*

Für $m \in \mathbb{N}$ gibt es also $\varphi\,(m)$ Zahlen $x_1,\ldots,x_{\varphi(m)} \in \{1,\ldots,m\}$, die zu m teilerfremd sind. Deren Restklassen in $\mathbb{Z}_m$ bilden bezüglich der Multiplikation eine Gruppe. Dies ist die Aussage des nächsten Satzes.

Satz 22. *Es sei $m \in \mathbb{N}$, $x_1,\ldots,x_{\varphi(m)}$ seien die zu m teilerfremden Zahlen aus $\{1,\ldots,m\}$ und $B = \{[x_1]_\sim,\ldots,[x_{\varphi(m)}]_\sim\} \subset \mathbb{Z}_m$. Dann ist $(B,\cdot)$ eine abelsche Gruppe.*

Beweis. B ist bezüglich der Multiplikation abgeschlossen, denn aus $[x_i]_\sim \in B$, $[x_j]_\sim \in B$ mit $x_i,x_j \in \{1,\ldots,m\}$ folgt nach (a), daß $x_i\,x_j$ teilerfremd zu m ist, und nach

(d) überträgt sich dies auf die Elemente der Restklasse $[x_i\, x_j]_\sim$, die somit ebenfalls zu B gehört. Weiterhin besitzt jede Restklasse $[x_i]_\sim$, $i = 1,\ldots,\varphi(m)$, auch eine Inverse bezüglich der Multiplikation. Denn zu x_i gibt es nach Satz 21 $z,y \in \mathbb{Z}$ mit $m\,z + y\,x_i = 1$. Also ist y teilerfremd zu m und $[y]_\sim \in \{[x_1]_\sim,\ldots,[x_{\varphi(m)}]_\sim\}$. Aus der obigen Darstellung der Eins folgt nun

$$[1]_\sim = [m\,z + y\,x_i]_\sim = [m\,z]_\sim + [y\,x_i]_\sim = [0]_\sim + [y]_\sim\,[x_i]_\sim = [y]_\sim\,[x_i]_\sim .$$

Also ist $[y]_\sim = [x]_\sim^{-1}$. Schließlich sind in $(B,\cdot)$ auch die restlichen Gruppenaxiome erfüllt, da $\mathbb{Z}_m$ ein kommutativer Ring mit Eins ist. ∎

Aus Satz 22 erhalten wir ein erstes Resultat zur Lösung von Problem (1).

Korollar 23 (Satz von Fermat–Euler). *Es seien a und m teilerfremde natürliche Zahlen. Dann gilt:*

$$a^{\varphi(m)} \equiv 1 \bmod m .$$

Beweis. Die abelsche Gruppe B aus Satz 22 besitzt genau $\varphi(m)$ Elemente $[x_1]_\sim,\ldots,[x_{\varphi(m)}]_\sim$, und $[a]_\sim$ ist in B. Weil $[x_1]_\sim\,[a]_\sim,\ldots,[x_{\varphi(m)}]_\sim\,[a]_\sim$ in B und paarweise verschieden sind, ist das Produkt dieser Elemente gleich $[x_1]_\sim\cdots[x_{\varphi(m)}]_\sim$. Daraus folgt $[a]_\sim^{\varphi(m)} = [1]_\sim$ und somit $[a^{\varphi(m)}]_\sim = [1]_\sim$. ∎

Um dieses Ergebnis zur Lösung von (1) einsetzen zu können, benötigen wir noch eine Darstellung von $\varphi(m)$, die eine Berechnung dieser Funktion erlaubt. Dazu wollen wir die Zahl m zunächst in Primfaktoren zerlegen.

Satz 24. *Jede natürliche Zahl $m > 1$ ist Produkt endlich vieler Primzahlen.*

Beweis: Ist m eine Primzahl, so ist die Behauptung trivial. Andernfalls ist $m \geq 4$ und wir erhalten durch vollständige Induktion mit dem Induktionsanfang $m = 4 = 2\cdot 2$ sofort die Behauptung. ∎

Für jede natürliche Zahl $m \geq 2$ gibt es also eine Primzahldarstellung

$$m = p_1^{n_1} \cdots p_k^{n_k} ,$$

wobei $k \in \mathbb{N}$, $p_1,\ldots,p_k$ paarweise verschiedene Primzahlen und $n_1,\ldots,n_k \in \mathbb{N}$ sind. Diese Darstellung ist bis auf die Reihenfolge der Faktoren eindeutig, wie man mit Hilfe von Folgerung (b) leicht beweist.

Weiterhin ist jeder nichttriviale Teiler von m ebenfalls Produkt von Primzahlen aus $\{p_1,\ldots,p_k\}$. Damit besitzt jede Zahl $k \in \{1,\ldots,m\}$, die nicht teilerfremd zu m ist, mindestens eine der Primzahlen $p_1,\ldots,p_k$ als Teiler, ist also Element einer der Mengen

$$A_i = \{1,\ldots,m\} \cap p_i \mathbb{N} \quad , \quad 1 \leq i \leq k \, .$$

Es gibt daher genau

$$m - \varphi(m) = |\bigcup_{i=1}^{k} A_i|$$

Zahlen aus $\{1,\ldots,m\}$, die nicht zu m teilerfremd sind. Nun nutzen wir Satz 3 der Vorbemerkungen über Mengen, Abbildungen, Relationen aus. Wir wissen

$$|A_i| = \frac{m}{p_i} \, , \qquad 1 \leq i \leq k \, ,$$

$$|A_i \cap A_j| = \frac{m}{p_i \cdot p_j} \, , \qquad 1 \leq i < j \leq k \, ,$$

$$\vdots$$

$$|A_{i_1} \cap \cdots \cap A_{i_{k-1}}| = \frac{m}{p_{i_1} \cdots p_{i_{k-1}}} \, , \qquad 1 \leq i_1 < \cdots < i_{k-1} \leq k \, ,$$

$$|A_1 \cap \cdots \cap A_k| = \frac{m}{p_1 \cdots p_k} \, .$$

Das ergibt

$$m - \varphi(m) = \sum_{i=1}^{k} \frac{m}{p_i} - \sum_{1 \leq i < j \leq k} \frac{m}{p_i \cdot p_j} + - \cdots + (-1)^{k+1} \frac{m}{p_1 \cdots p_k}$$

$$\varphi(m) = m \left(1 - \sum_{i=1}^{k} \frac{1}{p_i} + \sum_{1 \leq i < j \leq k} \frac{1}{p_i \cdot p_j} - + \cdots + (-1)^k \frac{1}{p_1 \cdots p_k}\right)$$

$$= m\,(1-\frac{1}{p_1})\cdots(1-\frac{1}{p_k}).$$

Dies ist die gewünschte Darstellung von φ und wir erhalten den folgenden Satz:

Satz 25. *Sei $m \in \mathbb{N}$ und seien $p_1,\ldots,p_k$ die Primzahlen, die m teilen. Dann gilt :*

$$\varphi\,(m) \;=\; m\,(1-\frac{1}{p_1})\cdots(1-\frac{1}{p_k})\,.$$

Beispiel. Für welches $x \in \{1,2,\ldots,14\}$ gilt $4^{10259} \equiv x \bmod 15$?

Es ist $\varphi\,(15) = 15(1-\frac{1}{5})(1-\frac{1}{3}) = 8$. Wegen $2^8 \equiv 1 \bmod 15$ und $4^{10259} = 2^{20518}$ $= (2^8)^{2564}\cdot 2^6$ folgt $4^{10259} \equiv 2^6 \bmod 15$. Also ist $x \equiv 2^6 \bmod 15$ und somit $x = 4$ die gesuchte Lösung.

Nun wenden wir uns Problem (2) zu und betrachten das System von Kongruenzen

$$(*)\qquad \begin{array}{l} x \equiv a_1 \bmod m_1 \\ \quad\vdots \qquad\quad \vdots \\ x \equiv a_k \bmod m_k \end{array} \qquad\qquad a_i \in \{0,1,\ldots,m_i-1\}.$$

Die Herleitung allgemeiner Lösungskriterien gehört zum Bereich der Zahlentheorie, weswegen wir uns hier auf den wichtigen Spezialfall beschränken, daß $m_1,\ldots,m_k$ paarweise teilerfremd sind.

Satz 26 (Chinesischer Restsatz). *Sind $m_1,\ldots,m_k$ paarweise teilerfremd, so hat (*) die Lösung*

$$x \;=\; m_1\cdots m_k \sum_{i=1}^{k} n_i \frac{a_i}{m_i}\,.$$

Dabei ist $n_i \in \{0,\ldots,m_i-1\}$ eindeutig durch

$$n_i \frac{m_1\cdots m_k}{m_i} \equiv 1 \bmod m_i$$

gegeben, $i = 1,\ldots,k$. Ist y eine weitere Lösung von (*), so gilt $y = x + k\,m_1\cdots m_k$, $k \in \mathbb{Z}$.

Beweis. Sei $i \in \{1,...,k\}$. Wiederholte Anwendung von Folgerung (a) ergibt, daß m_i und $\frac{m_1 \cdot \cdots \cdot m_k}{m_i}$ teilerfremd sind. Nach Satz 21 existieren dann Zahlen x_i , $y_i \in \mathbb{Z}$ mit

$$x_i\, m_i + \frac{m_1 \cdot \cdots \cdot m_k}{m_i}\, y_i = 1,$$

woraus

$$\frac{m_1 \cdot \cdots \cdot m_k}{m_i}\, y_i \equiv 1 \bmod m_i$$

folgt. Sei n_i die eindeutige Zahl aus $\{0,...m_i\text{–}1\}$, die $n_i \equiv y_i \bmod m_i$ erfüllt. Wir erhalten

$$\frac{m_1 \cdot \cdots \cdot m_k}{m_i}\, n_i \equiv 1 \bmod m_i\,,$$

und für

$$x = m_1 \cdot \cdots\, m_k \sum_{i=1}^{k} n_i \frac{a_i}{m_i}$$

gilt dann

$$x \equiv n_i\, a_i \frac{m_1 \cdot \cdots \cdot m_k}{m_i} \bmod m_i\,,$$

also

$$x \equiv a_i \bmod m_i\,, \quad i = 1,...,k\,.$$

Damit ist x Lösung von (*). Ist $y \in \mathbb{Z}$ ebenfalls Lösung von (*), so folgt $y \equiv x \bmod m_i$, $i = 1,...,k$, und weil die m_i teilerfremd sind, erhalten wir mit Folgerung (e) von Satz 21 $y \equiv x \bmod (m_1 \cdot \cdots \cdot m_k)$. ∎

Bemerkungen. **(a)** Der Beweis war konstruktiv, erlaubt also die Bestimmung von x . Zur Berechnung der n_i kann dabei Satz 25 und Korollar 23 benutzt werden. Danach ist ja

$$\frac{m_1 \cdot \cdots \cdot m_k}{m_i} \cdot \left(\frac{m_1 \cdot \cdots \cdot m_k}{m_i}\right)^{\varphi(m_i)-1} \equiv 1 \bmod m_i\,,$$

also

$$n_i \equiv \left(\frac{m_1 \cdot \ldots \cdot m_k}{m_i}\right)^{\varphi(m_i)-1} \bmod m_i .$$

(b) Satz 26 besagt, daß jede Zahl $x \in \{0,1,\ldots,m_1 \cdot \ldots \cdot m_k - 1\}$ in eindeutiger Weise durch das k–Tupel $(a_1,\ldots,a_k)$ dargestellt wird. Solche Zahlendarstellungen sind für die Informatik interessant, weil sich die Rechenoperationen + und · auf die einzelnen "Koordinaten" übertragen.

Beispiel. Wir suchen das kleinste $x \in \mathbb{N}$ mit

$$\begin{aligned} x &\equiv 5 \bmod 7 \\ x &\equiv 7 \bmod 11 \\ x &\equiv 3 \bmod 13. \end{aligned}$$

Es ist $m_1 = 7$, $m_2 = 11$, $m_3 = 13$, also $m_1 \cdot m_2 \cdot m_3 = 1001$. Weiterhin gilt $\varphi(m_1) = 6$, $\varphi(m_2) = 10$ und $\varphi(m_3) = 12$. Damit erhalten wir für die Zahlen n_1, n_2, n_3

$$n_1 \equiv (11 \cdot 13)^5 \equiv (4 \cdot 6)^5 \equiv 3^5 \equiv 5 \bmod 7, \text{ da } 3^6 \equiv 1 \bmod 7$$

$$n_2 \equiv (7 \cdot 13)^9 \equiv (7 \cdot 2)^9 \equiv 3^9 \equiv 4 \bmod 11, \text{ da } 3^{10} \equiv 1 \bmod 11$$

$$n_3 \equiv (7 \cdot 11)^{11} \equiv 12^{11} \equiv 12 \bmod 13, \text{ da } 12^{12} \equiv 1 \bmod 13,$$

also $n_1 = 5$, $n_2 = 4$, $n_3 = 12$, und für x ergibt sich

$$x \equiv 5 \cdot 5 \cdot 11 \cdot 13 + 4 \cdot 7 \cdot 7 \cdot 13 + 12 \cdot 3 \cdot 7 \cdot 11 \equiv 8895 \equiv 887 \bmod 1001.$$

Somit ist $x = 887$ und diese Zahl wird durch das Tripel (5,7,3) dargestellt.

Kapitel 2 Vektorräume

§ 1 Vektorräume und Untervektorräume

Bisher haben wir algebraische Strukturen betrachtet, bei denen innere Verknüpfungen, also Abbildungen $A \times A \longrightarrow A$ auf der Grundmenge A eine Rolle spielten. Nun wenden wir uns Mengen zu, auf denen neben einer inneren Verknüpfung (Addition) auch eine äußere Verknüpfung (Skalarmultiplikation) erklärt ist. Solche Mengen haben wir in Kapitel 1 mit dem Ring $\mathbb{K}^{n\times n}$ der Matrizen, dem Polynomring $\mathbb{K}[x]$ und den Lösungsmengen L_h homogener linearer Gleichungssysteme schon kennengelernt. Es sind erste Beispiele für den zentralen Begriff der linearen Algebra, den Vektorraum.

Definition. Es sei $\mathbb{K}$ ein Körper. Ein $\mathbb{K}$*–Vektorraum* oder auch *Vektorraum über dem Körper* $\mathbb{K}$ ist eine Menge V zusammen mit zwei Abbildungen $+ : V \times V \longrightarrow V$ und $\cdot : \mathbb{K} \times V \longrightarrow V$, wobei folgende Gesetze gelten:

(a) $(V,+)$ ist eine abelsche Gruppe,

(b) $a\cdot(x+y) = a\cdot x + a\cdot y$ für alle $a \in \mathbb{K}$, $x,y \in V$,

(c) $(a+b)\cdot x = a\cdot x + b\cdot x$ für alle $a,b \in \mathbb{K}$, $x \in V$,

(d) $a\cdot(b\cdot x) = (ab)\cdot x$ für alle $a,b \in \mathbb{K}$, $x \in V$,

(e) $1\cdot x = x$ für alle $x \in V$.

Die Elemente von V heißen *Vektoren*, das Neutralelement von $(V,+)$ ist der *Nullvektor* und die innere Verknüpfung $+$ heißt *Vektoraddition*.

Die Elemente von $\mathbb{K}$ werden *Skalare* genannt und die "äußere" Verknüpfung $\cdot$ heißt *Multiplikation mit Skalaren* .

Bemerkungen und Bezeichnungen. **(a)** Im Fall $\mathbb{K} = \mathbb{R}$ sprechen wir von einem *reellen Vektorraum*, im Fall $\mathbb{K} = \mathbb{C}$ von einem *komplexen Vektorraum* . Spielt der Körper $\mathbb{K}$ keine Rolle, so sagt man statt $\mathbb{K}$–Vektorraum auch kurz Vektorraum (VR).

(b) Sowohl Vektoren als auch Skalare werden mit kleinen lateinischen Buchstaben bezeichnet, insbesondere der Nullvektor mit o. Aus dem Kontext ist zu erkennen, was Skalare und was Vektoren sind. Statt $a \cdot x$ schreiben wir einfacher $a\,x$.

(c) Das Axiom (e) hat den Zweck, die sogenannte triviale Multiplikation mit Skalaren auszuschließen, bei der $a\,x = o$ für jedes $x \in V$ und jeden Skalar $a \in \mathbb{K}$ gesetzt wird. Eine solche Definition würde zu keiner sinnvollen Theorie führen.

(d) Es seien $x_1,\ldots,x_k$ endlich viele Vektoren aus V. Dann heißt jeder Vektor der Form

$$a_1\,x_1 + \cdots + a_k\,x_k\,, \quad \text{Kurzschreibweise:} \quad \sum_{i=1}^{k} a_i\,x_i\,,$$

mit $a_1,\ldots,a_k \in \mathbb{K}$ eine *Linearkombination* der Vektoren $x_1,\ldots,x_k$.

Beispiele. **(a)** $V = \mathbb{K}^n$, $n \in \mathbb{N}$, ist mit den folgenden Verknüpfungen ein $\mathbb{K}$–Vektorraum:

Addition: $(a_1,\ldots,a_n) + (b_1,\ldots,b_n) := (a_1 + b_1,\ldots,a_n + b_n)$ für alle $(a_1,\ldots,a_n) \in \mathbb{K}^n$ und alle $(b_1,\ldots,b_n) \in \mathbb{K}^n$,

Multiplikation mit Skalaren: $c\,(a_1,\ldots,a_n) := (ca_1,\ldots,ca_n)$ für alle $c \in \mathbb{K}$, $(a_1,\ldots,a_n) \in \mathbb{K}^n$.

(b) $V = \mathbb{K}^{m\times n}$, $m,n \in \mathbb{N}$, ist mit den in § 1.4 erklärten Verknüpfungen ein $\mathbb{K}$–Vektorraum.

Man beachte, daß $\mathbb{K}^{m\times n}$ und $\mathbb{K}^{mn}$ bis auf die Schreibweise übereinstimmen.

Bemerkung zur Schreibweise. Die drei $\mathbb{K}$–Vektorräume $\mathbb{K}^n$, $\mathbb{K}^{1\times n}$ und $\mathbb{K}^{n\times 1}$ unterscheiden sich nur durch die Schreibweise ihrer Elemente. Es ist für das weitere praktisch, $\mathbb{K}^n$ und $\mathbb{K}^{n\times 1}$ zu identifizieren, d.h. zu erlauben, daß n–Tupel wahlweise auch als Spalten geschrieben werden können (*Spaltenvektoren*):

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = (a_1,\ldots,a_n) \in \mathbb{K}^n.$$

Bei der Spaltenschreibweise lassen sich z.B. die Vektorraumverknüpfungen einfacher überblicken. Dagegen wollen wir $\mathbb{K}^{n\times 1}$ und $\mathbb{K}^{1\times n}$ auseinanderhalten, damit bei der Matrizenmultiplikation keine Mißverständnisse auftreten.

(c) Der Polynomring $\mathbb{K}[X]$ wird ein $\mathbb{K}$–Vektorraum, wenn die Multiplikation mit Skalaren komponentenweise erklärt wird:

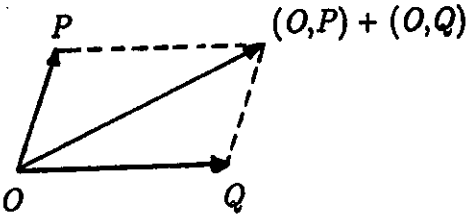
$$c\,(a_0,a_1,\ldots) := (ca_0,ca_1,\ldots) \text{ für alle } c \in \mathbb{K}, \text{ für alle } (a_0,a_1,\ldots) \in \mathbb{K}[X].$$

(d) Es seien A eine nichtleere Menge und $V = \mathbb{K}^A$. Erklären wir die Addition durch $(f,g) \longmapsto f+g$ mit $(f+g)(t) := f(t) + g(t)$ für alle $t \in A$ und die Multiplikation mit Skalaren durch $(a,f) \longmapsto af$ mit $(af)(t) := af(t)$ für alle $t \in A$, so ist V ein $\mathbb{K}$–Vektorraum.

(e) (leichte Verallgemeinerung von (d)) Es seien A eine nichtleere Menge und V ein $\mathbb{K}$–Vektorraum. Dann wird auch die Menge V^A mit analog erklärten Verknüpfungen ein $\mathbb{K}$–Vektorraum. Da $\mathbb{K}$ selbst ein $\mathbb{K}$–Vektorraum ist, erhält man für $V = \mathbb{K}$ wieder $\mathbb{K}^A$.

In den folgenden Beispielen legen wir den uns umgebenden Raum zugrunde und verwenden die Begriffe Punkt, Strecke, Länge, Richtung in ihrer anschaulichen Bedeutung. Eine gerichtete Strecke des Anschauungsraumes, zeichnerisch durch einen Pfeil symbolisiert und deshalb auch *Pfeil* genannt, ist durch ihren Anfangspunkt und ihre Spitze gekennzeichnet. Sie ist also nichts anderes als ein geordnetes Punktepaar.

Sei nun V die Menge aller Pfeile im Anschauungsraum mit einem beliebigen, aber festen Anfangspunkt O. Wir erklären die Addition zweier Pfeile mit Hilfe der "Parallelogrammregel"



und die Multiplikation mit Skalaren wie folgt.

Für $c > 0$ ist $c\,(O,P)$ der Pfeil mit Anfangspunkt O, c–facher Länge und gleicher Richtung wie (O,P), für $c < 0$ ist $c\,(O,P)$ der Pfeil mit Anfangspunkt O, $|c|$–facher Länge und entgegengesetzter Richtung wie (O,P) und für $c = 0$ ist $c\,(O,P)$ der Pfeil (O,O).

Damit wird V zu einem reellen Vektorraum. Seine Elemente heißen *Ortsvektoren.*

Die Menge aller physikalischen Kräfte, die an einem festen Punkt O angreifen, läßt sich durch diesen Vektorraum beschreiben.

Nun legen wir wieder den Anschauungsraum zugrunde, betrachten aber diesmal alle möglichen Pfeile. Dabei wollen wir zwei Pfeile (P,Q) und (R,S) als *gleich* (*äquivalent*) ansehen, wenn sie durch eine Parallelverschiebung (Translation) ineinander übergeführt werden können. Dies ist genau dann der Fall, wenn die Pfeile parallel, gleichlang und gleichgerichtet sind.

Die Punkte P des Anschauungsraums können als diejenigen speziellen Pfeile (P,P) aufgefaßt werden, bei denen Anfangspunkt und Spitze zusammenfallen. Jede Äquivalenzklasse $\overrightarrow{PQ}$ von Pfeilen heißt dann *Vektor* (*freier Vektor* im Gegensatz zu den Ortsvektoren).

Addition und Multiplikation mit Skalaren wird mit Hilfe geeigneter Pfeile wie bei den Ortsvektoren erklärt. Diese Definitionen sind unabhängig von der Wahl des Anfangspunktes O. Andere Anfangspunkte führen zu Figuren, die durch Parallelverschiebungen aus den alten hervorgehen.

Der so erklärte reelle Vektorraum heißt *Vektorraum der Pfeilklassen* oder der *freien Vektoren.* Er hat den Vorteil gegenüber dem Vektorraum der Ortsvektoren, daß man mit den freien Vektoren leichter rechnen kann. So gilt z.B. für beliebige Punkte P,Q,R des Anschauungsraumes die "Dreiecksgleichung" $\overrightarrow{PQ} + \overrightarrow{QR} + \overrightarrow{RP} = \overrightarrow{PP} = o$.

Bemerkung für Physiker. In der Physik treten viele Größen auf, wie z.B. Geschwindigkeit, Kraft, Impuls, elektrische und magnetische Feldstärke u.s.w., bei denen es

nicht genügt, ihre Größe bzw. Intensität durch eine einzige Zahl anzugeben, sondern bei denen auch die Angabe einer Richtung notwendig ist. Alle diese Größen werden von den Physikern meist als Vektoren bezeichnet. In unserem Sinn ist dies nicht immer ganz zutreffend. So sind z.B. zwei gleichgroße, gleichgerichtete Kräfte physikalisch gesehen durchaus etwas Verschiedenes, wenn sie an verschiedenen Punkten angreifen. Damit sind es aber weder Ortsvektoren noch freie Vektoren, sondern Pfeile.

Zur Einübung der Vektorraumaxiome wollen wir nun einige einfache Folgerungen beweisen.

Satz 1. *Es sei V ein $\mathbb{K}$–Vektorraum. Dann gilt*

(a) $a\,o = o$ *für alle* $a \in \mathbb{K}$,

(b) $0\,x = o$ *für alle* $x \in V$,

(c) *aus* $a\,x = o$ *folgt stets* $a = 0$ *oder* $x = o$,

(d) $(-1)\,x = -x$ *für alle* $x \in V$.

Beweis. (a) $a\,o = a\,(o + o) = a\,o + a\,o$. Daraus folgt $a\,o = o$.

(b) $0\,x = (0 + 0)\,x = 0\,x + 0\,x$. Daraus folgt $0\,x = o$.

(c) Aus $a\,x = o$, $a \neq 0$, folgt $o = a^{-1}(a\,x) = (a^{-1}a)\,x = 1 \cdot x = x$.

(d) Aus $o = 0x = (1 - 1)\,x = x + (-1)\,x$ folgt $(-1)\,x = -x$. ∎

Für den Rest dieses Paragraphen beschäftigen wir uns mit Teilmengen von Vektorräumen, die selbst wieder Vektorräume sind.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $U \subset V$ eine nichtleere Teilmenge. U heißt *Untervektorraum* (Kurzschreibweise: UVR) von V oder (*linearer*) *Unterraum* oder *Teilraum*, wenn U mit den auf $U \times U$ bzw. $\mathbb{K} \times U$ eingeschränkten Abbildungen $+$ und $\cdot$ ein $\mathbb{K}$–Vektorraum ist.

Der Nachweis, daß eine Menge $U \subset V$ Untervektorraum ist, wird meist mit dem folgenden einfachen Kriterium geführt, das sich unmittelbar aus Satz 1.4 ergibt.

Satz 2. *Es seien V ein $\mathbb{K}$–Vektorraum und $U \subset V$. Genau dann ist U ein Untervektorraum von V, wenn U nichtleer ist und wenn für alle $x,y \in U$ und alle $a \in \mathbb{K}$ gilt: $x + y \in U$ und $a\,x \in U$.*

Beispiele. **(a)** Für jeden Vektorraum V sind V und $\{o\}$ triviale Untervektorräume von V.

(b) Die Lösungsmenge eines homogenen linearen Gleichungssystems über $\mathbb{K}$ mit n Unbekannten ist ein Untervektorraum von $\mathbb{K}^n$. Die Lösungsmenge eines inhomogenen linearen Gleichungssystems mit n Unbekannten ist dagegen kein Untervektorraum von $\mathbb{K}^n$.

(c) Es seien A eine nichtleere Menge und $t \in A$. Dann ist $\{f \in \mathbb{K}^A \mid f(t) = 0\}$ ein Untervektorraum von $\mathbb{K}^A$, aber $\{f \in \mathbb{K}^A \mid f(t) = 1\}$ ist kein Untervektorraum von $\mathbb{K}^A$.

Mit Hilfe von Satz 2 folgt unmittelbar:

Korollar 3. *Der Durchschnitt beliebig vieler Untervektorräume von V ist ein Untervektorraum von V.*

Nach diesem Korollar gibt es zu jeder Teilmenge $A \subset V$ einen "kleinsten" Untervektorraum von V, der A enthält, nämlich den Durchschnitt aller Untervektorräume von V, die A enthalten. Wir geben diesem neuen Untervektorraum einen besonderen Namen.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $A \subset V$. Dann heißt

$$[A] := \bigcap_{\substack{U \text{ UVR von } V \\ A \subset U}} U$$

die *lineare Hülle* von A. Ist $A = \{x_1,\ldots,x_k\}$, so schreiben wir für $[A]$ auch $[x_1,\ldots,x_k]$ und sagen, daß die Vektoren $x_1,\ldots,x_k$ den Untervektorraum $[x_1,\ldots,x_k]$ *erzeugen* oder auch

aufspannen. Ist U ein Untervektorraum von V mit $U = [A]$, so heißt A *Erzeugendensystem* von U.

Beispiele. Die lineare Hülle des Vektorraumes V ist V selbst, die lineare Hülle der leeren Menge ist $\{o\}$ und die lineare Hülle eines Vektors $x \in V$ ist der Untervektorraum $\{a\,x \mid a \in \mathbb{K}\}$ von V.

Die lineare Hülle einer Teilmenge eines Vektorraumes läßt sich auch noch anders darstellen.

Satz 4. *Es seien V ein $\mathbb{K}$-Vektorraum und $\emptyset \neq A \subset V$. Dann ist $[A]$ die Menge aller Linearkombinationen von Vektoren aus A.*

Beweis. Sei U die Menge aller Linearkombinationen von Vektoren aus A. Mit Hilfe von Satz 2 überzeugt man sich leicht, daß U ein Untervektorraum von V ist. Wegen $A \subset U$ folgt somit $[A] \subset U$. Umgekehrt gilt $A \subset [A]$ und da $[A]$ ein Untervektorraum ist, folgt aus Satz 2 sofort, daß alle Linearkombinationen mit Vektoren aus A in $[A]$ liegen. Also gilt auch $U \subset [A]$. ∎

Bemerkung. Enthält A den Nullvektor, so benötigt man zur Beschreibung von $[A]$ nicht alle Linearkombinationen von Vektoren aus A, sondern nur diejenigen, bei denen die Summe der Koeffizienten konstant ist. Seien also $o \in A$ und $c \in \mathbb{K}$. Dann gilt:

$$[A] = \{ \sum_{i=1}^{k} a_i\, x_i \mid k \in \mathbb{N},\ a_1,\ldots,a_k \in \mathbb{K},\ x_1,\ldots,x_k \in A,\ \sum_{i=1}^{k} a_i = c \}\ .$$

Beweis. Die Menge rechts vom Gleichheitszeichen ist nach Satz 4 in $[A]$ enthalten. Umgekehrt liegt jede Linearkombination $x = a_1\, x_1 + \cdots + a_m\, x_m$ von Vektoren aus A wegen $x = a_1\, x_1 + \cdots + a_m\, x_m + (c - a_1 - \cdots - a_m)\, o$ auch in der rechten Menge. Also gilt Gleichheit. ∎

§ 2 Lineare Abhängigkeit und Unabhängigkeit

Der Nullvektor läßt sich stets als Linearkombination von k Vektoren $x_1,\dots,x_k$ darstellen, denn es gilt immer $o = 0\ x_1 + \cdots + 0\ x_k$. Gibt es neben dieser trivialen Linearkombination auch eine nichttriviale $o = a_1\ x_1 + \cdots + a_k\ x_k$ mit Skalaren $a_1,\dots,a_k$, die nicht alle Null sind, so wollen wir dieser Eigenschaft der Vektoren $x_1,\dots,x_k$ einen besonderen Namen geben.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum, $k \in \mathbb{N}$ und $x_1,\dots,x_k \in V$. Die Vektoren $x_1,\dots,x_k$ heißen *linear abhängig*, wenn es Skalare $a_1,\dots,a_k$ gibt, die nicht alle Null sind, so daß $a_1\ x_1 + \cdots + a_k\ x_k = o$ gilt. Die Vektoren $x_1,\dots,x_k$ heißen *linear unabhängig*, wenn für alle $a_1,\dots,a_k \in \mathbb{K}$ gilt: Aus $a_1\ x_1 + \cdots + a_k\ x_k = o$ folgt stets $a_1 = 0,\dots,\ a_k = 0$.

Bemerkungen. **(a)** k Vektoren $x_1,\dots,x_k \in V$ sind stets entweder linear abhängig oder linear unabhängig.

(b) Ist einer der Vektoren $x_1,\dots,x_k$ der Nullvektor, so sind sie linear abhängig. Ebenso, wenn zwei der Vektoren $x_1,\dots,x_k$ gleich sind.

(c) Die Vektoren $x_1,\dots,x_k$, $k \geq 2$, sind genau dann linear abhängig, wenn einer von ihnen Linearkombination der restlichen ist. Der Vektor x ist genau dann linear abhängig, wenn er der Nullvektor ist.

Die folgende Bemerkung ist nicht so direkt einzusehen. Wir werden sie deshalb beweisen. Es liege folgende Situation vor:

(d) In dem $\mathbb{K}$–Vektorraum V seien k Vektoren $x_1,\dots,x_k$ gegeben sowie m Linearkombinationen

$$y_1 = \sum_{i=1}^{k} a_{i1}\ x_i\ ,\dots,\ y_m = \sum_{i=1}^{k} a_{im}\ x_i\ .$$

Wir wollen feststellen, ob $y_1,\dots,y_m$ linear unabhängig oder linear abhängig sind. Hierzu machen wir den Ansatz $\sum_{j=1}^{m} t_j\ y_j = o$, $t_j \in \mathbb{K}$, und erhalten

$$(*) \qquad o = \sum_{j=1}^{m} t_j y_j = \sum_{j=1}^{m} t_j \Big(\sum_{i=1}^{k} a_{ij} x_i \Big) = \sum_{i=1}^{k} \Big(\sum_{j=1}^{m} t_j a_{ij} \Big) x_i .$$

Also gilt, falls die Vektoren $x_1,...,x_k$ linear unabhängig sind:

Die Vektoren $y_1,...,y_m$ aus V sind genau dann linear unabhängig, wenn die Spaltenvektoren

$$\hat{y}_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{k1} \end{bmatrix} \in \mathbb{K}^k, \ldots, \hat{y}_m = \begin{bmatrix} a_{1m} \\ \vdots \\ a_{km} \end{bmatrix} \in \mathbb{K}^k$$

linear unabhängig sind.

Beweis. Weil $x_1,...,x_k$ linear unabhängig sind, erhalten wir aus den obigen Gleichungen $\sum_{j=1}^{m} t_j a_{ij} = 0$ für $i = 1,...,k$, und umgekehrt hat dies $\sum_{i=1}^{k} \big(\sum_{j=1}^{m} t_j a_{ij} \big) x_i = o$ zur Folge. Also ist $\sum_{j=1}^{m} t_j y_j = o \in V$ äquivalent mit $\sum_{j=1}^{m} t_j \hat{y}_j = o \in \mathbb{K}^k$. Daraus folgt unmittelbar die Behauptung. ∎

Satz 5. *In jedem Vektorraum sind m Linearkombinationen von k Vektoren $x_1,...,x_k$ stets linear abhängig, falls $m \geq k + 1$.*

Beweis. Es seien

$$y_j = \sum_{i=1}^{k} a_{ij} x_i \ , \quad j = 1,...,m \, ,$$

beliebige m Linearkombinationen der Vektoren $x_1,...,x_k$. Wegen $k < m$ hat das homogene lineare Gleichungssystem

$$\begin{aligned} t_1 a_{11} + \cdots + t_m a_{1m} &= 0 \\ &\vdots \\ t_1 a_{k1} + \cdots + t_m a_{km} &= 0 \end{aligned}$$

nach Satz 1.18 eine nichttriviale Lösung $(t_1,...,t_m) \in \mathbb{K}^m$. Aus der obigen Gleichung

(*) folgt dann unmittelbar die Behauptung. ∎

Wie kann man nun die lineare Abhängigkeit bzw. Unabhängigkeit im konkreten Fall nachprüfen? Wir ziehen dazu unsere Kenntnisse über das Lösen linearer Gleichungssysteme heran.

Beispiele. **(a)** In dem reellen Vektorraum $V = \mathbb{R}^4$ seien die Vektoren

$$x_1 = \begin{bmatrix} 2 \\ -3 \\ 1 \\ 4 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

gegeben. Sind x_1, x_2, x_3 linear unabhängig? Der Ansatz

$$a_1 \begin{bmatrix} 2 \\ -3 \\ 1 \\ 4 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + a_3 \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

führt auf das homogene lineare Gleichungssystem

$$\begin{aligned} 2a_1 + a_2 - 2a_3 &= 0 \\ -3a_1 + \quad + a_3 &= 0 \\ a_1 + a_2 - a_3 &= 0 \\ 4a_1 + 2a_2 + a_3 &= 0 \end{aligned}$$

Wir wenden den Gaußschen Algorithmus an:

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{matrix} \\ \downarrow -2 \\ \\ \downarrow 3 \\ \downarrow -4 \end{matrix} \longrightarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 5 \end{bmatrix} \begin{matrix} \\ \\ \downarrow 3 \\ \downarrow -2 \end{matrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 5 \end{bmatrix} \begin{matrix} \\ \\ \mid (-1) \\ \downarrow \frac{5}{2} \end{matrix} \longrightarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Aus der Treppennormalform erhalten wir der Reihe nach $a_3 = 0$, $a_2 = 0$, $a_1 = 0$. Somit sind die Vektoren x_1, x_2, x_3 linear unabhängig.

(b) Es seien x_1, x_2, x_3, x_4 linear unabhängige Vektoren eines reellen Vektorraumes V und

$$\begin{aligned} y_1 &= x_1 - 2x_2 + x_3 - x_4 \\ y_2 &= -4x_1 - 2x_2 + 4x_4 \\ y_3 &= 2x_1 + 3x_2 - x_3 - 3x_4 \\ y_4 &= 17x_1 - 10x_2 + 11x_3 + x_4 \end{aligned}$$

Sind die Vektoren y_1, y_2, y_3, $y_4 \in V$ linear unabhängig? Dazu prüfen wir nach, ob die Spaltenvektoren

$$\begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ -2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 17 \\ -10 \\ 11 \\ 1 \end{bmatrix} \in \mathbb{R}^4$$

linear unabhängig sind. Dies ist nicht der Fall, da das entsprechende lineare Gleichungssystem mit $a_1 = 7$, $a_2 = 15$, $a_3 = 18$ und $a_4 = 1$ nichttrivial lösbar ist. Also gilt $7\,y_1 + 15\,y_2 + 18\,y_3 + y_4 = o$ und y_1, y_2, y_3, y_4 sind linear abhängig.

Für die folgenden Überlegungen ist es nötig, die Begriffe der linearen Abhängigkeit bzw. Unabhängigkeit auf unendlich viele Vektoren zu übertragen. Die Verallgemeinerung einer endlichen indizierten Anzahl von Vektoren wäre eine unendliche (indizierte) Familie von Vektoren. Wir wollen die obigen Definitionen aber nur auf beliebige Mengen von Vektoren übertragen.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $A \subset V$. Die Menge A heißt *linear abhängig*, wenn es (paarweise) verschiedene Vektoren $x_1, \ldots, x_k \in A$, $k \in \mathbb{N}$, gibt, die linear abhängig sind. Ist A nicht linear abhängig, so heißt A *linear unabhängig*.

Bemerkungen. **(a)** A ist genau dann linear unabhängig, wenn $A = \emptyset$ oder wenn $A \neq \emptyset$ und für jedes $k \in \mathbb{N}$ alle paarweise verschiedenen Vektoren $x_1, \ldots, x_k \in A$ linear unabhängig sind.

(b) Ist A endlich mit m Elementen, $A = \{x_1,...,x_m\}$, $m \in \mathbb{N}$, so gilt: A ist linear abhängig (linear unabhängig) genau dann, wenn $x_1,...,x_m$ linear abhängig (linear unabhängig) sind.

(c) Enthält A den Nullvektor, so ist A linear abhängig.

(d) Jede Obermenge einer linear abhängigen Menge ist linear abhängig. Jede Teilmenge einer linear unabhängigen Menge ist linear unabhängig.

Die folgenden zwei Bemerkungen sind wieder nicht so direkt einzusehen, weshalb wir sie beweisen.

(e) A ist genau dann linear abhängig, wenn es einen Vektor $x \in A$ gibt mit

$$[A] = [A \setminus \{x\}].$$

Beweis. Ist A linear abhängig, so gibt es k paarweise verschiedene Vektoren $x_1,...,x_k \in A$, $k \in \mathbb{N}$, die linear abhängig sind. Für $k = 1$ ist $x_1 = o$ und $[A \setminus \{x_1\}] = [A]$. Für $k \geq 2$ ist einer der Vektoren $x_1,...,x_k$ eine Linearkombination der anderen. Sei dieser Vektor etwa x_1. Ersetzen wir in jeder Linearkombination von Vektoren aus A, in der x_1 vorkommt, diesen Vektor durch die Linearkombination der Vektoren $x_2,...,x_k$, so erhalten wir $[A] \subset [A \setminus \{x_1\}]$ und somit $[A] = [A \setminus \{x_1\}]$.

Umgekehrt existiere ein Vektor $x \in A$ mit $[A] = [A \setminus \{x\}]$. Ist $A \setminus \{x\} = \emptyset$, so ist $[A \setminus \{x\}] = \{o\}$, also $x = o$. Nach (c) ist A dann linear abhängig. Ist $A \setminus \{x\} \neq \emptyset$, so ist x nach Satz 4 eine Linearkombination von Vektoren $x_1,...,x_k \in A \setminus \{x\}$, von denen wir o.E. annehmen können, daß sie paarweise verschieden sind. Dann sind x, $x_1,...,x_k$ ebenfalls paarweise verschieden und außerdem linear abhängig. Also ist auch A linear abhängig. ∎

(f) Es sei $x \in V$. Ist A linear unabhängig und $x \notin [A]$, so ist $A \cup \{x\}$ ebenfalls linear unabhängig. Umgekehrt folgt aus der linearen Unabhängigkeit von $A \cup \{x\}$ und $x \notin A$, daß A linear unabhängig ist und $x \notin [A]$.

Beweis. Wir beweisen nur die erste Behauptung und überlassen die zweite als Übung. Aus $x \notin [A]$ folgt zunächst $x \neq o$. Seien nun $x_1,...,x_k$ paarweise verschiedene

Vektoren aus $A \cup \{x\}$ und $a_1 x_1 + \cdots + a_k x_k = o$. Sind alle x_i von x verschieden, so gilt $x_i \in A$, $i = 1,...,k$, und aus der linearen Unabhängigkeit von A folgt $a_1 = \cdots = a_k = 0$. Ist etwa $x_k = x$, so muß $a_k = 0$ sein, da sonst $x \in [A]$ wäre. Dann folgt wie zuvor, daß auch $a_1 = \cdots = a_{k-1} = 0$ gilt. ∎

Beispiele. **(a)** Seien $V = \mathbb{K}^n$ und $A = \{e_1,...,e_n\}$ mit

$$e_i := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftarrow i\text{-te Stelle.}$$

A ist linear unabhängig, denn aus $o = a_1 e_1 + \cdots + a_n e_n$ folgt

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

also $a_1 = \cdots = a_n = 0$.

(b) In $V = \mathbb{K}[X]$ betrachten wir die Teilmenge $A = \{p_i \mid p_i = X^i , i \in \mathbb{N}_0\}$.

Es seien $k \in \mathbb{N}$, $p_{i_1},...,p_{i_k}$, $i_1 < i_2 < \cdots < i_k$, endlich viele verschiedene Vektoren aus A und $a_{i_1} p_{i_1} + \cdots + a_{i_k} p_{i_k} = o$. Dann folgt

$$(0,...,0,a_{i_1},0,...,0,a_{i_2},0,...,...,0,a_{i_k},0,...) = (0,0,...)$$

und somit $a_{i_1} = \cdots = a_{i_k} = 0$. Also ist A linear unabhängig.

§ 3 Basis und Dimension

In einem Vektorraum V sind besonders jene linear unabhängigen Mengen A wichtig, die die Eigenschaft haben, daß sich jeder Vektor aus V durch Vektoren von A linear kombinieren läßt.

Definition. Es sei V ein $\mathbb{K}$–Vektorraum.

(a) Ein Erzeugendensystem A von V heißt *minimal*, wenn keine echte Teilmenge von A Erzeugendensystem von V ist.

(b) Eine linear unabhängige Teilmenge $A \subset V$ heißt *maximal*, wenn in V jede echte Obermenge von A linear abhängig ist.

(c) Jedes linear unabhängige Erzeugendensystem von V heißt *Basis* von V.

Der Zusammenhang zwischen den obigen Definitionen wird im folgenden Satz hergestellt.

Satz 6. *Es seien V ein $\mathbb{K}$–Vektorraum und B eine nichtleere Teilmenge von V. Dann sind folgende Aussagen äquivalent:*

(a) *B ist Basis von V.*

(b) *B ist minimales Erzeugendensystem von V.*

(c) *B ist eine maximale linear unabhängige Teilmenge von V.*

(d) *Jeder Vektor $x \in V$ ist Linearkombination paarweise verschiedener Vektoren aus B, und jede derartige Linearkombination ist eindeutig, d.h. aus*

$$x = \sum_{i=1}^{k} a_i x_i = \sum_{i=1}^{k} b_i x_i$$

mit $a_1,\ldots,a_k$, $b_1,\ldots,b_k \in \mathbb{K}$, $x_1,\ldots,x_k \in B$ folgt $a_1 = b_1,\ldots, a_k = b_k$.

Bemerkung. Für $B = \emptyset$ sind (a), (b), (c) äquivalent (Übungsaufgabe).

Beweis. (a) $\Rightarrow$ (b): B ist nach Voraussetzung ein Erzeugendensystem von V. Ist B nicht minimal, so gibt es eine echte Teilmenge A von B, die Erzeugendensystem von

V ist. Für $A = \emptyset$ erhalten wir $V = \{o\}$ und damit $B = \emptyset$ im Widerspruch dazu, daß A echte Teilmenge von B ist. Also ist $A \neq \emptyset$. Dann gibt es einen Vektor aus $B \setminus A$, der Linearkombination von Vektoren aus A ist, und B ist linear abhängig im Widerspruch zur Voraussetzung.

(b) $\Rightarrow$ (c): B ist linear unabhängig. Anderfalls existiert ein Vektor $x \in B$ mit $[B] = [B \setminus \{x\}]$ und B ist nicht minimal. B muß auch maximale linear unabhängige Teilmenge von V sein, da es sonst in V eine echte Obermenge A von B gibt, die linear unabhängig ist, und somit einen Vektor $x \in A \setminus B$, der keine Linearkombination von Vektoren aus B ist, ein Widerspruch zur Voraussetzung.

(c) $\Rightarrow$ (d): Es ist $V = [B]$. Anderfalls existiert ein Vektor $x \in V$, $x \notin [B]$ und $B \cup \{x\}$ ist eine echte linear unabhängige Obermenge von B und B ist nicht maximal. Daher ist jeder Vektor $x \in V$ Linearkombination paarweise verschiedener Vektoren $x_1,\ldots,x_k$ aus B. Daß die Linearkombination eindeutig ist, folgt unmittelbar aus der linearen Unabhängigkeit der Vektoren $x_1,\ldots,x_k$.

(d) $\Rightarrow$ (a): B ist offensichtlich Erzeugendensystem von V. B ist auch linear unabhängig. Seien nämlich $x_1,\ldots,x_k$ paarweise verschiedene Vektoren aus B, $a_1,\ldots,a_k \in \mathbb{K}$ und $a_1 x_1 + \cdots + a_k x_k = o$. Da auch $0 x_1 + \cdots + 0 x_k = o$ gilt, folgt aus der Eindeutigkeit $a_1 = 0,\ldots, a_k = 0$. ∎

Bemerkung. Die Aussage (d) in Satz 6 kann verallgemeinert werden:

Es seien B eine Basis von V und $x_1,\ldots,x_k$ sowie $y_1,\ldots,y_m$ jeweils paarweise verschiedene Vektoren aus B. Ist $x \in V$ eine Linearkombination sowohl von $x_1,\ldots,x_k$ als auch von $y_1,\ldots,y_m$, so sind in beiden Linearkombinationen die vom Nullvektor verschiedenen Summanden bis auf die Reihenfolge gleich.

Beweis. Sei $x = \sum_{i=1}^{k} a_i x_i = \sum_{j=1}^{m} b_j y_j$ und o.B.d.A. $a_i \neq 0$, $b_j \neq 0$ für $i = 1,\ldots,k$ und $j = 1,\ldots,m$. Wäre $\{x_1,\ldots,x_k\} \cap \{y_1,\ldots,y_m\} = \emptyset$, so folgte, daß $x_1,\ldots,x_k, y_1,\ldots,y_m$ paarweise verschieden sind, und wegen $\sum_{i=1}^{k} a_i x_i - \sum_{j=1}^{m} b_j y_j = o$ erhielten wir $a_1 = \cdots = a_k = 0$,

$b_1 = \cdots = b_m = 0$, also einen Widerspruch zur Voraussetzung.

Seien also o.B.d.A. $x_1 = y_1,\ldots,x_l = y_l$, $1 \leq l \leq k$, $l \leq m$, und die restlichen Vektoren paarweise verschieden. Ist $l < m$, so sind $x_1,\ldots,x_k$, $y_{l+1},\ldots,y_m$ paarweise verschieden und linear unabhängig. Aus $\sum_{i=1}^{k} a_i x_i = \sum_{j=1}^{m} b_j y_j$ folgt dann $b_{l+1} = \cdots =$ $b_m = 0$, ein Widerspruch. Also ist $l = m$, und analog zeigen wir auch $l = k$.

Damit gilt $k = m = l$, also $x = \sum_{i=1}^{l} a_i x_i = \sum_{j=1}^{l} b_j x_j$. Daraus folgt schließlich $a_i = b_i$ für $i = 1,\ldots,l$. ∎

Beispiele. **(a)** Im Nullraum $V = \{o\}$ ist $\emptyset$ Basis.

(b) In $\mathbb{K}^n$ ist $B = \{e_1,\ldots,e_n\}$ eine Basis: B ist linear unabhängig; dies haben wir schon gezeigt. B ist auch Erzeugendensystem von $\mathbb{K}^n$:

$$\text{Für } x = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{K}^n \text{ gilt } x = \sum_{i=1}^{n} a_i e_i .$$

Also ist B Basis von $\mathbb{K}^n$. B heißt *Standardbasis* oder *kanonische Basis* von $\mathbb{K}^n$. Entsprechend definiert man die Standardbasis des Vektorraumes $\mathbb{K}^{m\times n}$.

(c) Seien $V = \mathbb{K}[X]$ und $B = \{p_i \mid p_i = X^i , i \in \mathbb{N}_0\}$.

B ist, wie wir schon wissen, linear unabhängig. B ist auch Erzeugendensystem von V. Dies zeigt man wie in Beispiel (b).

(d) Sei $V = \mathbb{K}^A$. Hier können wir keine Basis angeben, falls $|A| = \infty$. Ist beispielsweise $A = \mathbb{N}$, so ist $\mathbb{K}^A$ die Menge aller Folgen $(a_1,a_2,\ldots)$ mit Elementen aus $\mathbb{K}$. Die Menge

$$\{(0,0,\ldots,0,\underbrace{1}_{i\text{-te Stelle}},0,\ldots) \mid i \in \mathbb{N}\}$$

ist zwar linear unabhängig, aber sie ist kein Erzeugendensystem von $\mathbb{K}^{\mathbb{N}}$.

(e) Seien $V = \mathbb{R}^4$ und

$$A = \{\begin{bmatrix} 2 \\ -3 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}\}.$$

A ist linear unabhängig, aber kein Erzeugendensystem von V, da sich der Standardbasisvektor e_4 nicht aus den Elementen von A linear kombinieren läßt. Das entsprechende inhomogene lineare Gleichungssystem besitzt nämlich keine Lösung.

Wegen $e_4 \notin [A]$ ist $A \cup \{e_4\}$ linear unabhängig und wegen Satz 5 auch maximal, also eine Basis von V.

Wie das letzte Beispiel zeigt, kann es in einem Vektorraum mehrere Basen geben, man kann also nicht von **der** Basis sprechen. Unabhängig von der speziellen Wahl einer Basis ist aber die Anzahl ihrer Vektoren. Dies zeigt der folgende Satz.

Satz 7. *Es seien V ein $\mathbb{K}$–Vektorraum und B, B' Basen von V. Dann gilt $|B| = |B'|$.*

Beweis. Für $B = \emptyset$ ist $V = \{o\}$ und somit $B' = \emptyset$. Entsprechend folgt aus $B' = \emptyset$, daß auch $B = \emptyset$ gilt. Ist $|B| = k$, $k \in \mathbb{N}$, so muß $B \neq \emptyset$ sein, und wegen Satz 5 gilt $|B'| \leq k$. Daraus folgt wiederum wegen Satz 5 $|B| \leq |B'|$, insgesamt also $|B| = |B'|$. Entsprechend schließt man im Fall $|B'| = k$, $k \in \mathbb{N}$. Für $|B| = |B'| = \infty$ ist nichts zu beweisen. ∎

Wir wollen uns nun dem Problem zuwenden, ob es in jedem $\mathbb{K}$–Vektorraum V eine Basis gibt. Wir wissen auf jeden Fall, daß in V ein Erzeugendensystem existiert, nämlich V selbst. Gibt es dann auch ein minimales Erzeugendensystem in V? Dazu unterscheiden wir zwei Fälle:

1. **Fall** : In V gibt es ein endliches Erzeugendensystem.
2. **Fall** : Jedes Erzeugendensystem von V ist unendlich.

Im ersten Fall können wir jedes Erzeugendensystem von V zu einer Basis von V "abmagern". Wie das geht, zeigt der folgende Satz.

Satz 8. *Es seien V ein $\mathbb{K}-$Vektorraum und $A' \subset V$ ein endliches Erzeugendensystem von V. Dann gibt es zu jedem Erzeugendensystem A von V eine endliche Teilmenge $B \subset A$, die Basis von V ist.*

Beweis. Da jeder der endlich vielen Vektoren aus A' Linearkombination von endlich vielen Vektoren aus A ist, gibt es eine endliche Teilmenge $A_1 \subset A$, die Erzeugendensystem von V ist. Ist A_1 minimal, so sind wir fertig. Ist dies nicht der Fall, so gibt es eine echte Teilmenge $A_2 \subset A_1$, die Erzeugendensystem ist. Nach maximal $|A_1|$ Schritten erhalten wir so in V ein endliches Erzeugendensystem $B \subset A$, das auch minimal ist, also eine Basis von V. ∎

Korollar 9. *Es seien V ein $\mathbb{K}-$Vektorraum, B eine Basis von V und A ein Erzeugendensystem von V mit $|A| = |B| < \infty$. Dann ist A ebenfalls eine Basis von V.*

Beweis. Nach Satz 8 gibt es eine Teilmenge B' von A, die Basis von V ist. Nach Satz 7 gilt $|B'| = |B|$. Somit folgt $|B'| = |A|$ und schließlich $B' = A$. ∎

Im zweiten Fall ist die Vorgehensweise aus dem Beweis von Satz 8 unbrauchbar, da sie immer wieder zur Ausgangssituation eines unendlichen Erzeugendensystems zurückführen kann, das noch nicht minimal ist. Hier gibt es kein konstruktives Verfahren, das von dem unendlichen Erzeugendensystem A ausgehend zu einer Basis führt, aber man kann mit Hilfe des Zornschen Lemmas (siehe S. 26) auch in diesem Fall die Existenz einer Basis $B \subset A$ beweisen. Wir verzichten hier auf einen Beweis, da wir das Zornsche Lemma gleich bei dem Gegenstück von Satz 8, dem Basisergänzungssatz, verwenden werden.

Zunächst geben wir aufgrund der vorstehenden Überlegungen folgende Definition.

Definition. Es sei V ein $\mathbb{K}-$Vektorraum. Besitzt V ein endliches Erzeugendensystem, so heißt V *endlich dimensional* und die allen Basen von V gemeinsame Anzahl $n \in \mathbb{N}_0$ der Elemente heißt die *Dimension* von V. Schreibweise: $\dim V < \infty$ bzw. $\dim V = n$. Hat V kein endliches Erzeugendensystem, so heißt V *unendlich dimensional* und wir schreiben $\dim V = \infty$.

Bemerkung. ∞ ist nur ein Symbol. Es ist aber praktisch, ∞ als "unendlich große" Zahl aufzufassen und folgende Vereinbarung zu treffen:

$$\infty + \infty := \infty \ , \ \infty + a := \infty \text{ für alle } a \in \mathbb{R}.$$

Satz 10 (Basisergänzungssatz). *Es seien V ein $\mathbb{K}-$Vektorraum und $A \subset V$ eine linear unabhängige Teilmenge. Dann gibt es eine Basis von V, die A enthält.*

Beweis. Ist A maximal, so ist A eine Basis von V. Ist A nicht maximal, so gibt es eine echte linear unabhängige Obermenge A_1 von A.

Ist dim $V = n < \infty$, so gilt wegen Satz 5 $|A_1| \leq n$. Nach höchstens n Schritten erhalten wir so eine linear unabhängige Obermenge B von A mit $|B| = n$. Da jede echte Obermenge von B wegen Satz 5 linear abhängig sein muß, ist B maximal und somit eine Basis von V.

Ist dim $V = \infty$, so führt dieses Verfahren nicht zum Ziel. Hier müssen wir uns mit einem nichtkonstruktiven Existenzbeweis begnügen, der mit Hilfe des Zornschen Lemmas geführt wird: Wir betrachten dazu das Mengensystem

$$\mathcal{M} = \{A' | A \subset A' \subset V, A' \text{ linear unabhängig}\}$$

$\mathcal{M}$ ist bezüglich der Inklusion $\subset$ eine geordnete Menge. Um das Zornsche Lemma anwenden zu können, müssen wir zeigen, daß jede bezüglich $\subset$ totalgeordnete Teilmenge $\mathcal{N}$ von $\mathcal{M}$ eine obere Schranke C in $\mathcal{M}$ besitzt. Wir wählen

$$C = \bigcup_{A' \in \mathcal{N}} A'$$

Dann ist $A \subset C \subset V$, und C ist linear unabhängig: Für $C = \emptyset$ ist dies klar. Für $C \neq \emptyset$ seien $x_1,\ldots,x_k$ beliebige, paarweise verschiedene Vektoren aus C. Nach Definition von C gibt es Mengen $A_i' \in \mathcal{N}$ mit $x_i \in A_i'$ für $i = 1,\ldots,k$. Da $\mathcal{N}$ totalgeordnet ist, gibt es unter den k Mengen $A_1',\ldots,A_k'$ eine Menge $\tilde{A} = A_{i_0}'$ mit $A_i' \subset \tilde{A}$ für $i = 1,\ldots,k$. Somit gilt $x_1,\ldots,x_k \in \tilde{A}$, und weil $\tilde{A}$ linear unabhängig ist, sind es auch die Vektoren $x_1,\ldots,x_k$. Also ist C linear unabhängig und damit ein Element von $\mathcal{M}$. Da offensichtlich $A' \subset C$ für alle $A' \in \mathcal{N}$ gilt, ist C auch obere Schranke von $\mathcal{N}$.

Nach dem Zornschen Lemma existiert nun in $\mathcal{M}$ ein maximales Element B,

d.h. B ist linear unabhängig, es ist $A \subset B$ und aus $B \subset A'$ für $A' \in \mathcal{M}$ folgt stets $B = A'$. Damit ist B auch eine maximale linear unabhängige Teilmenge von V, also eine Basis von V. ∎

Bemerkung. Jeder $\mathbb{K}$–Vektorraum V besitzt eine Basis B. Es ist dim $V = |B|$.

Korollar 11. *Es seien V ein endlich dimensionaler $\mathbb{K}$–Vektorraum, B eine Basis von V und $A \subset V$ eine linear unabhängige Teilmenge mit $|A| = |B|$. Dann ist A ebenfalls eine Basis von V.*

Beweis. Nach Satz 10 gibt es eine Basis B' von V mit $A \subset B'$. Dann gilt $|A| = |B| = |B'|$ und somit $A = B'$. ∎

Der nächste Satz zeigt, daß die Dimension eines Untervektorraumes höchstens so groß sein kann wie die des Vektorraumes selbst.

Satz 12. *Es seien V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $U \subset V$ ein Untervektorraum. Dann gilt:*

(a) dim $U \leq n$.

(b) *Genau dann ist* dim $U = n$, *wenn* $U = V$.

Beweis. (a) Sei A eine beliebige linear unabhängige Teilmenge von U. Nach Satz 10 kann A zu einer Basis B von V ergänzt werden. Dann gilt $A \subset B$ und $|A| \leq n$. Somit gibt es auch eine maximale linear unabhängige Teilmenge $A \subset U$, die nach Satz 6 Basis von U ist. Daraus folgt dim $U \leq n$.

(b) Seien dim $U = n$, B eine Basis von U und $x \in V$. Gilt $x \notin [B]$, so ist $B \cup \{x\}$ linear unabhängig, ein Widerspruch zu Satz 5. Also ist $V \subset [B] = U$ und damit $V = U$. Die umgekehrte Richtung ist trivial. ∎

Bemerkung. Für unendlich dimensionale Vektorräume V ist die Ungleichung dim $U \leq$ dim V trivial, aus dim $U =$ dim V folgt aber nicht $U = V$.

Beispiel. Im $\mathbb{R}^5$ seien die Vektoren

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \quad x_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}$$

gegeben und es sei $U = [x_1,...,x_4]$.

(a) Wir wollen unter den Vektoren $x_1,...,x_4$ eine Basis von U finden. Dazu prüfen wir zunächst nach, ob die Vektoren $x_1,...,x_4$ linear unabhängig sind. Der Ansatz

$$a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 = o$$

führt auf ein lineares Gleichungssystem mit der zugehörigen Matrix

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix} \begin{matrix} \\ -2 \\ 1 \\ 1 \\ 1 \end{matrix}$$

Durch elementare Zeilenumformungen erhalten wir mit dem Gaußschen Algorithmus:

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -5 & -10 & 10 \\ 0 & 3 & 6 & -6 \\ 0 & 4 & 8 & -7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ :-5 \\ :3 \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ -1 \\ \\ -4 \end{matrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ 2 \\ \end{matrix} \begin{matrix} \\ \\ \\ \\ 1 \end{matrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ -2 \\ \\ \end{matrix} \longrightarrow \begin{bmatrix} \underline{1} & 0 & -1 & 0 \\ 0 & \underline{1} & 2 & 0 \\ 0 & 0 & 0 & \underline{1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Anhand der Normalform sehen wir, daß x_1, x_2, x_4 linear unabhängig sind, weil das lineare Gleichungssystem $a_1 x_1 + a_2 x_2 + a_4 x_4 = o$ nur trivial lösbar ist, und weiter, daß x_3 Linearkombination von x_1, x_2 ist, weil das lineare Gleichungssystem $b_1 x_1 + b_2 x_2 = x_3$ lösbar ist. Also gilt $U = [x_1, x_2, x_4]$ und $\{x_1, x_2, x_4\}$ ist Basis von U.

(b) Wir wollen eine möglichst einfache Basis von U finden, indem wir die Vek-

toren $x_1,\ldots,x_4$ durch geeignete Linearkombinationen ersetzen. Zur praktischen Durchführung schreiben wir die Vektoren $x_1,\ldots,x_4$ in die Zeilen einer Matrix und wenden wieder das Gaußsche Verfahren an.

$$\begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 2 & -1 & 1 & 2 & -2 \\ 3 & -4 & 3 & 5 & -3 \\ -1 & 8 & -5 & -6 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & -5 & 3 & 4 & 0 \\ 0 & -10 & 6 & 8 & 0 \\ 0 & 10 & -6 & -7 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & 1 & -3/5 & -4/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & -1 & 0 & -1 \\ 0 & 1 & -3/5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} \boxed{1} & 0 & 1/5 & 0 & -1 \\ 0 & \boxed{1} & -3/5 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Also ist

$$U = [\underbrace{\begin{bmatrix} 1 \\ 0 \\ 1/5 \\ 0 \\ -1 \end{bmatrix}}_{u_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ -3/5 \\ 0 \\ 0 \end{bmatrix}}_{u_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{u_3}] .$$

Die Vektoren u_1, u_2, u_3 bilden eine Basis von U, denn sie sind auch linear unabhängig: Aus $a_1\,u_1 + a_2\,u_2 + a_3\,u_3 = o$ folgt nämlich $a_1 = a_2 = a_3 = 0$.

Das in den beiden vorangehenden Beispielen angewendete Verfahren soll nun allgemein dargestellt werden. Insbesondere soll geklärt werden, wie sich die elementaren Zeilenumformungen auf die Spalten bzw. Zeilen einer Matrix auswirken. Sei

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{K}^{m\times n}.$$

Zu A kann man zwei Systeme von Vektoren aus $\mathbb{K}^m$ bzw. $\mathbb{K}^n$ betrachten. Zunächst bilden die Spalten

$$s_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \ldots, s_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

n Vektoren im $\mathbb{K}^m$. Sie spannen einen Untervektorraum $U \subset \mathbb{K}^m$ auf, $U = [s_1,\ldots,s_n]$.

Dann bilden auch die Zeilen von A, wenn man sie als Spaltenvektoren schreibt

$$z_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \end{bmatrix}, \ldots, z_m = \begin{bmatrix} a_{m1} \\ \vdots \\ a_{mn} \end{bmatrix}$$

m Vektoren im $\mathbb{K}^n$. Sie spannen einen Untervektorraum $W \subset \mathbb{K}^n$ auf, $W = [z_1,\ldots,z_m]$.

Für die Matrix A gilt somit

$$A = (s_1 \mid \cdots \mid s_n) = \begin{bmatrix} z_1^T \\ \cdots \\ z_m^T \end{bmatrix}.$$

Sei nun $\tilde{A}$ die beim Gaußschen Algorithmus durch Anwendung von Zeilenumformungen entstehende Endmatrix, also die Gaußsche Normalform von A, und seien

$$\tilde{s}_1,\ldots,\tilde{s}_n \text{ bzw. } \tilde{z}_1,\ldots,\tilde{z}_m$$

die zugehörigen Spalten– bzw. Zeilenvektoren. Wir wollen überlegen, welcher Zusammenhang zwischen den Spaltenvektoren s_j und $\tilde{s}_j$, $j = 1,\ldots,n$, bzw. den Zeilenvektoren z_i und $\tilde{z}_i$, $i = 1,\ldots,m$, besteht.

Satz 13. *Es ist* $[\tilde{z}_1,\ldots,\tilde{z}_m] = W$ *und diejenigen Vektoren* $\tilde{z}_i$, *welche vom Nullvektor verschieden sind, bilden eine (besonders einfache) Basis von* W.

Beweis. Da $\tilde{A}$ durch endlich viele elementare Zeilenumformungen aus A entstanden ist, sind die Vektoren $\tilde{z}_1,\ldots,\tilde{z}_m$ Linearkombinationen der ursprünglichen Vektoren $z_1,\ldots,z_m$. Also gilt $[\tilde{z}_1,\ldots,\tilde{z}_m] \subset W$. Da jede der angewendeten Zeilenumformungen wieder rückgängig gemacht werden kann, ensteht umgekehrt A durch endlich viele Zeilenumformungen aus $\tilde{A}$. Somit sind die Vektoren $z_1,\ldots,z_m$ Linearkombinationen der Vektoren $\tilde{z}_1,\ldots,\tilde{z}_m$ und es gilt $W \subset [\tilde{z}_1,\ldots,\tilde{z}_m]$.

Weiterhin erkennt man aus der Gestalt der Normalform

$$\tilde{A} = \left[\begin{array}{ccccccccccccc}
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * \\
0 & \cdots & \cdots & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots \\
\vdots & & & & & & & \ddots & & & & & \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & * & \cdots & * & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & * & \cdots & * \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & & & & & & & & & & & & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0
\end{array}\right] \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{l} \} k \\ \\ \\ \\ \\ \\ \} m-k \\ \\ \end{array}$$

unmittelbar, daß die ersten k Zeilen linear unabhängig sind. Das sind aber genau die Vektoren $\tilde{z}_i$, die von o verschieden sind. ∎

Nicht ganz so einfach ist der Zusammenhang zwischen den alten und den neuen Spaltenvektoren, denn im allgemeinen ist der Untervektorraum $[\tilde{s}_1,\ldots,\tilde{s}_n]$ von U verschieden. Es gilt aber:

Satz 14. *Es ist* $\dim [\tilde{s}_1,\ldots,\tilde{s}_n] = \dim U$. *Diejenigen Vektoren* $s_{j_1},\ldots,s_{j_k}$, *deren Indizes* $j_1,\ldots,j_k \in \{1,\ldots,n\}$ *zu den "Treppenstufen" in* $\tilde{A}$ *gehören, die also bei dem Gaußschen Algorithmus in Vektoren der Standardbasis*

$$\tilde{s}_{j_1} = e_1 , \ldots , \tilde{s}_{j_k} = e_k$$

übergehen, bilden eine Basis von U.

Beweis. Wir betrachten die linearen Gleichungssysteme $A\ y = o$ bzw. $\tilde{A}\ y = o$, $y = (y_1,\ldots,y_n) \in \mathbb{K}^n$, die wir in der Form

$$y_1\ s_1 + \cdots + y_n\ s_n = o \text{ bzw. } y_1\ \tilde{s}_1 + \cdots + y_n\ \tilde{s}_n = o$$

schreiben. Sie haben dieselbe Lösungsmenge. Aus der Gestalt von $\tilde{A}$ ergibt sich, daß die y_j mit $j \notin \{j_1,\ldots,j_k\}$ beliebig gewählt werden können. Damit ist jeder Vektor s_j mit $j \notin \{j_1,\ldots,j_k\}$ Linearkombination von $s_{j_1},\ldots,s_{j_k}$ und ebenso jeder Vektor $\tilde{s}_j$ mit $j \notin \{j_1,\ldots,j_k\}$ Linearkombination von $\tilde{s}_{j_1},\ldots,\tilde{s}_{j_k}$. Also gilt

$$U = [s_{j_1},\ldots,s_{j_k}] \text{ bzw. } [\tilde{s}_1,\ldots,\tilde{s}_n] = [\tilde{s}_{j_1},\ldots,\tilde{s}_{j_k}]$$

und die Vektoren $\tilde{s}_{j_1},\ldots,\tilde{s}_{j_k}$ sind linear unabhängig. Ist $y_{j_1}\ s_{j_1} + \cdots + y_{j_k}\ s_{j_k} = o$, $y_{j_1},\ldots,y_{j_k} \in \mathbb{K}$, so ist

$$y = (0,\ldots,0,\underbrace{y_{j_1}}_{j_1\text{-te Stelle}},0,\ldots,0,y_{j_2},\ldots,\ldots,\underbrace{y_{j_k}}_{j_k\text{-te Stelle}},0,\ldots,0)$$

Lösung von $A\ y = o$ also auch von $\tilde{A}\ y = o$. Damit folgt aber sofort $y_{j_1} = \cdots = y_{j_k} = 0$. Die Vektoren $s_{j_1},\ldots,s_{j_k}$ sind daher linear unabhängig und bilden eine Basis von U. Somit gilt auch $\dim U = \dim [\tilde{s}_{j_1},\ldots,\tilde{s}_{j_k}] = \dim [\tilde{s}_1,\ldots,\tilde{s}_n]$. ∎

Zusammenfassung. Es seien m Vektoren $x_1,\ldots,x_m \in \mathbb{K}^n$ gegeben und es soll der Untervektorraum $U = [x_1,\ldots,x_m] \subset \mathbb{K}^n$ untersucht werden.

Ist man an einer Basis von U in "Treppenform" interessiert, so wende man den Gaußschen Algorithmus auf die Matrix

$$\begin{bmatrix} x_1^T \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ x_m^T \end{bmatrix}$$

an. Man kann dann auch sagen, daß man *elementare Spaltenumformungen* auf $x_1,\ldots,x_m$ anwendet.

Ist man an der Frage interessiert, welche der Vektoren $x_1,\ldots,x_m$ eine Basis von

U bilden, so muß man den Gaußschen Algorithmus auf die Matrix

$$(x_1 \mid \cdots \mid x_m)$$

anwenden.

Zur Bestimmung der Dimension von U können beide Verfahren benutzt werden.

Es sei nun wieder eine Matrix $A \in \mathbb{K}^{m\times n}$ gegeben mit den Spalten $s_1,\ldots,s_n$ aus $\mathbb{K}^m$ und den Zeilen $z_1^\top,\ldots,z_m^\top$, $z_i \in \mathbb{K}^n$. Die Zahl dim $[s_1,\ldots,s_n]$ heißt der *Spaltenrang* von A, die Zahl dim $[z_1,\ldots,z_m]$ heißt der *Zeilenrang* von A.

Satz 15 und Definition. *Zeilenrang und Spaltenrang einer Matrix $A \in \mathbb{K}^{m\times n}$ sind gleich. Wir nennen diese Zahl den Rang von A, Schreibweise:* Rang A *oder* Rg A.

Beweis. Nach Satz 13 ist der Zeilenrang von A gleich der Anzahl k der Treppen in der Normalform $\tilde{A}$. Nach Satz 14 ist auch der Spaltenrang gleich k. ∎

Bemerkung. Der Rang einer Matrix A ist also die Maximalzahl linear unabhängiger Spalten (Zeilen) von A. Er läßt sich mit Hilfe des Gaußschen Algorithmus bestimmen.

Korollar 16. **(a)** *Für alle $A \in \mathbb{K}^{m\times n}$ gilt:* Rg A = Rg $A^\top$.

(b) *Für alle $A \in \mathbb{K}^{n\times n}$ gilt: A ist genau dann regulär, wenn* Rg $A = n$.

(c) *Für alle $A \in \mathbb{K}^{m\times n}$ und alle $b \in \mathbb{K}^m$ gilt: Das lineare Gleichungssystem $A\,x = b$ ist genau dann lösbar, wenn* Rg A = Rg $(A \mid b)$.

Bei homogenen linearen Gleichungssystemen können wir nun auch eine Aussage über die Dimension des Lösungsraumes machen. Es gilt

Korollar 17. *Es sei $A \in \mathbb{K}^{m\times n}$. Dann hat der Lösungsraum L des homogenen linearen Gleichungssystems $A\,x = o$ die Dimension $n -$ Rg A.*

Beweis. Die Matrix A habe die Normalform $\tilde{A}$ wie auf Seite 113. L und der Lösungsraum $\tilde{L}$ von $\tilde{A}\,x = o$ stimmen überein, und in jeder Lösung $x \in \tilde{L}$ sind die x_j mit $j \notin \{j_1,\ldots,j_k\}$ frei wählbar, und die restlichen x_j sind dann eindeutig festgelegt. Dabei

sind $j_1,\ldots,j_k$ wieder die Indizes, für die

$$\tilde{s}_{j_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \tilde{s}_{j_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots.$$

gilt. Somit folgt $\dim L = n - k = n - \operatorname{Rg} A$. ∎

Bemerkung. Jeder Untervektorraum $U \subset \mathbb{K}^n$ ist Lösungsraum eines geeigneten homogenen linearen Gleichungssystems:

Es sei $U = [x_1,\ldots,x_m] \subset \mathbb{K}^n$ und $\dim U = r$. Wir schreiben die Vektoren $x_1,\ldots,x_m$ als Zeilen einer Matrix

$$A = \begin{bmatrix} x_1^\top \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ x_m^\top \end{bmatrix} \in \mathbb{K}^{m\times n}$$

und betrachten das lineare Gleichungssystem $A\ y = o$, dessen Lösungsraum nach Korollar 17 die Dimension $k = n - \operatorname{Rg} A = n - \dim U = n - r$ hat.

In diesem Lösungsraum wählen wir eine Basis $\{y_1,\ldots,y_k\}$ und schreiben die Basisvektoren wieder als Zeilen einer Matrix

$$B = \begin{bmatrix} y_1^\top \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ y_k^\top \end{bmatrix} \in \mathbb{K}^{k\times n}$$

mit $\operatorname{Rg} B = k$. Dann ist U der Lösungsraum des linearen Gleichungssystems $B\ x = o$.

Beweis. Es sei L_h der Lösungsraum von $B\ x = o$. Nach Korollar 17 gilt dann $\dim L_h = n - \operatorname{Rg} B = n - k = r$. Also ist $\dim L_h = \dim U$. Aus $A\ y_j = o$, $j = 1,\ldots,k$, folgt $x_i^\top y_j = 0$ für $i = 1,\ldots,m$ und $j = 1,\ldots,k$ und somit auch $y_j^\top x_i = 0$ für $j = 1,\ldots,k$ und

$i = 1,...,m$. Also gilt $B\,x_i = o$, $i = 1,...,m$ und daher $U \subset L_h$. Aus der Gleichheit der Dimensionen folgt schließlich $U = L_h$. ∎

Beispiel. Das eben beschriebene Verfahren ist oft praktisch bei der Bestimmung des Durchschnitts von Untervektorräumen des $\mathbb{K}^n$.

Im $\mathbb{R}^5$ seien die Untervektorräume

$$U_1 = \left[\begin{bmatrix}1\\-1\\-1\\-2\\1\end{bmatrix}, \begin{bmatrix}0\\3\\3\\3\\0\end{bmatrix}, \begin{bmatrix}1\\-3\\1\\-2\\4\end{bmatrix}\right] , \quad U_2 = \left[\begin{bmatrix}-1\\0\\-4\\-5\\1\end{bmatrix}, \begin{bmatrix}-5\\-1\\2\\2\\-6\end{bmatrix}, \begin{bmatrix}1\\2\\-1\\3\\2\end{bmatrix}, \begin{bmatrix}3\\1\\0\\3\\3\end{bmatrix}\right]$$

gegeben. Wir suchen Ihren Durchschnitt $U_1 \cap U_2$. Hierzu stellen wir U_1 und U_2 durch lineare Gleichungssysteme dar und schreiben deshalb die erzeugenden Vektoren jeweils als Zeilen einer Matrix eines LGS:

$$A_1 = \begin{bmatrix}1 & -1 & -1 & -2 & 1\\0 & 3 & 3 & 3 & 0\\1 & -3 & 1 & -2 & 4\end{bmatrix} , \quad A_2 = \begin{bmatrix}-1 & 0 & -4 & -5 & 1\\-5 & -1 & 2 & 2 & -6\\1 & 2 & -1 & 3 & 2\\3 & 1 & 0 & 3 & 3\end{bmatrix}$$

Mit Hilfe des Gaußschen Verfahrens erhalten wir die Gaußschen Normalformen

$$\tilde{A}_1 = \begin{bmatrix}1 & 0 & 0 & -1 & 1\\0 & 1 & 0 & 1/2 & -3/4\\0 & 0 & 1 & 1/2 & 3/4\end{bmatrix} , \quad \tilde{A}_2 = \begin{bmatrix}1 & 0 & 0 & 0 & 12/13\\0 & 1 & 0 & 0 & 6/13\\0 & 0 & 1 & 0 & -5/13\\0 & 0 & 0 & 1 & -1/13\end{bmatrix}$$

Die Lösungsräume L_1 von $A_1\,x = o$ und L_2 von $A_2\,x = o$ sind

$$L_1 = \left[\begin{bmatrix}2\\-1\\-1\\2\\0\end{bmatrix}, \begin{bmatrix}-4\\3\\-3\\0\\4\end{bmatrix}\right] , \quad L_2 = \left[\begin{bmatrix}-12\\-6\\5\\1\\13\end{bmatrix}\right] .$$

Damit ist U_1 Lösungsraum des linearen Gleichungssystems

$$\begin{aligned} 2\,x_1 - x_2 - x_3 + 2\,x_4 &= 0 \\ -4\,x_1 + 3\,x_2 - 3\,x_3 + 4\,x_5 &= 0 \end{aligned}$$

und U_2 ist Lösungsraum von

$$-12\,x_1 - 6\,x_2 + 5\,x_3 + x_4 + 13\,x_5 = 0\,.$$

$U_1 \cap U_2$ ist nun Lösungsmenge des homogenen linearen Gleichungssystems mit der Matrix

$$\begin{bmatrix} 2 & -1 & -1 & 2 & 0 \\ -4 & 3 & -3 & 0 & 4 \\ -12 & -6 & 5 & 1 & 13 \end{bmatrix},$$

deren Normalform

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 \end{bmatrix}$$

lautet. Daraus folgt

$$U_1 \cap U_2 = \left[\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}\right].$$

§ 4 Summen und Faktorräume

Die Vereinigung von Untervektorräumen ist im allgemeinen kein Untervektorraum, wohl aber ihre Summe, die folgendermaßen definiert ist.

Definitionen. **(a)** Unter der *Summe* von k Teilmengen $A_1,...,A_k$ des $\mathbb{K}$–Vektorraumes V, $k \geq 2$, verstehen wir die Menge

$$A_1 + \cdots + A_k := \{x_1 + \cdots + x_k \mid x_i \in A_i,\ i = 1,...,k\}, \text{ Kurzschreibweise: } \sum_{i=1}^{k} A_i .$$

Ist $A = \{x\}$ und $B \subset V$, so schreiben wir statt $\{x\} + B$ kurz $x + B$.

(b) Sind $U_1,...,U_k$, $k \geq 2$, Untervektorräume von V und gilt für $i = 1,...,k$

$$U_i \cap \sum_{\substack{j=1 \\ j \neq i}}^{k} U_j = \{o\} ,$$

so heißt die Summe $U_1 + \cdots + U_k$ *direkt*. Wir schreiben in diesem Fall $U_1 \oplus \cdots \oplus U_k$.

Bemerkungen. **(a)** Die Summe von Teilmengen eines Vektorraumes ist im allgemeinen kein Untervektorraum. Man betrachte etwa die Summe $\{x\} + \{y\} = \{x + y\}$, die für $x + y \neq o$ kein Untervektorraum ist. Dagegen ist die Summe von Untervektorräumen eines Vektorraumes V stets ein Untervektorraum von V, wie man mit Hilfe von Satz 2 sofort erkennt.

(b) Für alle Teilmengen $A_1,...,A_k$ von V gilt $[A_1] + \cdots + [A_k] = [A_1 \cup \cdots \cup A_k]$.

Beweis. Nach Satz 4 ist $[A_1] + \cdots + [A_k] \subset [A_1 \cup \cdots \cup A_k]$. Umgekehrt gilt auch $A_1 \cup \cdots \cup A_k \subset [A_1] + \cdots + [A_k]$ und somit $[A_1 \cup \cdots \cup A_k] \subset [A_1] + \cdots + [A_k]$. ∎

Im folgenden Satz klären wir, wann eine Summe von Untervektorräumen direkt ist.

Satz 18. *Es seien V ein $\mathbb{K}$–Vektorraum und $U_1,\dots,U_k$, $k \geq 2$, Untervektorräume von V. Dann gilt:*

Die Summe $U = U_1 + \cdots + U_k$ ist genau dann direkt, wenn jeder Vektor $x \in U$ eine eindeutige Darstellung $x = u_1 + \cdots + u_k$ mit $u_i \in U_i$ für $i = 1,\dots,k$ besitzt.

Beweis. Sei $U = U_1 \oplus \cdots \oplus U_k$ und seien $x = u_1 + \cdots + u_k$ und $x = \tilde{u}_1 + \cdots + \tilde{u}_k$ Darstellungen von $x \in U$. Dann folgt $(u_1 - \tilde{u}_1) + \cdots + (u_k - \tilde{u}_k) = o$, also

$$u_i - \tilde{u}_i = \sum_{\substack{j=1 \\ j \neq i}}^{k} (\tilde{u}_j - u_j).$$

Auf der linken Seite steht ein Element von U_i, auf der rechten ein Element der Summe $U_1 + \cdots + U_{i-1} + U_{i+1} + \cdots + U_k$. Somit ist $u_i - \tilde{u}_i = o$. Dies gilt für alle $i = 1,\dots,k$, die Darstellung ist daher eindeutig.

Die Umkehrung beweisen wir indirekt und nehmen dazu an, die Summe sei nicht direkt. Es gibt somit ein $i \in \{1,\dots,k\}$ mit

$$U_i \cap \sum_{\substack{j=1 \\ j \neq i}}^{k} U_j \neq \{o\}.$$

Sei $u_i \neq o$ aus diesem Durchschnitt. Dann gilt $u_i = u_1 + \cdots + u_{i-1} + u_{i+1} + \cdots + u_k$ mit $u_j \in U_j$, $j = 1,\dots,k$, $j \neq i$, und $u_i = u_i$, ein Widerspruch zur Voraussetzung. ∎

Für die Summe zweier Untervektorräume gilt der folgende Dimensionssatz.

Satz 19 (Dimensionssatz). *Es seien V ein $\mathbb{K}$–Vektorraum und U, W Untervektorräume von V. Dann gilt:*

$$\dim (U + W) + \dim (U \cap W) = \dim U + \dim W.$$

Beweis. Für $\dim U = \infty$ oder $\dim W = \infty$ ist auch $\dim (U + W) = \infty$ und die Aussage ist trivial. Seien also $\dim U = m < \infty$ und $\dim W = n < \infty$. Nach Satz 12 gilt dann auch $\dim (U \cap W) = k < \infty$. Im Fall $k > 0$ sei $B_0 = \{u_1,\dots,u_k\}$ eine Basis von

$U \cap W$. Nach dem Basisergänzungssatz können wir B_0 zu einer Basis B_1 von U und zu einer Basis B_2 von W ergänzen:

$$B_1 = \{u_1,\dots,u_k\,,\ u_{k+1},\dots,u_m\}\,,\quad B_2 = \{u_1,\dots,u_k\,,\ w_{k+1},\dots,w_n\}.$$

Für $k = 0$ ist $B_0 = \emptyset$; hier können $B_1 = \{u_1,\dots,u_m\}$ und $B_2 = \{w_1,\dots,w_n\}$ beliebige Basen von U bzw. W sein. Nun gilt $U + W = [u_1,\dots,u_m\,,w_{k+1},\dots,w_n]$, d.h. $U + W$ ist ebenfalls endlich dimensional.

Wir zeigen, daß die Vektoren $u_1,\dots,u_m\,,w_{k+1},\dots,w_n$ linear unabhängig sind. Aus $a_1\,u_1 + \cdots + a_m\,u_m + b_{k+1}\,w_{k+1} + \cdots + b_n\,w_n = o$ erhalten wir $a_1\,u_1 + \cdots + a_m\,u_m = -b_{k+1}\,w_{k+1} - \cdots - b_n\,w_n$ und somit $b_{k+1}\,w_{k+1} + \cdots + b_n\,w_n \in U \cap W$. Daraus folgt $b_{k+1} = \cdots = b_n = 0$ und damit $a_1\,u_1 + \cdots + a_m\,u_m = o$. Wegen der linearen Unabhängigkeit von B_1 ist $a_1 = \cdots = a_m = 0$.

Also ist $B_1 \cup B_2$ Basis von $U + W$ und wir erhalten $\dim(U + W) = m + n - k$ $= \dim U + \dim W - \dim(U \cap W)$. ∎

Korollar 20. *Es gilt* $\dim(U \oplus W) = \dim U + \dim W$.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $U, W \subset V$ Untervektorräume. Gilt $V = U \oplus W$, so heißen U und W *komplementär*, W heißt ein *Komplementärraum* von U und analog heißt U ein *Komplementärraum* von W.

Wir zeigen, daß jeder Untervektorraum einen Komplementärraum besitzt.

Satz 21. *Es seien V ein $\mathbb{K}$–Vektorraum und U ein Untervektorraum von V. Dann gibt es zu U einen Komplementärraum.*

Beweis. Wir ergänzen eine Basis B' von U zu einer Basis B von V. Dann ist $W = [B \setminus B']$ ein solcher Komplementärraum. ∎

Bemerkung. In einem endlich dimensionalen Vektorraum V kann man zu jedem Untervektorraum U einen Komplementärraum konkret angeben. Im allgemeinen gibt es hierzu aber viele Möglichkeiten und kein Untervektorraum von V bietet sich

dann in natürlicher Weise als Komplementärraum von U an.

Beispiel. Sei $V = \mathbb{R}^5$. Der Untervektorraum U von V werde von den Vektoren

$$u_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \\ -2 \\ -1 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix}$$

erzeugt. Wir suchen einen Komplementärraum W von U. Dazu bestimmen wir eine Basis von U in "Treppenform", die durch Hinzunahme weiterer "Treppen" zu einer Basis von V ergänzt wird.

Der Gaußsche Algorithmus, angewendet auf die Matrix

$$\begin{bmatrix} 1 & 2 & -1 & 1 & -1 \\ 2 & -2 & 1 & -2 & -1 \\ 1 & 2 & -1 & -1 & -2 \end{bmatrix}$$

ergibt die Normalform

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1/2 \\ 0 & 1 & -1/2 & 0 & -1/2 \\ 0 & 0 & 0 & 1 & 1/2 \end{bmatrix}$$

Wir ergänzen diese zu einer Treppennormalform mit 5 Stufen

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1/2 \\ 0 & 1 & -1/2 & 0 & -1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Somit ist

$$W = \left[\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right] \text{ ein Komplementärraum von } U = \left[\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right].$$

Mit einer anderen Ergänzung der Treppennormalform erhält man einen anderen Komplementärraum, z.B.

$$W' = \left[\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right].$$

In unendlich dimensionalen Vektorräumen ist es im allgemeinen nicht möglich, Komplementärräume konkret anzugeben.

Wir beschreiben nun eine natürliche Konstruktion, die von U ausgehend zu einem neuen Vektorraum führt, der dann in allen praktischen Problemen die Rolle eines Komplementärraumes von U spielt. Diese Konstruktion ist unabhängig von der Dimension von V.

Gegeben seien ein $\mathbb{K}$–Vektorraum V und ein Untervektorraum $U \subset V$. Mit Hilfe von U wird auf V eine Relation $\sim$ erklärt durch

$$x \sim y \;:\Longleftrightarrow\; x - y \in U \;,\; x,y \in V \,.$$

Nach Kap. 1.2 (S.44) ist $\sim$ eine Äquivalenzrelation. Diese ist nun auch mit den Vektorraumverknüpfungen verträglich, d.h. es gilt für alle $x_1, x_2, y_1, y_2 \in V$ und alle $a \in \mathbb{K}$:

(*)
$$\begin{array}{c} \text{Aus } x_1 \sim y_1 \,,\, x_2 \sim y_2 \text{ folgt stets } x_1 + x_2 \sim y_1 + y_2 \,. \\ \\ \text{Aus } x_1 \sim y_1 \text{ folgt stets } a\, x_1 \sim a\, y_1 \,. \end{array}$$

Aus $x_1 \sim y_1$, $x_2 \sim y_2$ folgt nämlich $x_1 - y_1 \in U$, $x_2 - y_2 \in U$, und damit gilt auch $(x_1 + x_2) - (y_1 + y_2) \in U$. Ebenso folgt aus $x_1 \sim y_1$ zunächst $x_1 - y_1 \in U$ und daher auch $a\, x_1 - a\, y_1 = a\,(x_1 - y_1) \in U$.

Für die Faktormenge $V/_{\sim}$ schreiben wir wieder $V/_U$. Jede Äquivalenzklasse $[x]_{\sim}$ ist eine Summe:

$$[x]_{\sim} = \{y \mid y - x \in U\} = x + U.$$

Die Faktormenge $V/_U$ wird mit den folgenden Verknüpfungen zu einem $\mathbb{K}$–Vektorraum:

$$[x]_\sim + [y]_\sim := [x+y]_\sim \quad \text{für alle } [x]_\sim, [y]_\sim \in V/_U,$$
$$a\,[x]_\sim := [a\,x]_\sim \quad \text{für alle } a \in \mathbb{K}, [x]_\sim \in V/_U.$$

Wegen der Eigenschaften (*) sind diese Definitionen unabhängig von der speziellen Wahl der Repräsentanten.

Satz 22 und Definition. *Es seien V ein $\mathbb{K}$–Vektorraum und U ein Untervektorraum von V. Dann ist $V/_U$ mit den durch*

$$[x]_\sim + [y]_\sim := [x+y]_\sim, \quad [x]_\sim, [y]_\sim \in V/_U,$$
$$a[x]_\sim := [a\,x]_\sim, \quad a \in \mathbb{K}, [x]_\sim \in V/_U,$$

gegebenen Verknüpfungen ein $\mathbb{K}$–Vektorraum; er heißt Faktor– oder Quotientenraum von V nach U.

Beweis. Die Vektorraumaxiome lassen sich unmittelbar nachprüfen, wobei das Neutralelement in $(V/_U, +)$ die Klasse $[o]_\sim$ und das Inverse zur Klasse $[x]_\sim$ die Klasse $[-x]_\sim$ ist. ∎

Bemerkung. Für $U = \{o\}$ gilt $[x]_\sim = x + \{o\} = \{x\}$. In diesem Fall kann $V/_{\{o\}}$ mit V identifiziert werden. Für $U = V$ erhalten wir $V/_U = \{[o]_\sim\}$.

Satz 23. *Es seien V ein $\mathbb{K}$–Vektorraum und U ein Untervektorraum von V. Dann gilt:*

$$\dim V/_U + \dim U = \dim V.$$

Beweis. Sei B eine Basis von U. Wir ergänzen B zu einer Basis $B \cup B'$ von V mit $B \cap B' = \emptyset$. Dann gilt $V = U \oplus [B']$. Wir zeigen, daß $\tilde{B} := \{[x]_\sim \mid x \in B'\}$ eine Basis von $V/_U$ ist und daß $|\tilde{B}| = |B'|$ gilt:

$\tilde{B}$ ist ein Erzeugendensystem: Sei $[v]_\sim \in V/_U$ beliebig. Der Vektor $v \in V$ hat

die Darstellung $v = a_1 u_1 + \cdots + a_k u_k + b_1 x_1 + \cdots + b_m x_m$ mit $u_i \in B$, $a_i \in \mathbb{K}$ für $i = 1,\ldots,k$ und mit $x_j \in B'$, $b_j \in \mathbb{K}$ für $j = 1,\ldots,m$. Daraus folgt $[v]_\sim = b_1 [x_1]_\sim + \cdots + b_m [x_m]_\sim$ mit $[x_i]_\sim \in \tilde{B}$ für $i = 1,\ldots,m$.

$\tilde{B}$ ist linear unabhängig und es gilt $|\tilde{B}| = |B'|$: Hierzu genügt es zu zeigen, daß für alle $k \in \mathbb{N}$ und alle linear unabhängigen Vektoren $x_1,\ldots,x_k \in B'$ die Vektoren $[x_1]_\sim,\ldots,[x_k]_\sim$ ebenfalls linear unabhängig sind:

Aus $a_1 [x_1]_\sim + \cdots + a_k [x_k]_\sim = [o]_\sim$ folgt $[a_1 x_1 + \cdots + a_k x_k]_\sim = [o]_\sim$, also $a_1 x_1 + \cdots + a_k x_k \in U$. Wegen $U \cap [B'] = \{o\}$ folgt $a_1 x_1 + \cdots + a_k x_k = o$. Daraus ergibt sich $a_1 = \cdots = a_k = 0$.

Somit gilt: $\dim V = \dim U + \dim [B'] = \dim U + \dim {}^{V}/_{U}$. ∎

Aus dem Beweis von Satz 23 ergibt sich unmittelbar die Aussage (a) des folgenden Korollars.

Korollar 24. **(a)** *Es seien V ein $\mathbb{K}$-Vektorraum, U ein Untervektorraum von V und B eine Basis von U. Ist $B \cup B'$ eine Basis von V und $B \cap B' = \emptyset$, so ist $\{[x]_\sim \mid x \in B'\}$ eine Basis von ${}^{V}/_{U}$.*

(b) *Für jeden Komplementärraum W von U gilt ${}^{V}/_{U} \cong W$.*

Beispiel. Sei $U \subset \mathbb{R}^5$ wieder der Untervektorraum aus dem vorigen Beispiel. Dann bilden die Vektoren

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{bmatrix} \quad \text{bzw.} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

eine Basis von U bzw. von einem Komplementärraum W von U. Also ist

$$\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + U, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + U \right\}$$

eine Basis von ${}^{V}/_{U}$.

§ 5 Affine Unterräume eines Vektorraumes

Bisher haben wir die Äquivalenzklassen $x + U$, $x \in V$, U Untervektorraum von V, hauptsächlich im Zusammenhang mit dem Begriff Faktorraum als Elemente eines neuen Vektorraumes kennengelernt. In diesem Abschnitt dagegen wollen wir ihre geometrischen Eigenschaften als Teilmengen des Vektorraumes V untersuchen. Um dies auch durch die Sprechweise deutlich zu machen, geben wir ihnen gleich einen anderen Namen.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum, $x \in V$ und U ein Untervektorraum von V. Dann heißt die Teilmenge $L = x + U$ von V *affiner Unterraum* von V, und U heißt *Richtungsraum* oder kurz *Richtung* von L.

Bemerkung. Sind $L = x + U$ und $\tilde{L} = \tilde{x} + \tilde{U}$ affine Unterräume des $\mathbb{K}$–Vektorraumes V, so ist L genau dann eine Teilmenge von $\tilde{L}$, wenn $U \subset \tilde{U}$ und $x - \tilde{x} \in \tilde{U}$.

Beweis. Sei $x + U \subset \tilde{x} + \tilde{U}$. Dann ist $x \in \tilde{x} + \tilde{U}$, also $x - \tilde{x} \in \tilde{U}$. Für $y \in U$ folgt $x + y \in x + U \subset \tilde{x} + \tilde{U}$, also $x + y - \tilde{x} \in \tilde{U}$. Somit ist $y \in \tilde{U}$, und es gilt $U \subset \tilde{U}$. Umgekehrt folgt aus $x + y \in x + U$ wegen $x - \tilde{x} \in \tilde{U}$ und $U \subset \tilde{U}$ sofort $x + y = \tilde{x} + (x - \tilde{x}) + y \in \tilde{x} + \tilde{U}$. ∎

Aus $L = x + U = \tilde{x} + \tilde{U}$ folgt somit $U = \tilde{U}$, d.h. der Richtungsraum von L ist eindeutig bestimmt. Für den Vektor x gilt dies jedoch nicht, denn für jedes $\tilde{x} \in x + U$ gilt $\tilde{x} - x \in U$, also $x + U = \tilde{x} + U$.

Bezeichnungen und Bemerkungen. **(a)** Die *Dimension* eines affinen Unterraumes $L = x + U$ definieren wir durch $\dim L := \dim U$.

Die nulldimensionalen affinen Unterräume von V heißen *Punkte*. Es sind genau die einelementigen Mengen $x + \{o\} = \{x\}$, $x \in V$. Weil wir sie üblicherweise mit ihrem einzigen Element x identifizieren, nennen wir die Vektoren von V, insbesondere wenn wir geometrische Sachverhalte beschreiben wollen, ebenfalls *Punkte*. Der Punkt x in der Darstellung $L = x + U$ heißt dann auch der *Aufpunkt* von L.

Die eindimensionalen affinen Unterräume heißen *Geraden* und die zweidimensionalen affinen Unterräume heißen *Ebenen*. Ist V n–dimensional und $\dim L = n-1$, so heißt L eine *Hyperebene*.

(b) Affine Unterräume werden oft in Form einer Parameterdarstellung beschrieben:

Es sei $L = x_0 + U$ ein k–dimensionaler affiner Unterraum von V. Ist $\{u_1,...,u_k\}$ eine Basis von U, so läßt sich jeder Punkt $x \in L$ in der Form

$$x = x_0 + a_1 u_1 + \cdots + a_k u_k$$

mit $a_1,...,a_k \in \mathbb{K}$ darstellen. Man nennt dies eine *Parameterdarstellung* von L mit den *Richtungsvektoren* $u_1,...,u_k$ und den *Parametern* $a_1,...,a_k$.

Beispiele. Parameterdarstellung einer Geraden: $x = x_0 + a\,u$, $a \in \mathbb{K}$, $u \neq o$.
Parameterdarstellung einer Ebene: $x = x_0 + a_1 u_1 + a_2 u_2$, $a_1, a_2 \in \mathbb{K}$, u_1, u_2 linear unabhängig.

(c) Seien $A \in \mathbb{K}^{m\times n}$ und $b \in \mathbb{K}^m$. Die Lösungsmenge des linearen Gleichungssystems $A\,x = b$ ist entweder die leere Menge oder ein affiner Unterraum von $\mathbb{K}^n$ der Dimension $n - \operatorname{Rg} A$. Insbesondere ist die Lösungsmenge der linearen Gleichung $a_1 x_1 + \cdots + a_n x_n = b$, $b \in \mathbb{K}$, für $(a_1,...,a_n) \neq (0,...,0)$ eine Hyperebene in $\mathbb{K}^n$.

(d) Umgekehrt ist in $\mathbb{K}^n$ jeder k–dimensionale affine Unterraum auch Lösungsmenge eines linearen Gleichungssystems $A\,x = b$ mit $A \in \mathbb{K}^{m\times n}$, $\operatorname{Rg} A = n - k$ und $b \in \mathbb{K}^m$.

Beweis. Sei $L = x_0 + U$, $\dim U = k$. Der Untervektorraum U ist nach der Bemerkung von S.116 Lösungsmenge eines LGS $A\,x = o$ mit $A \in \mathbb{K}^{m\times n}$ und $\operatorname{Rg} A = n - k$. Damit gilt $x \in L$ genau dann, wenn $A\,(x - x_0) = o$, d.h. wenn $A\,x = A\,x_0 =: b$. ∎

Die Bemerkungen (c) und (d) lassen sich auch geometrisch formulieren: Der Schnitt endlich vieler Hyperebenen in $\mathbb{K}^n$ ist entweder leer oder ein affiner Unterraum von $\mathbb{K}^n$. Umgekehrt ist jeder k–dimensionale affine Unterraum in $\mathbb{K}^n$ Schnitt von endlich vielen Hyperebenen.

Allgemein gilt für den Schnitt affiner Unterräume die folgende Aussage.

Satz 25. *Es seien V ein $\mathbb{K}-$Vektorraum und $\mathcal{M}$ ein nichtleeres System affiner Unterräume von V. Dann ist der Schnitt*

$$M = \bigcap_{L \in \mathcal{M}} L$$

entweder leer oder ein affiner Unterraum von V mit Richtungsraum

$$U_M = \bigcap_{L \in \mathcal{M}} U_L .$$

Dabei bezeichnet U_L den Richtungsraum von L.

Beweis. Sei $M \neq \emptyset$. Dann gibt es einen Punkt $x \in V$, der in allen Unterräumen $L \in \mathcal{M}$ liegt. Diese lassen sich damit in der Form $L = x + U_L$ darstellen. Es folgt

$$\bigcap_{L \in \mathcal{M}} L = \bigcap_{L \in \mathcal{M}} x + U_L = x + \bigcap_{L \in \mathcal{M}} U_L = x + U_M .$$

Da nach Korollar 3 der Schnitt U_M ein Untervektorraum von V ist, gilt die Behauptung. ∎

Beispiel. Im $\mathbb{R}^4$ seien die affinen Unterräume L_1 und L_2 gegeben:

$$L_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \left[\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right] , \quad L_2 = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \left[\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right] .$$

Wir wollen ihren Schnitt bestimmen. Dies kann mit Hilfe der obigen Bemerkung (d) geschehen oder aber wie folgt:

$x \in L_1 \cap L_2 \iff$ es gibt reelle Zahlen a_1, a_2, a_3, b_1, b_2 mit

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} + a_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = x = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} .$$

Wir erhalten daraus ein inhomogenes lineares Gleichungssystem mit den Unbekannten a_1, a_2, a_3, b_1, b_2 und der zugehörigen erweiterten Matrix

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{array}\right]$$

Der Gaußsche Algorithmus führt nach einfacher Rechnung zu folgender Treppennormalform

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -2 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

Daraus lesen wir ab: $b_2 = 1$, $b_1 \in \mathbb{R}$ beliebig. Somit gilt

$$x = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} + b_1 \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \quad b_1 \in \mathbb{R}.$$

Also ist $L_1 \cap L_2$ eine Gerade.

Zum Abschluß dieses Paragraphen wollen wir die gegenseitige Lage affiner Unterräume an einigen Beispielen erläutern. Hierbei spielt der Begriff der Parallelität eine wichtige Rolle.

Definition. Die affinen Unterräume L_1 und L_2 heißen *parallel*, in Zeichen $L_1 \parallel L_2$, falls für die zugehörigen Richtungsräume U_1, U_2 gilt: $U_1 \subset U_2$ oder $U_2 \subset U_1$.

Bemerkungen. **(a)** Parallelität ist offensichtlich eine reflexive und symmetrische Relation, sie ist aber im allgemeinen nicht transitiv. So sind zwei sich schneidende Geraden in einer Ebene des Raumes zwar zu dieser Ebene parallel, aber sie selbst

sind nicht parallel. Beschränkt man sich jedoch auf affine Unterräume gleicher Dimension, so ist die Parallelität auch transitiv, also eine Äquivalenzrelation.

(b) Parallele affine Unterräume, die nicht ineinander enthalten sind, haben einen leeren Durchschnitt.

Speziell bei Geraden ist noch eine weitere Bezeichnung üblich: Zwei Geraden heißen *windschief*, falls sie weder parallel sind noch einen Punkt gemeinsam haben.

Zwei Geraden sind also genau dann windschief, wenn sie keinen Punkt gemeinsam haben und ihre Richtungen verschieden sind. In diesem Fall gibt es keine Ebene, die beide Geraden enthält (vgl. folgende Beispiele).

Beispiele. **(a)** Gegenseitige Lage zweier Geraden g und h in $\mathbb{R}^n$, $n \geq 2$:

Es seien $g = x + U_g$ und $h = y + U_h$.

$n = 2$: Ist $g \cap h \neq \emptyset$, so erhalten wir für $\dim (U_g \cap U_h) = 0$ als Schnittmenge einen Punkt und für $\dim (U_g \cap U_h) = 1$ folgt $U_g = U_h$, also $g = h$.

Ist $g \cap h = \emptyset$, so folgt aus $\dim (U_g \cap U_h) = 1$ wieder $U_g = U_h$, also $g \parallel h$, $g \neq h$. Der Fall $\dim (U_g \cap U_h) = 0$ kann hier nicht auftreten. Andernfalls wäre $\mathbb{R}^2$ direkte Summe von U_g und U_h und wir erhielten $x - y = z_1 + z_2$ mit $z_1 \in U_g$ und $z_2 \in U_h$. Also wäre $x - z_1 = y + z_2 \in g \cap h$ im Widerspruch zur Vorausetzung $g \cap h = \emptyset$.

$n \geq 3$: Für $g \cap h \neq \emptyset$ ergibt sich wie oben, daß entweder $g = h$ ist oder g und h sich in einem Punkt schneiden.

Für $g \cap h = \emptyset$ erhalten wir im Fall $\dim (U_g \cap U_h) = 0$, daß g und h windschief sind. Dann gibt es keine Ebene L, die sowohl g als auch h enthält. Andernfalls wäre $U_L = U_g \oplus U_h$ und $x - y \in U_g \oplus U_h$, woraus wieder $g \cap h \neq \emptyset$ folgte. Gilt $\dim (U_g \cap U_h) = 1$, so sind g und h parallel.

(b) Gegenseitige Lage zweier Ebenen L_1 und L_2 in $\mathbb{R}^n$, $n \geq 3$:

Es seien $L_1 = x_1 + U_1$ und $L_2 = x_2 + U_2$.

$n = 3$: Ist $L_1 \cap L_2 \neq \emptyset$, so ist der Schnitt nach Satz 25 ein affiner Unterraum. Je nach der Dimension des zugehörigen Richtungsraumes $U = U_1 \cap U_2$ erhalten wir

die folgenden Fälle: Für dim $U = 2$ ist $U_1 = U_2$, also $L_1 = L_2$ und für dim $U = 1$ ist der Schnitt eine Gerade. Der Fall dim $U = 0$ kann hier nicht auftreten, da andernfalls die Summe von U_1 und U_2 direkt wäre, also die Dimension 4 hätte, im Widerspruch dazu, daß $U_1 \oplus U_2$ ein Untervektorraum von $\mathbb{R}^3$ ist.

Ist $L_1 \cap L_2 = \emptyset$, so erhalten wir für dim $U = 2$, daß die Ebenen parallel sind. Die restlichen Fälle können nicht auftreten: Gilt nämlich dim $U = 1$, so ergänzen wir eine Basis $\{u\}$ von U mit geeigneten Vektoren $u_1 \in U_1$ und $u_2 \in U_2$ zu einer Basis $\{u_1, u_2, u\}$ von $\mathbb{R}^3$. Für die Aufpunkte $x_1 \in L_1$ und $x_2 \in L_2$ erhalten wir dann

$$x_1 - x_2 = a_1 u_1 + a_2 u_2 + a u, \text{ also } x_1 - a_1 u_1 - \frac{a}{2} u = x_2 + a_2 u_2 + \frac{a}{2} u$$

und somit den Widerspruch $L_1 \cap L_2 \neq \emptyset$. Der Fall dim $U = 0$ führt wieder zu dem Widerspruch dim $(U_1 \oplus U_2) = 4$.

$n = 4$: Hier kann für $L_1 \cap L_2 \neq \emptyset$ der Schnitt auch noch ein Punkt sein, und für $L_1 \cap L_2 = \emptyset$ ist jetzt auch der Fall dim $U = 1$ möglich. Dann sind die Ebenen nicht parallel und haben auch keinen Punkt gemeinsam, aber es gibt eine Gerade, die sowohl zu L_1 als auch L_2 parallel ist. Der Fall $L_1 \cap L_2 = \emptyset$ und dim $U = 0$ kann für $n = 4$ ebenfalls nicht auftreten, er führt wieder zu dem Widerspruch $L_1 \cap L_2 \neq \emptyset$.

Für $n > 4$ sind alle Fälle möglich.

Mit den gleichen Methoden wird die gegenseitige Lage von Geraden und Ebenen in $\mathbb{R}^n$, $n \geq 3$, untersucht. Wir überlassen die Einzelheiten der Durchführung als Übungsaufgabe.

Eine ausführlichere Behandlung affiner Unterräume wird im Rahmen der affinen Geometrie in Kapitel 6 erfolgen.

Kapitel 3 Lineare Abbildungen

Ebenso wie bei den Gruppen und den Körpern wollen wir auch bei den Vektorräumen die strukturverträglichen Abbildungen studieren. Wir werden sehen, daß sich die endlich dimensionalen Vektorräume als "isomorph" zu einem der Standardräume $\mathbb{K}^n$ erweisen (mit $\mathbb{K}^0 := \{0\}$), so daß wir damit alle "Typen" endlich dimensionaler Vektorräume erfaßt haben.

Zur Vereinfachung und auch zur Präzisierung der Dimension werden wir im folgenden statt von endlich dimensionalen Vektorräumen meist von n–dimensionalen Vektorräumen V sprechen und dabei den Fall $n = 0$, also $V = \{o\}$, einschließen (wenn es nicht ausdrücklich anders gesagt wird). Um lästige Fallunterscheidungen zu vermeiden, werden wir uns aber bei den Beweisen auf Dimensionen $n \geq 1$ konzentrieren und dann mit Basisvektoren $x_1,\ldots,x_n \in V$ arbeiten. Die Übertragung der Aussagen auf den Fall $V = \{o\}$ ist meist trivial und wird dem Leser überlassen.

§ 1 Definition und Eigenschaften linearer Abbildungen

Es seien V und W Vektorräume über demselben Körper $\mathbb{K}$. Eine Abbildung Φ von V nach W erhält die Vektorraumstruktur, wenn

$$\Phi\,(x+y) = \Phi\,(x) + \Phi\,(y)\,,$$
$$\Phi\,(a\,x) = a\,\Phi\,(x)$$

für alle $x,\ y \in V$ und alle $a \in \mathbb{K}$ gilt. Wir können beide Gleichungen zu einer einzigen zusammenfassen:

Definition. Es seien V und W $\mathbb{K}$–Vektorräume. Eine Abbildung $\Phi : V \longrightarrow W$ heißt *linear* oder (*Vektorraum–*) *Homomorphismus*, wenn

$$\Phi\,(a\,x + b\,y) = a\,\Phi\,(x) + b\,\Phi\,(y)$$

für alle $x,\ y \in V$ und alle $a,b \in \mathbb{K}$ gilt. Eine bijektive lineare Abbildung heißt (*Vektor-*

raum–) Isomorphismus. Zwei $\mathbb{K}$–Vektorräume V und W heißen *isomorph*, wenn es einen Isomorphismus $\Phi : V \longrightarrow W$ gibt, Schreibweise: $V \cong W$.

Offensichtlich ist $\cong$ eine reflexive Relation für $\mathbb{K}$–Vektorräume. Wie wir im nächsten Paragraphen sehen werden, ist $\cong$ auch symmetrisch und transitiv, also eine Äquivalenzrelation.

Weitere Bezeichnungen. Für spezielle lineare Abbildungen gibt es in der Literatur eine Reihe weiterer Bezeichnungen:

$\Phi : V \longrightarrow V$ linear : *Endomorphismus, lineare Selbstabbildung, lineare Transformation, linearer Operator* ,

$\Phi : V \longrightarrow V$ linear und bijektiv : *Automorphismus* ,

$\Phi : V \longrightarrow W$ linear und injektiv : *Monomorphismus* ,

$\Phi : V \longrightarrow W$ linear und surjektiv : *Epimorphismus* .

Bemerkungen. **(a)** Eine lineare Abbildung Φ ist insbesondere ein Homomorphismus der additiven Gruppen V und W. Damit gilt speziell $\Phi(o) = o$.

(b) Zu jeder linearen Abbildung $\Phi : V \longrightarrow W$ gehören in natürlicher Weise die Mengen

$$\text{Bild } \Phi := \Phi(V) \text{ und } \text{Kern } \Phi := \Phi^{-1}(\{o\}) .$$

Bild Φ ist ein Untervektorraum von W und Φ ist genau dann surjektiv, wenn Bild $\Phi = W$. Kern Φ ist ein Untervektorraum von V und Φ ist genau dann injektiv, wenn Kern $\Phi = \{o\}$.

Beispiele. **(a)** Es seien V, W $\mathbb{K}$–Vektorräume und $w_0 \in W$. Die konstante Abbildung

$$\begin{aligned} \Phi : V &\longrightarrow W \\ v &\longmapsto w_0 \end{aligned}$$

ist genau dann linear, wenn $w_0 = o$ gilt. Dann ist Bild $\Phi = \{o\}$ und Kern $\Phi = V$. Φ ist in diesem Fall die Nullabbildung, $\Phi = O$.

(b) Sei $A \in \mathbb{K}^{m \times n}$. Dann ist die Abbildung $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^m$, $x \longmapsto A\,x$, linear. Für

$A = (s_1|\cdots|s_n)$ erhalten wir

$$\text{Bild } \Phi = \{A\,x|\ x \in \mathbb{K}^n\} = \{x_1\,s_1 + \cdots + x_n\,s_n|\ x_1,\ldots,x_n \in \mathbb{K}\} = [s_1,\ldots,s_n].$$

Kern Φ ist der Lösungsraum des homogenen linearen Gleichungssystems $A\,x = o$.

So ist zum Beispiel die Abbildung

$$\Phi : \mathbb{R}^4 \longrightarrow \mathbb{R}^2\ ,\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \longmapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_1 + x_4 \end{bmatrix}$$

linear, und es gilt

$$\text{Bild } \Phi = \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right] = \mathbb{R}^2$$

sowie

$$\text{Kern } \Phi = \left[\begin{bmatrix} 0 \\ 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1/2 \\ 0 \\ 1 \end{bmatrix}\right].$$

(c) Die Abbildungen $\Phi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3\ ,(x_1,\ x_2,\ x_3) \longmapsto (x_1x_2,\ x_2,\ x_1x_3)$ und $\Phi : \mathbb{R} \longrightarrow \mathbb{R}^2\ ,\ x \longmapsto (x,\ x+2)$ sind nicht linear.

(d) Die Abbildung

$$D : \mathbb{R}[X] \longrightarrow \mathbb{R}[X]\ ,\ \sum_{i=0}^{n} a_i X^i \longmapsto \sum_{i=1}^{n} a_i \cdot i \cdot X^{i-1}$$

ist linear (*Ableitungsoperator*). Hier ist Bild $D = \mathbb{R}[X]$ und Kern D ist die Menge der konstanten Polynome, also Kern $D \cong \mathbb{R}$.

Schränken wir D ein auf den Untervektorraum $U = [X,X^2,\ldots]$ von $\mathbb{R}[X]$, so ist $D|_U : U \longrightarrow \mathbb{R}[X]$ sowohl surjektiv als auch injektiv, also sind U und $\mathbb{R}[X]$ isomorph.

(e) Es sei $V = \{f\colon [0,1] \longmapsto \mathbb{R} \mid f \text{ stetig}\}$. Die Abbildung

$$I\colon V \longrightarrow \mathbb{R}\ ,\ f \longmapsto \int_0^1 f(t)\,\mathrm{d}t$$

ist linear (*Integrationsoperator*). Hier ist Bild $I = \mathbb{R}$.

(f) Es sei $V = \{ (a_i) \in \mathbb{R}^{\mathbb{N}} \mid (a_i) \text{ konvergent} \}$. Die Abbildung

$$\Phi : V \longrightarrow \mathbb{R} \ , \ (a_i) \longmapsto \lim_{i \to \infty} a_i$$

ist linear. Hier ist Bild $\Phi = \mathbb{R}$ und Kern Φ ist die Menge der Nullfolgen.

Die Untervektorräume Kern $\Phi \subset V$ bzw. Bild $\Phi \subset W$ sind insbesondere auch Untergruppen von V bzw. W und Φ ist ein Gruppenhomomorphismus. Somit gilt der Homomorphiesatz für Gruppen (S.46). Weil V und W Vektorräume sind und Φ als lineare Abbildung auch mit der Skalarmultiplikation verträglich ist, können wir diesen Satz sofort auf Vektorräume übertragen. Des besseren Verständnisses wegen werden wir ihn nochmals vollständig beweisen.

Satz 1 (Homomorphiesatz für Vektorräume). *Es seien V und W zwei $\mathbb{K}$-Vektorräume und $\Phi : V \longrightarrow W$ eine lineare Abbildung. Dann gilt:*

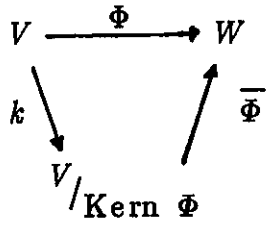
(a) *Die kanonische Abbildung $k : V \longrightarrow {}^{V}/_{\text{Kern } \Phi}$ ist linear.*

(b) *Es gibt eine injektive lineare Abbildung $\overline{\Phi} : {}^{V}/_{\text{Kern } \Phi} \longrightarrow W$ mit $\Phi = \overline{\Phi} \circ k$.*

(c) *Ist Φ surjektiv, so sind ${}^{V}/_{\text{Kern } \Phi}$ und W isomorph.*

Beweis. (a) $k(x + y) = [x + y]_{\sim} = [x]_{\sim} + [y]_{\sim} = k(x) + k(y)$ nach Definition der Addition im Faktorraum ${}^{V}/_{\text{Kern } \Phi}$ und $k(a\,x) = [a\,x]_{\sim} = a\,[x]_{\sim} = a\,k(x)$ nach Definition der Multiplikation mit Skalaren in ${}^{V}/_{\text{Kern } \Phi}$.

(b) Wir betrachten das Diagramm



und erklären die gesuchte Abbildung $\overline{\Phi}$ durch $\overline{\Phi}([x]_{\sim}) := \Phi(x)$. Diese Definition ist repräsentantenunabhängig, da $x \sim y$ nach Definition $x - y \in \text{Kern } \Phi$ bedeutet und

daraus $\Phi(x-y) = o$, also $\Phi(x) = \Phi(y)$ folgt.

$\overline{\Phi}$ ist linear: $\overline{\Phi}(a\,[x]_\sim + b\,[y]_\sim) = \overline{\Phi}([a\,x + b\,y]_\sim) = \Phi(a\,x + b\,y) = a\,\Phi(x) + b\,\Phi(y) = a\,\overline{\Phi}([x]_\sim) + b\,\overline{\Phi}([y]_\sim)$.

$\overline{\Phi}$ ist injektiv: Sei $\overline{\Phi}([x]_\sim) = \overline{\Phi}([y]_\sim)$. Dann gilt $\Phi(x) = \Phi(y)$ und $x - y \in$ Kern Φ, d.h. $x \sim y$ und somit $[x]_\sim = [y]_\sim$. Nach Definition von $\overline{\Phi}$ ist schließlich $\overline{\Phi} \circ k = \Phi$.

(c) Ist Φ surjektiv, so gibt es zu jedem $y \in W$ ein $x \in V$ mit $\Phi(x) = y$. Dann gilt $\overline{\Phi}([x]_\sim) = \Phi(x) = y$, d.h. auch $\overline{\Phi}$ ist surjektiv. Da $\overline{\Phi}$ nach (b) injektiv ist, folgt insgesamt, daß $\overline{\Phi}$ ein Isomorphismus ist. ∎

Korollar 2. *Es seien V, W $\mathbb{K}$-Vektorräume und $\Phi : V \longrightarrow W$ eine lineare Abbildung. Dann sind die Vektorräume $V/_{\text{Kern}\,\Phi}$ und* Bild Φ *isomorph.*

Beispiele. **(a)** $V/_V \cong \{o\}$ (Kor. 2 auf die Nullabbildung angewendet), $V/_{\{o\}} \cong V$ (Kor. 2 auf die Identität angewendet).

(b) Es sei V der Vektorraum der reellen konvergenten Folgen, und die lineare Abbildung $\Phi : V \longrightarrow \mathbb{R}$ sei gegeben durch $(x_i) \longmapsto \lim\limits_{i \to \infty} x_i$. Dann ist Bild $\Phi = \mathbb{R}$ und Kern Φ ist der Vektorraum der Nullfolgen in $\mathbb{R}$. Also gilt $V/_{\text{Kern}\,\Phi} \cong \mathbb{R}$.

Korollar 3. *Es seien V ein $\mathbb{K}$-Vektorraum und U, W Untervektorräume von V mit $V = U \oplus W$. Dann gilt*

$$V/_U \cong W \quad \textit{und} \quad V/_W \cong U.$$

Beweis. Wegen $V = U \oplus W$ besitzt jeder Vektor $x \in V$ eine eindeutige Darstellung $x = u + w$ mit $u \in U$ und $w \in W$. Die Abbildung $\pi : V \longrightarrow V$, $x \longmapsto w$ mit $x = u + w$ ist linear. Es ist Kern $\pi = U$, Bild $\pi = W$ und somit $V/_U = V/_{\text{Kern}\,\pi}$ isomorph zu W. Analog beweist man die andere Behauptung. ∎

Bemerkungen und Definitionen. **(a)** Die Abbildung π heißt *Projektion* von V auf den Untervektorraum W. Sie hat die Eigenschaft, daß $\pi^2 = \pi$ gilt. Allgemein nennt man jede lineare Abbildung $\pi : V \longrightarrow V$ mit $\pi^2 = \pi$ eine *Projektion.*

(b) Ist $\pi : V \longrightarrow V$ eine Projektion, so gilt $V = \text{Kern } \pi \oplus \text{Bild } \pi$.

Beweis. Sei $x \in V$. Wir setzen $x_1 := x - \pi(x)$ und $x_2 := \pi(x)$. Dann gilt $x_2 \in \text{Bild } \pi$ und $\pi(x_1) = \pi(x - \pi(x)) = \pi(x) - \pi^2(x) = o$, d.h. $x_1 \in \text{Kern } \pi$. Somit ist $x = x_1 + x_2$ und $x_1 \in \text{Kern } \pi$, $x_2 \in \text{Bild } \pi$.

Die Summe ist auch direkt: Sei $x \in \text{Kern } \pi \cap \text{Bild } \pi$. Dann gilt $\pi(x) = o$ und es existiert ein $y \in V$ mit $x = \pi(y)$. Daraus folgt $x = \pi(y) = \pi^2(y) = \pi(x) = o$. ∎

Aus $V = \text{Kern } \Phi \oplus \text{Bild } \Phi$ muß aber umgekehrt nicht folgen, daß Φ eine Projektion ist. So sind beispielsweise die bijektiven linearen Abbildungen von V, die von der Identität verschieden sind, keine Projektionen.

Wir wollen nun zeigen, wie lineare Abbildungen auf Basen wirken. Zur Vorbereitung dient der folgende Satz.

Satz 4. *Es seien V, W $\mathbb{K}$-Vektorräume, B eine Basis von V sowie $\Phi' : B \longrightarrow W$ eine beliebige Abbildung. Dann kann Φ' auf genau eine Weise zu einer linearen Abbildung $\Phi : V \longrightarrow W$ fortgesetzt werden. Insbesondere ist jede lineare Abbildung $\Phi : V \longrightarrow W$ durch ihre Werte auf B eindeutig festgelegt.*

Beweis. Jedes $x \in V$ besitzt nach der Bemerkung von S. 104/105 eine eindeutige Darstellung $x = a_1 v_1 + \cdots + a_n v_n$, $v_i \in B$, $a_i \in \mathbb{K}$. Wir setzen

$$(*) \qquad \Phi(x) := a_1 \Phi'(v_1) + \cdots + a_n \Phi'(v_n).$$

Dann ist Φ linear und $\Phi = \Phi'$ auf B. Ist $\Psi : V \longrightarrow W$ linear und gilt $\Psi = \Phi'$ auf B, so folgt aus (*), daß $\Psi = \Phi$ auf ganz V gilt. ∎

Satz 5. *Es seien V, W $\mathbb{K}$-Vektorräume, $\dim V = n$, $B = \{v_1, \ldots, v_n\}$ eine Basis von V und $\Phi : V \longrightarrow W$ linear. Dann gilt:*

(a) *Φ ist genau dann injektiv, wenn die Vektoren $\Phi(v_1), \ldots, \Phi(v_n)$ linear unabhängig sind.*

(b) *Φ ist genau dann surjektiv, wenn die Vektoren $\Phi(v_1), \ldots, \Phi(v_n)$ ein Erzeugendensystem von W bilden.*

(c) *Φ ist genau dann bijektiv, wenn die Vektoren $\Phi(v_1),...,\Phi(v_n)$ linear unabhängig und ein Erzeugendensystem von W sind, also eine Basis von W bilden.*

Beweis. (a) Die Vektoren $\Phi(v_1),...,\Phi(v_n)$ sind genau dann linear unabhängig, wenn aus $a_1\,\Phi(v_1) +...+ a_n\,\Phi(v_n) = o$ stets $a_1 = \cdots = a_n = 0$ folgt. Weil $v_1,...,v_n$ linear unabhängig sind, ist dies äquivalent zu der Aussage, daß aus $\Phi\,(a_1\,v_1 +\cdots+ a_n\,v_n) = o$ stets $a_1\,v_1 +\cdots+ a_n\,v_n = o$ folgt, daß also Kern $\Phi = \{o\}$ gilt und Φ somit injektiv ist.
(b) $\Phi(v_1),...,\Phi(v_n)$ bilden genau dann ein Erzeugendensystem von W, wenn jeder Vektor $y \in W$ sich als Linearkombination in der Form

$$y = a_1\,\Phi(v_1) +...+ a_n\,\Phi(v_n) = \Phi\,(a_1\,v_1 +\cdots+ a_n\,v_n)$$

darstellen läßt. Weil $v_1,...,v_n$ ein Erzeugendensystem von V bilden, ist dies wiederum äquivalent dazu, daß es zu jedem $y \in W$ ein $x \in V$ gibt mit $\Phi(x) = y$, daß also Φ surjektiv ist.
(c) folgt aus (a) und (b). ∎

Ist dim V = dim $W = n$, so sind nach Korollar 2.9 und Korollar 2.11 die Aussagen in Satz 5 (a),(b),(c) auf der rechten Seite alle äquivalent. Also erhalten wir direkt das folgende Isomorphiekriterium.

Korollar 6. *Es seien V, W $\mathbb{K}$-Vektorräume mit* dim V = dim $W = n$ *und Φ eine lineare Abbildung von V nach W. Dann sind folgende Aussagen äquivalent:*

(a) *Φ ist bijektiv,*

(b) *Φ ist surjektiv,*

(c) *Φ ist injektiv.*

Bemerkung. Für dim V = dim $W = \infty$ ist dieser Satz nicht mehr richtig. So ist z.B. der Ableitungsoperator aus Beispiel (d) von S. 134 zwar surjektiv, aber nicht injektiv.

Nun können wir einen Zusammenhang zwischen dem Isomorphiebegriff und der Dimension herstellen, falls die betrachteten $\mathbb{K}$-Vektorräume endlich dimensional sind.

Satz 7. *Es seien V und W endlich dimensionale $\mathbb{K}$–Vektorräume. Dann gilt:*

Die Vektorräume V und W sind genau dann isomorph, wenn sie dieselbe Dimension haben.

Beweis. Es seien $\dim V = n$, $\{v_1,...,v_n\}$ eine Basis von V und $\Phi : V \longrightarrow W$ ein Isomorphismus. Nach Satz 5 bilden $\Phi(v_1),...,\Phi(v_n)$ eine Basis von W. Also gilt $\dim V = \dim W$.

Ist umgekehrt $\dim V = \dim W = n$, so wählen wir Basen $\{v_1,...,v_n\}$ von V und $\{w_1,...,w_n\}$ von W. Nach Satz 4 existiert eine lineare Abbildung $\Phi : V \longrightarrow W$ mit $\Phi(v_i) = w_i$, $i = 1,...,n$. Nach Satz 5 ist Φ ein Isomorphismus. ∎

Jeder n–dimensionale $\mathbb{K}$–Vektorraum V ist also zu $\mathbb{K}^n$ isomorph. Wir werden deshalb in Zukunft oft den $\mathbb{K}^n$ als Prototyp des n–dimensionalen $\mathbb{K}$–Vektorraumes zugrunde legen.

Bemerkung. Für unendlich dimensionale Vektorräume gilt der obige Satz nur noch teilweise. Zwar besitzen isomorphe Vektorräume stets dieselbe Dimension, aber die unendlich dimensionalen Vektorräume $\mathbb{K}[X]$ und $\mathbb{K}^{\mathbb{N}}$ sind zum Beispiel nicht isomorph.

Satz 8 (Dimensionssatz für lineare Abbildungen). *Es seien V, W $\mathbb{K}$–Vektorräume und $\Phi : V \longrightarrow W$ linear. Dann gilt:*

$$\dim \text{Kern } \Phi + \dim \text{Bild } \Phi = \dim V .$$

Beweis. Nach Korollar 2 gilt

$$V/_{\text{Kern } \Phi} \cong \text{Bild } \Phi .$$

Daraus folgt

$$\dim V/_{\text{Kern } \Phi} = \dim \text{Bild } \Phi$$

und nach Satz 2.23 somit

$$\dim V = \dim V/_{\text{Kern } \Phi} + \dim \text{Kern } \Phi = \dim \text{Bild } \Phi + \dim \text{Kern } \Phi. \quad ∎$$

§ 2 Vektorräume linearer Abbildungen

Zu Beginn unserer Betrachtungen über Vektorräume haben wir schon Beispiele dieser Strukturen kennengelernt, bei denen die Elemente Abbildungen waren, etwa $W^A = \{f : A \longrightarrow W \mid f \text{ Abbildung}\}$, wobei A eine nichtleere Menge ist und W ein $\mathbb{K}$–Vektorraum. Von besonderem Interesse ist der Fall, daß die Menge A ebenfalls ein Vektorraum ist und $f : A \longrightarrow W$ linear.

Satz 9. *Es seien V, W, X $\mathbb{K}$–Vektorräume. Dann gilt :*

(a) *Sind $\Phi : V \longrightarrow W$ und $\Psi : W \longrightarrow X$ linear, so auch $\Psi \circ \Phi : V \longrightarrow X$.*

(b) *Ist $\Phi : V \longrightarrow W$ ein Isomorphismus, so auch $\Phi^{-1} : W \longrightarrow V$.*

(c) *Sind $\Phi, \Psi : V \longrightarrow W$ linear, so auch $\Phi + \Psi$ und $a\,\Phi$, $a \in \mathbb{K}$.*

Beweis. (a) und (c) sind trivial und werden als Übung überlassen.

(b) Es fehlt nur noch die Linearität von Φ^{-1} : Aus $\Phi\,(\Phi^{-1}(a\,x + b\,y)) = a\,x + b\,y = a\,\Phi\,(\Phi^{-1}(x)) + b\,\Phi\,(\Phi^{-1}(y)) = \Phi(a\,\Phi^{-1}(x) + b\,\Phi^{-1}(y))$ für alle $x,\ y \in W$ und alle $a, b \in \mathbb{K}$, folgt wegen der Injektivität $\Phi^{-1}(a\,x + b\,y) = a\,\Phi^{-1}(x) + b\,\Phi^{-1}(y)$. ∎

Bemerkungen und Bezeichnungen. **(a)** Wegen Satz 9 (a),(b) ist die Isomorphie von $\mathbb{K}$–Vektorräumen transitiv und symmetrisch.

(b) Nach Satz 9 (c) ist die Menge $\{\Phi : V \longrightarrow W \mid \Phi \text{ linear}\}$ ein Untervektorraum des $\mathbb{K}$–Vektorraumes W^V ; Bezeichnung: Hom (V, W) .

(c) Nach Satz 9 (a),(c) ist Hom (V, V) ein $\mathbb{K}$–Vektorraum und ein Ring mit Eins. Außerdem gilt in Hom (V, V) :

$$a\,(\Phi \circ \Psi) = (a\,\Phi) \circ \Psi = \Phi \circ (a\,\Psi) \quad , a \in \mathbb{K} .$$

Eine solche Struktur heißt *Algebra* ($\mathbb{K}$–Algebra). In der Literatur wird auch oft End (V) statt Hom (V, V) geschrieben.

(d) Nach Satz 9 (a),(b) ist die Teilmenge der Automorphismen in Hom (V, V) bezüglich der Komposition "$\circ$" eine Gruppe ; Bezeichnung: Aut(V).

Sind V und W endlich dimensional, so gilt dies aufgrund des folgenden Satzes auch für Hom (V,W).

Satz 10. *Es seien V und W endlich dimensionale $\mathbb{K}$-Vektorräume. Dann ist auch der Vektorraum* Hom (V,W) *endlich dimensional, und es gilt:*

$$\dim \text{Hom}\,(V,W) \;=\; \dim V \cdot \dim W.$$

Beweis. Für dim $V = 0$ oder dim $W = 0$ ist der Satz trivial. Seien nun dim $V = n$, dim $W = m$ und $\{v_1,\ldots,v_n\}$ bzw. $\{w_1,\ldots,w_m\}$ Basen von V bzw. W. Durch

$$\Phi_{ij}\,(v_k) := \delta_{jk}\, w_i \qquad (i = 1,\ldots,m\,;\, j,k = 1,\ldots,n)$$

werden nach Satz 4 $m \cdot n$ lineare Abbildungen $\Phi_{ij} : V \longrightarrow W$ definiert, von denen wir nun zeigen wollen, daß sie eine Basis von Hom (V,W) bilden.

1. Schritt : $\Phi_{11},\ldots,\Phi_{mn}$ sind linear unabhängig (und damit auch paarweise verschieden): Aus

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\,\Phi_{ij} \;=\; O \quad ,\; a_{ij} \in \mathbb{K}\,,$$

erhalten wir für alle $k = 1,\ldots,n$:

$$o \;=\; O\,(v_k) \;=\; \Big(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\,\Phi_{ij}\Big)(v_k) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\,\Phi_{ij}(v_k)$$

$$=\; \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\,\delta_{jk}\,w_i \;=\; \sum_{i=1}^{m} a_{ik}\,w_i\,.$$

Aus der linearen Unabhängigkeit der Vektoren $w_1,\ldots,w_m$ folgt somit $a_{ik} = 0$ für $i = 1,\ldots,m$ und $k = 1,\ldots,n$. Also sind die Abbildungen $\Phi_{11},\ldots,\Phi_{mn}$ linear unabhängig.

2. Schritt : $\{\Phi_{ij} \mid i = 1,\ldots,m\;;\; j = 1,\ldots,n\}$ ist ein Erzeugendensystem des Vektorraumes Hom (V,W): Sei $\Phi \in$ Hom (V,W) beliebig. Dann gibt es für jedes $k = 1,\ldots,n$ Zahlen $a_{1k},\ldots,a_{mk} \in \mathbb{K}$ mit

$$\Phi(v_k) = \sum_{i=1}^{m} a_{ik} w_i .$$

Daraus folgt für $k = 1,\ldots,n$

$$\Big(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\Phi_{ij}\Big)(v_k) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\Phi_{ij}(v_k)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\delta_{jk} w_i = \sum_{i=1}^{m} a_{ik} w_i = \Phi(v_k) ,$$

also nach Satz 4

$$\Phi = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\Phi_{ij} .$$

Somit ist $\{\Phi_{ij} \mid i = 1,\ldots,m\,;\, j = 1,\ldots,n\}$ eine Basis von Hom (V,W), und es gilt außerdem $\dim \mathrm{Hom}(V,W) = n \cdot m = \dim V \cdot \dim W$. ∎

Wir wenden uns nun einem wichtigen Spezialfall zu. Da der Körper $\mathbb{K}$ selbst ein $\mathbb{K}$–Vektorraum ist, können wir $W = \mathbb{K}$ setzen. Statt Hom $(V,\mathbb{K})$ schreiben wir V^*, und statt Φ schreiben wir häufig x^*.

V^* heißt der *Dualraum* von V, und $x^* \in V^*$ heißt *lineares Funktional* oder *Linearform* auf V.

Beispiele. **(a)** Die Beispiele (e) und (f) aus § 1 enthielten Linearformen, nämlich Integral und Grenzwert.

(b) Es seien A eine nichtleere Menge, $t_0 \in A$ und $V = \mathbb{K}^A$. Die durch

$$\Phi : \mathbb{K}^A \longrightarrow \mathbb{K},\ f \longmapsto \Phi(f) := f(t_0)$$

erklärte Abbildung ist eine Linearform (*Auswertungsfunktional*).

(c) Die Abbildung $\Phi : \mathbb{K}^{n\times n} \longrightarrow \mathbb{K}$, $A = (\!(a_{ij})\!) \longmapsto \sum_{i=1}^{n} a_{ii} =:$ Spur A ist eine Linearform.

Bemerkungen und Definitionen. **(a)** Gilt dim $V = n$, so ist nach Satz 10 auch dim $V^* = n$, also $V \cong V^*$.

Ist $B = \{v_1,...,v_n\}$ eine Basis von V, so wird, wenn wir in $\mathbb{K}$ als Basis $\{1\}$ wählen, nach dem Beweis von Satz 10 durch

$$v_j^*(v_k) = \delta_{jk}\ , \qquad j,k = 1,...,n$$

eine Basis $B^* = \{v_1^*,...,v_n^*\}$ von V^* festgelegt. B^* heißt die zur Basis B *duale Basis*. Sie ist nach Satz 4 eindeutig bestimmt. Der Basisvektor $v_j^* \in B^*$ heißt *j–tes Koordinatenfunktional* bezüglich der Basis B.

Dieser Name erklärt sich dadurch, daß für alle $x \in V$ und alle $x^* \in V^*$ gilt:

$$x = v_1^*(x)\, v_1 + \cdots + v_n^*(x)\, v_n$$

$$x^* = x^*(v_1)\, v_1^* + \cdots + x^*(v_n)\, v_n^*.$$

(b) Speziell für $V = \mathbb{K}^n$ und die Standardbasis $B = \{e_1,...,e_n\}$ erhalten wir, daß das j–te Koordinatenfunktional e_j^* durch das n–Tupel

$$(e_j^*(e_1),...,e_j^*(e_n)) = (0,\ldots,0,\underbrace{1}_{j\text{–te Stelle}},0,...,0)$$

festgelegt ist und ein beliebiges $x^* \in (\mathbb{K}^n)^*$ wegen $x^* = \sum_{j=1}^{n} x_j\, e_j^*$ und $x^*(e_i) = x_i$ durch das n–Tupel $(x_1,...,x_n)$. Weil wir n–Tupel und Spaltenvektoren identifizieren, können wir e_j^* als Standardbasisvektor e_j auffassen und x^* als Spaltenvektor $x = (x_1 \cdots x_n)^T$. Damit wird $(\mathbb{K}^n)^*$ in natürlicher Weise mit $\mathbb{K}^n$ identifiziert und B^* mit B. Die Wirkung von x^* auf

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

wird dann durch

$$x^*(y) = x_1 y_1 + \cdots + x_n y_n = x^\top y =: \langle x,y \rangle$$

gegeben. Das so definierte Produkt $\langle \cdot,\cdot \rangle$ heißt das *Standardskalarprodukt* auf $\mathbb{K}^n$. Es wird in Kapitel 5 weiter behandelt.

(c) Unendlich dimensionale Vektorräume sind im allgemeinen nicht zu ihrem Dualraum isomorph:

Sei beispielsweise $V = \mathbb{R}[X]$, und sei $B = \{1, X, X^2, X^3, \ldots\}$ die kanonische Basis von V. Dann ist $x^* \in V^*$ eindeutig bestimmt durch die Folge $(x^*(1), x^*(X), x^*(X^2), \ldots) \in \mathbb{R}^{\mathbb{N}_0}$. Ist umgekehrt $(a_0, a_1, a_2, \ldots)$ eine Folge reeller Zahlen, so wird durch $x^*(X^i) = a_i$, $i = 0,1,2,\ldots$, eine Linearform auf V definiert. Die Abbildung

$$V^* \longrightarrow \mathbb{R}^{\mathbb{N}_0}, \quad x^* \longmapsto (x^*(1), x^*(X), x^*(X^2), \ldots)$$

ist dann ein Isomorphismus. Weil V und $\mathbb{R}^{\mathbb{N}_0}$ nicht isomorph sind, ist V damit auch nicht zu V^* isomorph.

Nun gehen wir einen Schritt weiter und betrachten den Dualraum von V^*. Statt $(V^*)^*$ schreiben wir V^{**} und nennen V^{**} den *Bidualraum* von V.

Es gibt einen natürlichen Zusammenhang zwischen dem Vektorraum V und seinem Bidualraum V^{**}. Wir erhalten ihn, indem wir $x^*(x)$, $x \in V$, $x^* \in V^*$, betrachten und bei festem x die Linearform $x^* \in V^*$ variieren lassen.

Bemerkung. Sei $x \in V$. Gilt $x^*(x) = 0$ für alle $x^* \in V^*$, so ist $x = o$.

Beweis. Ist $x \neq o$, so können wir x zu einer Basis B von V ergänzen und dann eine nichttriviale Linearform $x^* \in V^*$ definieren durch $x^*(x) = 1$ und $x^*(x') = 0$ für alle $x' \in B \setminus \{x\}$. ∎

Satz 11. *Es seien V ein $\mathbb{K}$-Vektorraum und V^{**} der Bidualraum. Dann ist die kanonische Abbildung*

$$\begin{aligned} \Psi : V &\longrightarrow V^{**} \\ x &\longmapsto x^{**}, \end{aligned}$$

wo $x^{**}(x^*) := x^*(x)$ *für alle* $x^* \in V^*$, *ein injektiver Homomorphismus. Ist* V *endlich dimensional, so ist* Ψ *ein Isomorphismus.*

Beweis. Ψ ist linear: Für alle $x, y \in V$ und alle $a,b \in \mathbb{K}$ ist $\Psi(a\,x + b\,y)(x^*) = x^*(a\,x + b\,y) = a\,x^*(x) + b\,x^*(y) = a\,\Psi(x)(x^*) + b\,\Psi(y)(x^*) = (a\,\Psi(x) + b\,\Psi(y))(x^*)$. Dies gilt für alle $x^* \in V^*$, also folgt $\Psi(a\,x + b\,y) = a\,\Psi(x) + b\,\Psi(y)$.

Ψ ist injektiv: Sei $x \in V$ und $\Psi(x)$ der Nullvektor in V^{**}. Nach Definition von Ψ erhalten wir $x^*(x) = \Psi(x)(x^*) = 0$ für alle $x^* \in V^*$. Nach der obigen Bemerkung folgt $x = o$, also ist Ψ injektiv.

Ist V endlich dimensional, so ist Ψ nach Korollar 6 dann auch bijektiv. ∎

Ist V endlich dimensional, so werden wir wegen des natürlichen Isomorphismus zwischen V und V^{**} in Zukunft V mit seinem Bidualraum V^{**} identifizieren.

Ist V unendlich dimensional, so ist Ψ eine Einbettung von V in V^{**}, und wir können deswegen V als Untervektorraum von V^{**} auffassen.

Zum Abschluß dieses Paragraphen wollen wir noch jeder linearen Abbildung $\Phi : V \longrightarrow W$ eine lineare Abbildung $\Phi^\mathsf{T} : W^* \longrightarrow V^*$ zuordnen.

Definition. Es seien V, W $\mathbb{K}$–Vektorräume und $\Phi : V \longrightarrow W$ linear. Dann heißt die durch

$$\begin{aligned} \Phi^\mathsf{T} : W^* &\longrightarrow V^* \\ y^* &\longmapsto y^* \circ \Phi \end{aligned}$$

erklärte lineare Abbildung Φ^T die *transponierte* oder *duale Abbildung* von Φ.

Bemerkung. Der Name transponierte Abbildung erklärt sich von selbst, wenn wir im nächsten Paragraphen den Zusammenhang zwischen linearen Abbildungen und Matrizen behandeln.

Für transponierte Abbildungen gelten folgende Rechenregeln.

Satz 12. *Es seien V, W, X $\mathbb{K}$-Vektorräume. Dann gilt:*

(a) $(\mathrm{id}_V)^\tau = \mathrm{id}_{V^*}$,

(b) $(\Phi + \Psi)^\tau = \Phi^\tau + \Psi^\tau$, $(a\,\Phi)^\tau = a\,\Phi^\tau$ *für alle* $\Phi, \Psi \in \mathrm{Hom}(V, W)$ *und alle* $a \in \mathbb{K}$,

(c) $(\Psi \circ \Phi)^\tau = \Phi^\tau \circ \Psi^\tau$ *für alle* $\Phi \in \mathrm{Hom}(V, W)$ *und alle* $\Psi \in \mathrm{Hom}(W, X)$,

(d) $\Phi \in \mathrm{Hom}(V, W)$ *ist genau dann Isomorphismus, wenn* $\Phi^\tau \in \mathrm{Hom}(W^*, V^*)$ *Isomorphismus ist. In diesem Fall gilt* $(\Phi^{-1})^\tau = (\Phi^\tau)^{-1}$.

Beweis. (a) $(\mathrm{id}_V)^\tau(x^*) = x^* \circ \mathrm{id}_V = x^*$ für alle $x^* \in V^*$.

(b) $(\Phi + \Psi)^\tau(y^*) = y^* \circ (\Phi + \Psi) = y^* \circ \Phi + y^* \circ \Psi = \Phi^\tau(y^*) + \Psi^\tau(y^*) = (\Phi^\tau + \Psi^\tau)(y^*)$ für alle $y^* \in W^*$. Analog folgt $(a\,\Phi)^\tau(y^*) = y^* \circ (a\,\Phi) = a\,(y^* \circ \Phi) = a\,(\Phi^\tau(y^*)) = (a\,\Phi^\tau)(y^*)$ für alle $y^* \in W^*$.

(c) Es ist $\Psi \circ \Phi \in \mathrm{Hom}(V, X)$ und $(\Psi \circ \Phi)^\tau \in \mathrm{Hom}(X^*, V^*)$. Dann gilt für alle $z^* \in X^*$: $(\Psi \circ \Phi)^\tau(z^*) = z^* \circ (\Psi \circ \Phi)$ und $(\Phi^\tau \circ \Psi^\tau)(z^*) = \Phi^\tau(\Psi^\tau(z^*)) = \Psi^\tau(z^*) \circ \Phi = (z^* \circ \Psi) \circ \Phi = z^* \circ (\Psi \circ \Phi)$.

(d) Ist $\Phi \in \mathrm{Hom}(V, W)$ bijektiv, so existiert Φ^{-1} und es gilt $\Phi^{-1} \circ \Phi = \mathrm{id}_V$, $\Phi \circ \Phi^{-1} = \mathrm{id}_W$. Daraus folgt $\Phi^\tau \circ (\Phi^{-1})^\tau = (\Phi^{-1} \circ \Phi)^\tau = (\mathrm{id}_V)^\tau = \mathrm{id}_{V^*}$ und $(\Phi^{-1})^\tau \circ \Phi^\tau = (\Phi \circ \Phi^{-1})^\tau = (\mathrm{id}_W)^\tau = \mathrm{id}_{W^*}$. Also ist auch Φ^τ bijektiv, und es gilt $(\Phi^\tau)^{-1} = (\Phi^{-1})^\tau$.

Die Umkehrung ergibt sich aus folgenden zwei Behauptungen:

1. Beh.: Ist $\Phi^\tau \in \mathrm{Hom}(W^*, V^*)$ surjektiv, so ist $\Phi \in \mathrm{Hom}(V, W)$ injektiv.

Beweis. Sei $\Phi(x) = o$. Weil Φ^τ surjektiv ist, gibt es zu jedem $x^* \in V^*$ ein $y^* \in W^*$ mit $\Phi^\tau(y^*) = x^*$. Damit folgt $x^*(x) = \Phi^\tau(y^*)(x) = y^*(\Phi(x)) = 0$ für alle $x^* \in V^*$. Nach der obigen Bemerkung ist dann $x = o$, also ist Φ injektiv.

2. Beh.: Ist $\Phi^\tau \in \mathrm{Hom}(W^*, V^*)$ injektiv, so ist $\Phi \in \mathrm{Hom}(V, W)$ surjektiv.

Beweis. Wir führen einen Widerspruchsbeweis und nehmen an, daß Φ nicht surjektiv sei. Somit gilt $\Phi(V) \neq W$. Wir ergänzen nun eine Basis C von $\Phi(V)$ zu einer Basis C' von W, wählen ein festes $y \in C' \backslash C$ und definieren eine Linearform $y^* \in W^*$ durch $y^*(y) = 1$ und $y^*(y') = 0$ für $y' \in C' \backslash \{y\}$. Dann gilt $\Phi^\tau(y^*)(x) = y^*(\Phi(x)) = 0$ für alle $x \in V$, also $\Phi^\tau(y^*) = o^*$. Wegen der Injektivität von Φ^τ folgt hieraus, daß y^* die Nullform ist, ein Widerspruch zur Definition von y^*. Somit ist Φ surjektiv. ∎

§ 3 Darstellung linearer Abbildungen durch Matrizen

Ein n–dimensionaler $\mathbb{K}$–Vektorraum V ist nach Satz 7 isomorph zum $\mathbb{K}^n$, und wenn wir eine Basis $\{v_1,...,v_n\}$ von V gegeben haben, können wir einen Isomorphismus konkret angeben. Da es im folgenden auf die Reihenfolge der Basisvektoren ankommt, schreiben wir

$$B = (v_1,...,v_n)$$

und nennen dieses n–Tupel von Vektoren eine *geordnete Basis* von V.

Bemerkung. Bei einer geordneten Basis $(v_1,...,v_n)$ sind die Vektoren $v_1,...,v_n$ paarweise verschieden.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $B = (v_1,...,v_n)$ eine geordnete Basis von V. Sei $x \in V$ und

$$x = x_1 v_1 + \cdots + x_n v_n \,,\; x_i \in \mathbb{K}$$

die eindeutige Darstellung von x bezüglich B. Dann heißen die Zahlen $x_1,...,x_n$ die *Koordinaten* von x bezüglich B und der Vektor

$$\hat{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{K}^n$$

heißt *Koordinatenvektor* oder *Koordinatendarstellung* von x bezüglich B.

Bemerkung. Die Abbildung $V \longrightarrow \mathbb{K}^n$, die jedem $x \in V$ den zugehörigen Koordinatenvektor $\hat{x}$ bezüglich B zuordnet, ist linear und bildet insbesondere v_i auf den Standardbasisvektor $e_i \in \mathbb{K}^n$ ab, $i = 1,...,n$. Sie ist daher ein Isomorphismus.

Nun können wir einen Zusammenhang zwischen linearen Abbildungen endlich dimensionaler Vektorräume und Matrizen herstellen.

Definition. Es seien V, W $\mathbb{K}$–Vektorräume, $B = (v_1,...,v_n)$ bzw. $C = (w_1,...,w_m)$

geordnete Basen von V bzw. W und $\Phi : V \longrightarrow W$ linear. Für $j \in \{1,...,n\}$ sei

$$\Phi(v_j) = a_{1j} w_1 + \cdots + a_{mj} w_m$$

die eindeutige Darstellung von $\Phi(x_j)$ bezüglich C. Dann heißt die (m,n)–Matrix

$$A_\Phi := (\!(a_{ij})\!)$$

Abbildungsmatrix von Φ (bezüglich der geordneten Basen B von V und C von W).

Bemerkungen **(a)** Die Koordinaten von $\Phi(v_j)$ bilden gerade die j–te Spalte von A_Φ.

(b) Es ist Rg A_Φ = dim Bild Φ. Aus diesem Grunde verwenden wir statt dim Bild Φ auch die Bezeichnung Rg Φ und sprechen vom *Rang* der Abbildung Φ.

Satz 13. *Es seien V, W $\mathbb{K}$–Vektorräume*, dim $V = n$, dim $W = m$ *und* $B = (v_1,...,v_n)$ *bzw.* $C = (w_1,...,w_m)$ *geordnete Basen von V bzw. W. Dann ist die Abbildung*

$$\Phi \longmapsto A_\Phi ,$$

die jedem $\Phi \in$ Hom (V,W) die Abbildungsmatrix A_Φ bezüglich der Basen B,C zuordnet, ein Isomorphismus von Hom (V,W) *auf* $\mathbb{K}^{m\times n}$.

Beweis. Ist $A_\Phi = (\!(a_{ij})\!)$, $A_{\Phi'} = (\!(b_{ij})\!)$, so gilt für alle $a,a' \in \mathbb{K}$ und $j = 1,...,n$

$$(a\,\Phi + a'\,\Phi')(v_j) = a\,\Phi(v_j) + a'\,\Phi'(v_j)$$

$$= a \sum_{i=1}^{m} a_{ij} w_i + a' \sum_{i=1}^{m} b_{ij} w_i = \sum_{i=1}^{m} (a\,a_{ij} + a'\,b_{ij})\, w_i .$$

Also folgt $A_{a\,\Phi + a'\Phi'} = a\,A_\Phi + a'A_{\Phi'}$, und die Abbildung $\Phi \longmapsto A_\Phi$ ist daher linear.

Für jedes $A = (\!(a_{ij})\!) \in \mathbb{K}^{m\times n}$ wird durch

$$\Phi(v_j) = \sum_{i=1}^{m} a_{ij} w_i , \; j = 1,...,n,$$

eine lineare Abbildung $\Phi : V \longrightarrow W$ eindeutig festgelegt (Satz 4), die $A_\Phi = A$ erfüllt. Also ist die Abbildung $\Phi \longmapsto A_\Phi$ ein Isomorphismus. ∎

Bemerkung. Aus diesem Satz folgt wieder das schon bekannte Ergebnis über die Dimension von $\mathrm{Hom}(V,W)$: $\dim \mathrm{Hom}\,(V,W) = \dim V \cdot \dim W$.

Die Wirkung der Abbildungsmatrix wird durch den folgenden Satz deutlich.

Satz 14. *Es seien V, W endlich dimensionale $\mathbb{K}$-Vektorräume mit geordneten Basen B bzw. C, $\Phi : V \longrightarrow W$ linear und A_Φ die Abbildungsmatrix von Φ. Weiterhin sei $\hat{x}$ der Koordinatenvektor von $x \in V$ und $\hat{y}$ der Koordinatenvektor von $y = \Phi(x)$. Dann gilt*

$$\hat{y} = A_\Phi \cdot \hat{x}.$$

Beweis. Seien $B = (v_1,\ldots,v_n)$, $C = (w_1,\ldots,w_m)$ und $A_\Phi = (\!(a_{ij})\!)$. Für $x = x_1 v_1 + \cdots + x_n v_n$ folgt $\Phi(x) = y_1 w_1 + \cdots + y_m w_m$ sowie

$$\Phi(x) = \sum_{j=1}^{n} x_j \Phi(v_j) = \sum_{j=1}^{n} x_j \Big(\sum_{i=1}^{m} a_{ij} w_i \Big) = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} x_j \Big) w_i .$$

Also gilt

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

für $i = 1,\ldots,m$, d.h.

$$\hat{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = A_\Phi \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A_\Phi \, \hat{x} . \qquad ∎$$

Bemerkung. Im Fall $V = \mathbb{K}^n$, $W = \mathbb{K}^m$ gilt bezüglich der Standardbasen

$$A_\Phi = (\Phi(e_1) \,|\cdots|\, \Phi(e_n)) \quad \text{und} \quad \Phi(x) = A_\Phi \cdot x \ , \ x \in \mathbb{K}^n.$$

Beispiele. **(a)** Es seien V, W reelle Vektorräume , (v_1, v_2, v_3) eine geordnete Basis von V sowie (w_1, w_2, w_3, w_4) eine geordnete Basis von W. Weiter sei $\Phi \in \mathrm{Hom}\,(V, W)$ festgelegt durch

$$\begin{aligned} \Phi(v_1) &= w_1 - w_2 + 3\,w_3 - w_4 \\ \Phi(v_2) &= 2\,w_1 + w_2 + 7\,w_3 + 2\,w_4 \\ \Phi(v_3) &= 3\,w_2 + w_3 + 4\,w_4\,. \end{aligned}$$

Die Abbildungsmatrix von Φ bezüglich der gegebenen Basen ist

$$A_\Phi = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}.$$

Wir wollen Kern Φ berechnen. Wegen Kern $\Phi = \{x \in V \mid A_\Phi\, \hat{x} = o\}$ müssen wir das homogene lineare Gleichungssystem $A_\Phi\, \hat{x} = o$ lösen. Es gilt

$$L_h = \{a \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \mid a \in \mathbb{R}\}\,,$$

also folgt

$$\text{Kern } \Phi = \{a\,(2\,v_1 - v_2 + v_3) \mid a \in \mathbb{R}\}.$$

(b) Sei $V \subset \mathbb{R}[X]$ der Untervektorraum der Polynome vom Grad $\leq n-1$, $n \in \mathbb{N}$. Dann gilt dim $V = n$. Der Ableitungsoperator $D : V \longrightarrow V$ hat bezüglich der Standardbasis $(1, X, \ldots, X^{n-1})$ von V die Abbildungsmatrix

$$A_D = \begin{bmatrix} 0 & 1 & 0 & \cdots\cdots & 0 \\ \vdots & \ddots & 2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ \vdots & & & \ddots & n-1 \\ 0 & \cdots\cdots\cdots\cdots & & & 0 \end{bmatrix}$$

Wir wollen nun zeigen, wie sich das Verketten linearer Abbildungen, die Inver-

senbildung sowie der Übergang zur transponierten Abbildung bei den zugehörigen Abbildungsmatrizen auswirkt.

Satz 15. *Es seien V,W,X $\mathbb{K}$-Vektorräume mit* dim $V = n$, dim $W = m$, dim $X = k$ *und mit geordneten Basen B_1,B_2,B_3. Weiter seien $\Phi : V \longrightarrow W$ und $\Phi' : W \longrightarrow X$ linear. Dann gilt:*

$$A_{\Phi' \circ \Phi} = A_{\Phi'} \cdot A_{\Phi} .$$

Beweis. Es seien $B_1 = (v_1,...,v_n)$, $B_2 = (w_1,...,w_m)$, $B_3 = (x_1,...,x_k)$ sowie $A_{\Phi} = (\!(a_{ij})\!)$, $A_{\Phi'} = (\!(b_{ij})\!)$ und $A_{\Phi' \circ \Phi} = (\!(c_{ij})\!)$. Dann gilt

$$\Phi' \circ \Phi\,(v_j) = \Phi'(\Phi(v_j)) = \Phi'(\sum_{i=1}^{m} a_{ij}\, w_i) = \sum_{i=1}^{m} a_{ij}\, \Phi'(w_i)$$

$$= \sum_{i=1}^{m} a_{ij} (\sum_{l=1}^{k} b_{li}\, x_l) = \sum_{l=1}^{k} (\sum_{i=1}^{m} b_{li}\, a_{ij})\, x_l .$$

Damit erhalten wir

$$c_{lj} = \sum_{i=1}^{m} b_{li}\, a_{ij} ,$$

also gilt die Behauptung. ∎

Satz 16. *Es seien V,W n-dimensionale $\mathbb{K}$-Vektorräume mit geordneten Basen B,C und $\Phi : V \longrightarrow W$ eine lineare Abbildung. Φ ist genau dann ein Isomorphismus, wenn die Abbildungsmatrix A_{Φ} von Φ bezüglich B,C regulär ist. In diesem Fall gilt für die Abbildungsmatrix von Φ^{-1} bezüglich C,B*

$$A_{\Phi^{-1}} = A_{\Phi}^{-1} .$$

Beweis. Ist Φ ein Isomorphismus, so folgt $\Phi^{-1} \circ \Phi = \mathrm{id}_V$, $\Phi \circ \Phi^{-1} = \mathrm{id}_W$, also nach Satz 15

$$A_\Phi{}^{-1} \cdot A_\Phi = A_{\mathrm{id}_V} = E_n ,$$

$$A_\Phi \cdot A_\Phi{}^{-1} = A_{\mathrm{id}_W} = E_n .$$

Somit gilt $A_\Phi{}^{-1} = A_\Phi^{-1}$. Nun sei umgekehrt A_Φ regulär. Wir betrachten die Abbildung $\Phi' : W \longrightarrow V$, die nach Satz 13 zu A_Φ^{-1} gehört. Dann folgt wieder nach Satz 15

$$E_n = A_\Phi \cdot A_\Phi^{-1} = A_{\Phi \circ \Phi'} ,$$

$$E_n = A_\Phi^{-1} \cdot A_\Phi = A_{\Phi' \circ \Phi} .$$

Nach Satz 13 ist folglich

$$\Phi \circ \Phi' = \mathrm{id}_W , \quad \Phi' \circ \Phi = \mathrm{id}_V ,$$

also $\Phi' = \Phi^{-1}$. ∎

Satz 17. *Es seien V, W $\mathbb{K}$-Vektorräume mit* dim $V = n$, dim $W = m$ *und mit geordneten Basen $B = (v_1,\ldots,v_n)$, $C = (w_1,\ldots,w_m)$. Weiter seien $\Phi : V \longrightarrow W$ eine lineare Abbildung und A_Φ die Abbildungsmatrix von Φ bezüglich B,C. Ist $A_\Phi{}^\top$ die Abbildungsmatrix von $\Phi^\top$ bezüglich der dualen Basen C^* und B^* , so gilt*

$$A_\Phi{}^\top = A_\Phi^\top .$$

Beweis. Es gilt für $A_\Phi{}^\top = (\!(c_{kl})\!)$ nach Definition der Abbildungsmatrix

$$\Phi^\top(w_l^*) = \sum_{j=1}^{n} c_{jl}\, v_j^* , \text{ also } (\Phi^\top(w_l^*))(v_k) = \sum_{j=1}^{n} c_{jl}\, v_j^*(v_k) = \sum_{j=1}^{n} c_{jl}\, \delta_{jk} = c_{kl} .$$

Andererseits ist nach Definition $\Phi^\top(w_l^*) = w_l^* \circ \Phi$, also gilt mit $A_\Phi = (\!(a_{ij})\!)$

$$(\Phi^\top(w_l^*))(v_k) = w_l^*(\Phi(v_k)) = w_l^*\Big(\sum_{i=1}^{m} a_{ik}\, w_i\Big) = \sum_{i=1}^{m} a_{ik}\, w_l^*(w_i) = a_{lk} .$$

Damit folgt $c_{kl} = a_{lk}$ für alle k,l . ∎

Zum Schluß dieses Abschnitts wollen wir uns noch überlegen, wie sich die Abbildungsmatrix einer linearen Abbildung $\Phi : V \longrightarrow W$ ändert, wenn man in den Vektorräumen V und W zu anderen Basen übergeht. Es seien

$$B = (v_1,...,v_n) \quad , \quad \tilde{B} = (\tilde{v}_1,...,\tilde{v}_n)$$

geordnete Basen von V und

$$C = (w_1,...,w_m) \quad , \quad \tilde{C} = (\tilde{w}_1,...,\tilde{w}_m)$$

geordnete Basen von W. Weiterhin sei $A_\Phi = (\!(a_{ij})\!) \in \mathbb{K}^{m\times n}$ die Abbildungsmatrix von $\Phi \in \mathrm{Hom}\,(V,W)$ bezüglich der Basen B,C und $\tilde{A}_\Phi = (\!(\tilde{a}_{ij})\!)$ sei die Abbildungsmatrix von Φ bezüglich der neuen Basen $\tilde{B},\tilde{C}$.

Die neuen Basisvektoren lassen sich mit Hilfe der alten linear kombinieren:

$$\tilde{v}_j = s_{1j}\, v_1 + \cdots + s_{nj}\, v_n \quad , \quad j = 1,...,n\,,$$

$$\tilde{w}_k = t_{1k}\, w_1 + \cdots + t_{mk}\, w_m \quad , \quad k = 1,...,m\,.$$

Die Matrix $S = (\!(s_{ij})\!) \in \mathbb{K}^{n\times n}$ ist regulär, denn sie ist die Abbildungsmatrix von id_V bezüglich der Basen $\tilde{B},B$. Ebenso ist auch die Matrix $T = (\!(t_{ij})\!) \in \mathbb{K}^{m\times m}$ regulär als Abbildungsmatrix von id_W bezüglich der Basen $\tilde{C},C$.

Nun gilt für alle $j = 1,...,n$:

$$\Phi(\tilde{v}_j) = \sum_{k=1}^{m} \tilde{a}_{kj}\, \tilde{w}_k = \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{i=1}^{m} t_{ik}\, w_i = \sum_{i=1}^{m} \Big(\sum_{k=1}^{m} t_{ik}\, \tilde{a}_{kj} \Big)\, w_i$$

bzw.

$$\Phi(\tilde{v}_j) = \sum_{k=1}^{n} s_{kj}\, \Phi(v_k) = \sum_{k=1}^{n} s_{kj} \sum_{i=1}^{m} a_{ik}\, w_i = \sum_{i=1}^{m} \Big(\sum_{k=1}^{n} a_{ik}\, s_{kj} \Big)\, w_i\,.$$

Daraus folgt für alle $j = 1,...,n$ und alle $i = 1,...,m$:

$$\sum_{k=1}^{m} t_{ik}\, \tilde{a}_{kj} = \sum_{k=1}^{n} a_{ik}\, s_{kj}$$

d.h.

$$T\,\tilde{A}_{\Phi} \;=\; A_{\Phi}\,S$$

und somit

$$\tilde{A}_{\Phi} \;=\; T^{-1}\,A_{\Phi}\,S\,.$$

Die Art und Weise wie A_{Φ} und $\tilde{A}_{\Phi}$ zusammenhängen, führt zu folgender Definition.

Definition. Zwei Matrizen $A,\tilde{A} \in \mathbb{K}^{m\times n}$ heißen *äquivalent*, wenn es reguläre Matrizen $S \in \mathbb{K}^{n\times n}$ und $T \in \mathbb{K}^{m\times m}$ gibt mit $\tilde{A} \;=\; T^{-1}A\,S$.

Zusammenfassung. Durch Basiswechsel in V (B wird ersetzt durch $\tilde{B}$) und in W (C wird ersetzt durch $\tilde{C}$) geht die Abbildungsmatrix A_{Φ} einer linearen Abbildung $\Phi : V \longrightarrow W$ in eine äquivalente Matrix $\tilde{A}_{\Phi}$ über,

$$\tilde{A}_{\Phi} \;=\; T^{-1}\,A_{\Phi}\,S\,,$$

wobei die regulären Matrizen $S = (\!(s_{ik})\!)$ bzw. $T = (\!(t_{ik})\!)$ den Basiswechsel in V bzw. in W beschreiben und durch

$$\tilde{v}_j \;=\; \sum_{i\,=\,1}^{n} s_{ij}\,v_i\;,$$

$j = 1,\ldots,n$, bzw.

$$\tilde{w}_k \;=\; \sum_{i\,=\,1}^{m} t_{ik}\,w_i\;,$$

$k = 1,\ldots,m$, gegeben sind.

Beispiel. Die lineare Abbildung $\Phi : \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ sei bezüglich der Standardbasen durch die Matrix

$$A_{\Phi} = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}$$

beschrieben. Als neue Basen betrachten wir

$$\tilde{B} = (\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}) \text{ in } \mathbb{R}^3 \text{ sowie } \tilde{C} = (\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}) \text{ in } \mathbb{R}^4 .$$

Dann ist

$$S = \begin{bmatrix}1 & 0 & 1\\1 & 1 & 0\\0 & 1 & 1\end{bmatrix}, \quad T = \begin{bmatrix}1 & 1 & 0 & 1\\1 & 0 & 1 & 0\\0 & 1 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}$$

und

$$\tilde{A}_\Phi = T^{-1} A_\Phi S = \begin{bmatrix}1/2 & 1/2 & -1/2 & -1/2\\1/2 & -1/2 & 1/2 & -1/2\\-1/2 & 1/2 & 1/2 & 1/2\\0 & 0 & 0 & 1\end{bmatrix} \begin{bmatrix}3 & 2 & 1\\0 & 4 & 2\\10 & 8 & 4\\1 & 6 & 3\end{bmatrix} = \begin{bmatrix}-4 & -4 & -2\\6 & 0 & 0\\4 & 8 & 4\\1 & 6 & 3\end{bmatrix}$$

Bemerkungen. **(a)** Die Äquivalenz von Matrizen ist eine Äquivalenzrelation auf der Menge $\mathbb{K}^{m\times n}$.

(b) Durch elementare Zeilen– oder Spaltenumformungen geht eine Matrix $A \in \mathbb{K}^{m\times n}$ in eine äquivalente Matrix A' über.

Beweis. A ist die Abbildungsmatrix der linearen Abbildung $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^m$, $x \longmapsto A\,x$, bezüglich der Standardbasen. Elementare Zeilenumformungen entsprechen einem Basiswechsel in $\mathbb{K}^m$, elementare Spaltenumformungen einem Basiswechsel in $\mathbb{K}^n$. Daraus folgt die Behauptung. ∎

(c) Jede Matrix $A \in \mathbb{K}^{m\times n}$ ist zu ihrer Gaußschen Normalform äquivalent.

Beweis. Die Aussage folgt direkt aus (b). ∎

(d) Zwei Matrizen $A,B \in \mathbb{K}^{m\times n}$ sind genau dann äquivalent, wenn sie gleichen Rang haben.

Beweis. Durch elementare Zeilen– **und** Spaltenumformungen kann jede Matrix A auf die Form

$$A' = \begin{bmatrix} E_r & O \\ O & O \end{bmatrix}$$

mit $r = \text{Rg } A$ gebracht werden. Nach (b) sind A und A' äquivalent. Gilt also $\text{Rg } A = \text{Rg } B$, so sind auch A und B äquivalent. Gilt umgekehrt $B = T^{-1} A S$ mit regulären Matrizen $S \in \mathbb{K}^{n\times n}$ und $T \in \mathbb{K}^{m\times m}$, so ist A Abbildungsmatrix von $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^m$, $x \longmapsto A x$, bezüglich der Standardbasen in $\mathbb{K}^n$ und $\mathbb{K}^m$ und B ist die Abbildungsmatrix von Φ bezüglich der geordneten Basen $(S e_1,\ldots,S e_n)$ in $\mathbb{K}^n$ und $(T e_1,\ldots,T e_m)$ in $\mathbb{K}^m$. Damit folgt $\text{Rg } A = \dim \text{Bild } \Phi = \text{Rg } B$. ∎

Zum Schluß betrachten wir noch den Fall, daß eine lineare Abbildung $\Phi : V \longrightarrow V$ bezüglich einer Basis B in V dargestellt ist. Wird die Basis B gewechselt, so geht A_Φ über in $\tilde{A}_\Phi = S^{-1} A_\Phi S$ (wegen $T = S$). Auch für diese Beziehung verwenden wir einen eigenen Namen:

Definition. Zwei Matrizen $A,\tilde{A} \in \mathbb{K}^{n\times n}$ heißen *ähnlich*, wenn es eine reguläre Matrix $S \in \mathbb{K}^{n\times n}$ gibt mit

$$\tilde{A} = S^{-1} A S.$$

Bemerkungen. **(a)** Die Ähnlichkeit von Matrizen ist ein Äquivalenzrelation auf der Menge $\mathbb{K}^{n\times n}$.

(b) Ähnliche Matrizen sind äquivalent. Umgekehrt gibt es aber äquivalente quadratische Matrizen, die nicht ähnlich sind. In jeder Klasse äquivalenter Matrizen in $\mathbb{K}^{n\times n}$ gibt es nämlich nach der obigen Bemerkung (d) eine Matrix, die Diagonalgestalt hat, aber es gibt nicht zu jeder (n,n)–Matrix eine ähnliche Diagonalmatrix. Wann dies der Fall ist, werden wir im nächsten Kapitel sehen.

§ 4 Affine Abbildungen eines Vektorraumes

In diesem Abschnitt wollen wir zunächst wieder die affine Struktur des Vektorraumes V in den Vordergrund stellen und die zugehörigen strukturverträglichen Abbildungen studieren. Es sind dies die affinen Abbildungen von V. In Kapitel 6.1 werden wir dann die gewonnenen Ergebnisse auf allgemeine affine Räume übertragen.

Der affine Unterraum $L = x_0 + U$ geht aus seinem Richtungsraum durch die Abbildung $\tau : x \longmapsto x + x_0$ hervor. Diese Abbildung von V in sich ist nur für $x_0 = o$ linear; τ ist eine bijektive affine Abbildung von V in sich. Sie heißt *Translation* mit *Translationsvektor* x_0.

Definition. Es seien V, W $\mathbb{K}$–Vektorräume, $\Phi \in \mathrm{Hom}\,(V,W)$ und $w \in W$. Dann heißt die Abbildung

$$\begin{aligned} \varphi : V &\longrightarrow W \\ x &\longmapsto \Phi(x) + w \end{aligned}$$

eine *affine Abbildung* von V in W. Der Vektor w heißt *Translationsvektor* von φ.

Jede bijektive affine Abbildung $\varphi : V \longrightarrow V$ heißt eine *Affinität* von V.

Bemerkungen. **(a)** Jede affine Abbildung $\varphi : V \longrightarrow W$ ist also Komposition einer linearen Abbildung $\Phi : V \longrightarrow W$ mit einer Translation τ in W, $\varphi = \tau \circ \Phi$. Die Abbildungen Φ und τ sind dabei eindeutig bestimmt.

φ ist genau dann bijektiv, wenn es die zugehörige lineare Abbildung Φ ist. In diesem Fall gilt $\varphi^{-1}(y) = \Phi^{-1}(y) - \Phi^{-1}(w)$ für alle $y \in W$, und die Umkehrfunktion $\varphi^{-1} : W \longrightarrow V$ ist daher ebenfalls affin.

(b) Die Komposition affiner Abbildungen $\varphi : V \longrightarrow W$, $\varphi' : W \longrightarrow X$ ist affin.

Beweis. Seien $\varphi(v) = \Phi(v) + w_0$, $\Phi \in \mathrm{Hom}\,(V,W)$, $w_0 \in W$ und $\varphi'(w) = \Phi'(w) + x_0$, $\Phi' \in \mathrm{Hom}\,(W,X)$, $x_0 \in X$. Dann gilt $\varphi' \circ \varphi\,(v) = \Phi' \circ \Phi\,(v) + \Phi'(w_0) + x_0$ mit $\Phi' \circ \Phi \in \mathrm{Hom}\,(V,X)$ und $\Phi'(w_0) + x_0 \in X$. ∎

(c) Die Affinitäten von V bilden bezüglich der Komposition eine Gruppe, die *affine Gruppe* von V. Wir bezeichnen sie mit $\mathrm{Aff}(V)$.

Affine Abbildungen lassen die geometrische Struktur invariant. Genauer bedeutet dies:

Satz 18. *Es seien V, W endlich dimensionale $\mathbb{K}$-Vektorräume und $\varphi : V \longrightarrow W$ eine affine Abbildung. Dann gilt:*

(a) *Ist L ein affiner Unterraum von V, so ist $\varphi(L)$ ein affiner Unterraum von W und es ist* $\dim \varphi(L) \leq \dim L$.

(b) *Ist M ein affiner Unterraum von W, so ist $\varphi^{-1}(M)$ entweder die leere Menge oder ein affiner Unterraum von V.*

(c) *φ erhält die Parallelität, d.h. für alle affinen Unterräume L_1, L_2 von V folgt aus $L_1 \parallel L_2$ stets $\varphi(L_1) \parallel \varphi(L_2)$.*

Beweis. Es sei $\varphi(x) = \Phi(x) + w$ mit $\Phi \in \mathrm{Hom}\,(V, W)$ und $w \in W$.

(a) Mit $L = x_0 + U$ erhalten wir $\varphi(L) = \Phi(U) + \Phi(x_0) + w$, also ist $\varphi(L)$ ein affiner Unterraum von W, und es gilt $\dim \varphi(L) = \dim \Phi(U) \leq \dim U = \dim L$.

(b) Sei $M = y' + U'$ mit $U' \subset W$ und $y' \in W$. Ist $\varphi^{-1}(M) \neq \emptyset$, so gibt es ein $x_0 \in V$ mit $\varphi(x_0) \in M$. Damit folgt $\Phi(x_0) \in y' - w + U'$ und nach der Bemerkung von S.126 $\Phi(x_0) + U' = y' - w + U'$. Es gilt dann:

$$x \in \varphi^{-1}(M) \iff \varphi(x) \in M \iff \Phi(x) \in y' - w + U' = \Phi(x_0) + U'$$

$$\iff \Phi(x - x_0) \in U' \iff x \in x_0 + \Phi^{-1}(U'),$$

d.h. $\varphi^{-1}(M)$ ist ein affiner Unterraum von V.

(c) Es seien $L_1 = x_1 + U_1$, $L_2 = x_2 + U_2$ und o.B.d.A. sei $U_1 \subset U_2$. Dann folgt $\Phi(U_1) \subset \Phi(U_2)$ und somit $\varphi(L_1) \parallel \varphi(L_2)$. ∎

Nun wenden wir uns den affinen Selbstabbildungen von V zu. Die einfachsten Beispiele neben der Identität sind die Translationen.

Weitere Beispiele sind die *Streckungen* mit *Zentrum* z und *Streckungsfaktor* c. Sie bilden z auf sich ab und ordnen jedem von z verschiedenen Punkt x denjenigen

Punkt y auf der Geraden durch x und z zu, für den $y - z = c\,(x - z)$ gilt. Somit hat eine solche Streckung $\sigma : V \longrightarrow V$ die Darstellung $\sigma(x) = y = c\,x + (1 - c)\,z$, ist also tatsächlich eine affine Abbildung.

Für $c = 0$ bildet σ alle Punkte auf z ab, für $c = 1$ ist σ die Identität. Für char $\mathbb{K} \neq 2$ und $c = -1$ schließlich gilt $z = \frac{1}{2}(x + \sigma(x))$, d.h. z ist der *Mittelpunkt* der Punkte x und $\sigma(x)$. In diesem Fall heißt σ *Punktspiegelung* an z.

Translationen und Streckungen mit Streckungsfaktor $c \neq 0$ sind Affinitäten.

Fixelemente affiner Selbstabbildungen

Für die Klassifikation der affinen Selbstabbildungen von V nach geometrischen Gesichtspunkten sind die Fixelemente (Fixpunkte, Fixräume, Fixrichtungen) einer solchen Abbildung wichtig.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $\varphi : V \longrightarrow V$ eine affine Abbildung. Jeder Punkt x mit $\varphi(x) = x$ heißt *Fixpunkt* von φ. Allgemein heißt ein affiner Unterraum $L \subset V$ der Dimension k ein *k–dimensionaler Fixraum* von φ, falls $\varphi(L) = L$ gilt.

Bemerkungen. **(a)** Fixräume bleiben unter φ im allgemeinen nicht punktweise fest.

(b) $L = x_0 + U$ ist genau dann Fixraum der affinen Abbildung $\varphi : x \longmapsto \Phi(x) + v$, $x \in V$, wenn $\Phi(U) = U$ und $\varphi(x_0) \in L$ gilt.

(c) x ist genau dann Fixpunkt von φ, wenn x die Gleichung $(\Phi - \mathrm{id}_V)(x) = -v$ erfüllt. Die Menge aller Fixpunkte von φ ist also entweder leer oder ein affiner Unterraum L von V mit Richtungsraum $\mathrm{Kern}(\Phi - \mathrm{id}_V)$; L heißt dann *Fixpunktraum* von φ.

(d) Ist V endlich dimensional, so besitzt φ genau dann genau einen Fixpunkt, wenn $\Phi - \mathrm{id}_V$ injektiv ist. In diesem Fall enthält jeder Fixraum L von φ diesen Fixpunkt:

Sei $L = x_0 + U$. Aus $\Phi(U) = U$ folgt zunächst $(\Phi - \mathrm{id}_V)(U) \subset U$. Weil $\Phi - \mathrm{id}_V$ bijektiv ist, gilt sogar Gleichheit, also auch $(\Phi - \mathrm{id}_V)^{-1}(U) = U$. Aus $\varphi(x_0) \in L$ folgt $(\Phi - \mathrm{id}_V)(x_0) = -v + u$, $u \in U$, und aus $\varphi(x) = x$ folgt $(\Phi - \mathrm{id}_V)(x) = -v$. Durch Subtraktion erhalten wir $(\Phi - \mathrm{id}_V)(x - x_0) = -u \in U$, also $x \in x_0 + (\Phi - \mathrm{id}_V)^{-1}(U) = L$.

Beispiele. **(a)** Translationen, die von der Identität verschieden sind, besitzen keine Fixpunkte. Die k–dimensionalen Fixräume, $k \geq 1$, sind genau die k–dimensionalen affinen Unterräume, die zur Translationsrichtung parallel sind.

(b) Die Streckungen mit Streckungsfaktor $c \neq 0$ und $c \neq 1$ besitzen das Zentrum als einzigen Fixpunkt, und genau die affinen Unterräume, die das Zentrum enthalten, sind Fixräume.

(c) Sei $\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, $\varphi(x) = A\,x + v$ mit $A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$, $v = \begin{bmatrix} 6 \\ -4 \end{bmatrix}$.

Wir bestimmen die Fixpunkte: $\varphi(x) = x \iff (A - E_2)\,x = -v$. Die einzige Lösung dieses LGS ist $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$ ist. Somit ist $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$ einziger Fixpunkt von φ.

Um die Fixgeraden von φ zu erhalten, müssen wir nach der obigen Bemerkung (b) diejenigen eindimensionalen Unterräume $[x]$ berechnen, für die $A\,([x]) = [x]$ gilt. Dies führt uns auf das Problem, alle $c \in \mathbb{R}$, $c \neq 0$, und alle $x \in \mathbb{R}^2$, $x \neq o$, zu bestimmen, für die $A\,x = c\,x$ gilt. Jedes solche $x \neq o$ heißt ein *Eigenvektor* zum *Eigenwert* c. Nun gilt

$$A\,x = c\,x,\ x \neq o,\ \text{ist lösbar} \iff (A - c\,E_2)\,x = o \ \text{ist nichttrivial lösbar}$$

$$\iff \operatorname{Rg}(A - c\,E_2) < 2 \iff c = \pm 2.$$

Für $c = 2$ erhalten wir als Fixrichtung $[\begin{bmatrix} 3 \\ 1 \end{bmatrix}]$, für $c = -2$ erhalten wir als Fixrichtung $[\begin{bmatrix} -1 \\ 1 \end{bmatrix}]$. Da in diesem Fall jede Fixgerade den einzigen Fixpunkt enthalten muß, sind

$$\begin{bmatrix} 0 \\ -2 \end{bmatrix} + [\begin{bmatrix} 3 \\ 1 \end{bmatrix}] \quad \text{und} \quad \begin{bmatrix} 0 \\ -2 \end{bmatrix} + [\begin{bmatrix} -1 \\ 1 \end{bmatrix}]$$

die beiden einzigen Fixgeraden von φ.

(d) Sei $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, $\varphi(x) = A\,x + v$ mit $A = \begin{bmatrix} 7 & -6 & 11 \\ 0 & 1 & -1 \\ -4 & 4 & -7 \end{bmatrix}$, $v = \begin{bmatrix} 4 \\ -2 \\ -4 \end{bmatrix}$.

Zur Bestimmung der Fixpunkte lösen wir das LGS $(A - E_3)\,x = -v$ mit der erwei-

terten Matrix

$$\left[\begin{array}{rrr|r} 6 & -6 & 11 & -4 \\ 0 & 0 & -1 & 2 \\ -4 & 4 & -8 & 4 \end{array}\right]$$

Es hat als Lösungsmenge die Gerade

$$h = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + \left[\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right],$$

d.h. φ besitzt eine Fixpunktgerade.

Die Bestimmung der Fixgeraden $g = x_0 + [x]$ erfolgt in zwei Schritten.

1.Schritt: Wir berechnen wieder alle eindimensionalen Fixrichtungen $[x]$, d.h. alle Eigenvektoren zu dem Eigenwert $c \neq 0$:

$$(A - c\,E_3)\,x = 0 \text{ ist nichttrivial lösbar} \iff \operatorname{Rg}(A - c\,E_3) < 3 \iff c = \pm 1.$$

Für $c = -1$ bzw. $c = 1$ erhalten wir als zugehörige Fixrichtungen

$$[x_1] = \left[\begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}\right] \text{ bzw. } [x_2] = \left[\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right].$$

2. Schritt: Zu den Fixrichtungen $[x_1]$ bzw. $[x_2]$ suchen wir alle Punkte x_0 mit $\varphi(x_0) \in x_0 + [x_1]$ bzw. alle Punkte y_0 mit $\varphi(y_0) \in y_0 + [x_2]$:

$$\varphi(x_0) \in x_0 + [x_1] \iff A\,x_0 + v = x_0 + a\,x_1,\ a \in \mathbb{R} \iff (A - E_3)\,x_0 - a\,x_1 = -v.$$

Dies ist ein inhomogenes lineares Gleichungssystem mit den Unbekannten x_0 und a und der zugehörigen erweiterten Matrix

$$\left[\begin{array}{rrrr|r} 6 & -6 & 11 & 2 & -4 \\ 0 & 0 & -1 & -1 & 2 \\ -4 & 4 & -8 & -2 & 4 \end{array}\right]$$

Als Lösungen erhalten wir

$$x_0 = \begin{bmatrix} 3 + \ b + 3/2\ a \\ b \\ -2 - a \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + a \begin{bmatrix} 3/2 \\ 0 \\ -1 \end{bmatrix} = \tilde{a} \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \tilde{a}, b \in \mathbb{R}\,.$$

Dies ist eine Ebene L_1 durch den Nullpunkt o, die die Fixpunktgerade h enthält. Weil x_1 Linearkombination der Vektoren

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix}$$

ist, liegen alle Fixgeraden mit Richtung $[x_1]$ in L_1 . Sie schneiden dann alle die Fixpunktgerade, weil x_1 und $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ linear unabhängig sind. Somit erhalten wir als Fixgeraden mit Richtung $[x_1]$ die in der Ebene L_1 liegende Geradenschar

$$\begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + [x_1]\,, \ b \in \mathbb{R}.$$

Auf die gleiche Art und Weise ergeben sich als Fixgeraden mit Richtung $[x_2]$ die zu der Fixpunktgeraden h parallelen Geraden

$$\begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + b \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + [x_2]\,, \ b \in \mathbb{R},$$

die alle in einer von L_1 verschiedenen Ebene

$$L_2 = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix} + [\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}]$$

durch h liegen.

Kapitel 4 Determinanten und Eigenwerte

§ 1 Determinanten

Wir wollen bei der Einführung der Determinante einer quadratischen Matrix $A \in \mathbb{K}^{n\times n}$ nicht den historischen Weg wählen, sondern vielmehr eine Definition geben, die es erlaubt, theoretische Überlegungen einfach durchzuführen. Zu diesem Zweck müssen wir uns nochmals kurz mit Permutationen beschäftigen.

Es sei

$$\pi = \begin{bmatrix} 1 & \cdots & i & \cdots & j & \cdots & m \\ \pi(1) & \cdots & \pi(i) & \cdots & \pi(j) & \cdots & \pi(m) \end{bmatrix} \in S_m .$$

Dann heißt jedes Paar $(i,j) \in \{1,...,m\} \times \{1,...,m\}$ mit $i < j$ und $\pi(i) > \pi(j)$ ein *Fehlstand* von π . Die Anzahl $F(\pi)$ der Fehlstände von π heißt die *Fehlstandszahl* von π.

Beispiel. Die Permutation

$$\pi = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{bmatrix}$$

besitzt die Fehlstände (1,2), (1,3), (1,4) und (2,3), also ist $F(\pi) = 4$.

Eigenschaften der Fehlstandszahl

(a) Für alle $\pi \in S_m$ gilt

$$\prod_{1 \le i < j \le m} \frac{\pi(j) - \pi(i)}{j - i} = (-1)^{F(\pi)}.$$

Beweis. Die Gleichung folgt, weil links im Zähler und im Nenner bis auf Vorzeichen dieselben Faktoren stehen und $\pi(j) - \pi(i)$ genau dann negativ ist, wenn (i,j) ein Fehlstand ist. ∎

(b) Für alle $\sigma, \pi \in S_m$ gilt: $(-1)^{F(\sigma\circ\pi)} = (-1)^{F(\sigma)} \cdot (-1)^{F(\pi)}$.

Beweis.

$$(-1)^{F(\sigma\circ\pi)} = \prod_{1 \le i < j \le m} \frac{\sigma(\pi(j)) - \sigma(\pi(i))}{j - i}$$

$$= \prod_{1 \le i < j \le m} \frac{\sigma(\pi(j)) - \sigma(\pi(i))}{\pi(j) - \pi(i)} \cdot \frac{\pi(j) - \pi(i)}{j - i}$$

$$= \prod_{1 \le \pi(i) < \pi(j) \le m} \frac{\sigma(\pi(j)) - \sigma(\pi(i))}{\pi(j) - \pi(i)} \prod_{1 \le i < j \le m} \frac{\pi(j) - \pi(i)}{j - i}$$

$$= (-1)^{F(\sigma)} \cdot (-1)^{F(\pi)}. \qquad \blacksquare$$

(c) $(-1)^{F(\pi)} = (-1)^{F(\pi^{-1})}$.

(d) $F(\tau^{(i,j)}) = 2\,(j-i) - 1$.

Beweis. Die Fehlstände der Transposition $\tau^{(i,j)}$ sind gerade die Paare $(i,i+1),\ldots,$ $(i,j-1)$, (i,j) und $(i+1,j),\ldots,(j-1,j)$. Dies ergibt zusammen $(j-i) + (j-1-i) =$ $2\,(j-i) - 1$ Fehlstände. $\blacksquare$

Bemerkung. Aus diesen Eigenschaften folgt direkt der Satz : *Eine Permutation ist genau dann gerade, wenn ihre Fehlstandszahl gerade ist.*

Beweis. Ist π eine gerade Permutation, so ist π Produkt einer geraden Anzahl von Transpositionen $\tau_1,\ldots,\tau_{2k}$. Mit (b) folgt dann $(-1)^{F(\pi)} = (-1)^{F(\tau_1)} \cdots (-1)^{F(\tau_{2k})} =$ $(-1)^{2k} = 1$, also ist $F(\pi)$ gerade. Ebenso zeigt man, daß für ungerades π auch $F(\pi)$ ungerade ist. $\blacksquare$

Nun können wir die angekündigte Definition geben.

Definition. Es sei $A = (\!(a_{ij})\!) \in \mathbb{K}^{n\times n}$. Dann heißt

$$\det A := \sum_{\pi \in S_n} (-1)^{F(\pi)}\, a_{\pi(1),1} \cdots a_{\pi(n),n}$$

Determinante der Matrix A. Statt $\det A$ schreiben wir auch $|A|$ bzw.

$$\det A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

Bemerkungen. **(a)** Für $n = 1$ gilt $\det A = a_{11}$.

Für $n = 2$ erhalten wir $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\, a_{22} - a_{21}\, a_{12}$.

Für $n = 3$ müssen wir über die Elemente der Gruppe S_3 summieren. Dies ergibt die folgende Regel (*Regel von Sarrus*):

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}$$

$$- a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}\,.$$

(b) Ist A eine obere bzw. untere Dreiecksmatrix, so folgt aus der Definition der Determinante sofort $\det A = a_{11} \cdots a_{nn}$.

(c) $\det E_n = 1$.

Satz 1. *Es sei* $A \in \mathbb{K}^{n\times n}$. *Dann gilt* $\det A = \det A^{\mathsf{T}}$.

Beweis. Es seien $A = (\!(a_{ij})\!)$ und $A^{\mathsf{T}} = (\!(\tilde{a}_{ij})\!)$. Dann ist

$$\det A^{\mathsf{T}} = \sum_{\pi \in S_n} (-1)^{F(\pi)}\, \tilde{a}_{\pi(1),1} \cdots \tilde{a}_{\pi(n),n}$$

$$= \sum_{\pi \in S_n} (-1)^{F(\pi)}\, a_{1,\pi(1)} \cdots a_{n,\pi(n)}$$

$$= \sum_{\pi \in S_n} (-1)^{F(\pi)} \, a_{\pi^{-1}\circ\pi(1),\pi(1)} \cdots a_{\pi^{-1}\circ\pi(n),\pi(n)}$$

$$= \sum_{\pi^{-1} \in S_n} (-1)^{F(\pi^{-1})} \, a_{\pi^{-1}(1),1} \cdots a_{\pi^{-1}(n),n} = \det A. \qquad \blacksquare$$

Bemerkung. Aus dem voranstehenden Beweis folgt unmittelbar

$$\det A = \sum_{\pi \in S_n} (-1)^{F(\pi)} \, a_{1,\pi(1)} \cdots a_{n,\pi(n)} .$$

Als nächstes wollen wir einige Eigenschaften der Abbildung $A \longmapsto \det A$ beweisen. Dabei wird insbesondere die Abhängigkeit von $\det A$ von den Spalten (Zeilen) von A eine Rolle spielen. Um diese Abhängigkeit besser überblicken zu können, definieren wir folgende Funktionen auf $(\mathbb{K}^n)^n$:

$$\Delta : (\mathbb{K}^n)^n \longrightarrow \mathbb{K} \ , \ \Delta(x_1,\ldots,x_n) := \det (x_1 \mid \cdots \mid x_n) .$$

$$\tilde{\Delta} : (\mathbb{K}^n)^n \longrightarrow \mathbb{K} \ , \ \tilde{\Delta}(y_1,\ldots,y_n) := \det \begin{bmatrix} y_1^T \\ \cdots \\ y_n^T \end{bmatrix}$$

Wegen Satz 1 gilt aber $\Delta = \tilde{\Delta}$, es genügt daher nur Δ zu untersuchen.

Satz 2. *Die Abbildung* $\Delta : (\mathbb{K}^n)^n \longrightarrow \mathbb{K}$ *hat folgende Eigenschaften:*

(a) Δ *ist* n*−fach multilinear, d.h. es gilt für alle* $j \in \{1, \ldots, n\}$, *alle* $a,b \in \mathbb{K}$ *und alle* $x_j , \tilde{x}_j \in \mathbb{K}^n$:

$$\Delta(\ldots,x_{j-1},a\,x_j + b\,\tilde{x}_j,x_{j+1},\ldots) = a\,\Delta(\ldots,x_{j-1},x_j,x_{j+1},\ldots) + b\,\Delta(\ldots,x_{j-1},\tilde{x}_j,x_{j+1},\ldots) .$$

(b) *Für alle* $\pi \in S_n$ *gilt:* $\Delta(x_{\pi(1)},\ldots,x_{\pi(n)}) = (-1)^{F(\pi)} \Delta(x_1,\ldots,x_n)$.

Insbesondere ist Δ *alternierend, d.h. für alle* $i,j \in \{1,\ldots,n\}$, $i \neq j$, *gilt:*

$$\Delta(\ldots,\underbrace{x_i}_{\substack{i-te\\Stelle}},\ldots,\underbrace{x_j}_{\substack{j-te\\Stelle}},\ldots) = -\,\Delta(\ldots,\underbrace{x_j}_{\substack{i-te\\Stelle}},\ldots,\underbrace{x_i}_{\substack{j-te\\Stelle}},\ldots)$$

(c) Δ *ist normiert, d.h.* $\Delta(e_1,\ldots,e_n) = 1$.

(d) *Die Vektoren* $x_1,\ldots,x_n$ *sind genau dann linear abhängig, wenn* $\Delta(x_1,\ldots,x_n) = 0$.

Beweis. (a) folgt direkt aus der Definition der Determinante.

(b) Nach Definition von Δ und der Determinante gilt für alle $\pi \in S_n$

$$\Delta(x_{\pi(1)},\ldots,x_{\pi(n)}) = \sum_{\sigma \in S_n} (-1)^{F(\sigma)}\, x_{\sigma(1),\pi(1)} \cdots x_{\sigma(n),\pi(n)}$$

$$\underset{\sigma=\tau\circ\pi}{=} \sum_{\tau \in S_n} (-1)^{F(\tau\circ\pi)}\, x_{\tau(1),1} \cdots x_{\tau(n),n} = (-1)^{F(\pi)}\, \Delta(x_1,\ldots,x_n)\,.$$

Der zweite Teil der Behauptung ergibt sich nun als Spezialfall, wenn man $\pi = \tau^{(i,j)}$ wählt und beachtet, daß die Fehlstandszahl einer Transposition ungerade ist.

(c) $\Delta(e_1,\ldots,e_n) = \det E_n = 1$.

(d) Die Vektoren $x_1,\ldots,x_n$ seien linear abhängig. Wir beweisen zunächst den folgenden *Spezialfall:*

Ist $x_i = x_j$ *für* $i \neq j$, *so ist* $\Delta(x_1,\ldots,x_n) = 0$.

Beweis. In $\Delta(x_1,\ldots,x_n)$ tritt mit jedem Summanden

$$(-1)^{F(\pi)}\, x_{\pi(1),1} \cdots x_{\pi(i),i} \cdots x_{\pi(j),j} \cdots x_{\pi(n),n}$$

auch der Summand

$$(-1)^{F(\pi\circ\tau^{(i,j)})}\, x_{\pi(1),1} \cdots x_{\pi(j),i} \cdots x_{\pi(i),j} \cdots x_{\pi(n),n}$$

auf, der wegen $x_i = x_j$ die Form

$$(-1)^{F(\pi)}\,(-1)^{F(\tau^{(i,j)})}\, x_{\pi(1),1} \cdots x_{\pi(i),i} \cdots x_{\pi(j),j} \cdots x_{\pi(n),n}$$

hat, also das $(-1)-$fache des ersteren ist. Somit gilt $\Delta(x_1,\ldots,x_n) = 0$. (Für Körper $\mathbb{K}$ der Charakteristik $\neq 2$ folgt der Spezialfall auch direkt aus (b).)

Nun können wir den allgemeinen Fall beweisen. Sind die Vektoren $x_1,\ldots,x_n$

linear abhängig, so gibt es ein $i \in \{1,...,n\}$ mit $x_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j x_j$, und wir erhalten mit Hilfe von (a) und dem Spezialfall, daß

$$\Delta(x_1,...,x_n) = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_j \Delta(x_1,...,x_{i-1},x_j,x_{i+1}, \ldots ,x_n) = 0.$$

Es sei nun umgekehrt $\Delta(x_1,...,x_n) = 0$. Sind die Vektoren $x_1,...,x_n$ linear unabhängig, so bilden sie eine Basis von $\mathbb{K}^n$ und die Standardbasisvektoren $e_1,...,e_n$ lassen sich in der Form $e_j = \sum_{k=1}^{n} a_{kj} x_k$ darstellen, $j = 1,...,n$. Mit (a), dem Spezialfall, (b) und der Voraussetzung folgt dann

$$\Delta(e_1,...,e_n) = \sum_{k_1, \ldots ,k_n=1}^{n} a_{k_1 1} \cdots a_{k_n n} \Delta(x_{k_1},...,x_{k_n})$$

$$= \sum_{\pi \in S_n} a_{\pi(1),1} \cdots a_{\pi(n),n} \Delta(x_{\pi(1)},...,x_{\pi(n)}) = 0,$$

ein Widerspruch zu (c). Also sind die Vektoren $x_1,...,x_n$ linear abhängig. ∎

Für das praktische Rechnen mit Determinanten ergeben sich aus Satz 2 zusammen mit Satz 1 folgende nützliche Regeln:

Satz 3. *Es sei $A \in \mathbb{K}^{n \times n}$. Dann gilt:*

(a) *Addition des Vielfachen einer Spalte (Zeile) zu einer anderen ändert* det A *nicht.*

(b) *Multiplikation einer Spalte (Zeile) mit $a \in \mathbb{K}$ vervielfacht* det A *um den Faktor a.*

(c) *Vertauschen von Spalten (Zeilen) ändert das Vorzeichen von* det A.

(d) *A ist genau dann regulär, wenn* det A *von Null verschieden ist.*

Wegen Satz 3 läßt sich das Prinzip des Gaußschen Algorithmus auch zur Berechnung von det A einsetzen.

Beispiel.

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

$$= \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & -1 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = 6.$$

Die Berechnung von $\det A$ läßt sich dadurch vereinfachen, daß man zuerst geeignet umformt und dann nach einer Spalte (Zeile) mit möglichst vielen Nullen "entwickelt". Diese Möglichkeit ergibt sich aus dem nachstehenden Satz, zu dessen Formulierung wir folgende Abkürzung verwenden:

Es seien $A \in \mathbb{K}^{n\times n}$ und $i,j \in \{1,\ldots,n\}$. Dann bezeichnen wir mit $A_{i,j}$ diejenige Matrix, die aus A durch Streichen der i–ten Zeile und der j–ten Spalte hervorgeht.

Satz 4. *Es sei $A = (\!(a_{ij})\!) \in \mathbb{K}^{n\times n}$. Dann gilt für $j = 1,\ldots,n$:*

(a) $$\det A = \sum_{k=1}^{n} (-1)^{k+j}\, a_{kj} \det A_{k,j}\,.$$ *("Entwicklung nach der j–ten Spalte")*

(b) $$\det A = \sum_{k=1}^{n} (-1)^{k+j}\, a_{jk} \det A_{j,k}\,.$$ *("Entwicklung nach der j–ten Zeile")*

Beweis. Zunächst beweisen wir wieder einen Spezialfall:

(*) $$\begin{vmatrix} b_{11} & \cdots & b_{1,n-1} & 0 \\ \vdots & & \vdots & \vdots \\ b_{n-1,1} & \cdots & b_{n-1,n-1} & 0 \\ b_{n,1} & \cdots & b_{n,n-1} & 1 \end{vmatrix} = \begin{vmatrix} b_{11} & \cdots & b_{1,n-1} \\ \vdots & & \vdots \\ b_{n-1,1} & \cdots & b_{n-1,n-1} \end{vmatrix}$$

Beweis. Für die linke Determinante gilt

$$\sum_{\pi \in S_n} (-1)^{F(\pi)} b_{\pi(1),1} \cdots b_{\pi(n),n} = \sum_{\substack{\pi \in S_n \\ \pi(n)=n}} (-1)^{F(\pi)} b_{\pi(1),1} \cdots b_{\pi(n-1),n-1}$$

$$= \sum_{\sigma \in S_{n-1}} (-1)^{F(\sigma)} b_{\sigma(1),1} \cdots b_{\sigma(n-1),n-1} .$$

Die letzte Summe ist aber nach Definition gerade die rechte Determinante.

(a) Wegen Satz 3 gilt für $j = 1,\ldots,n$

$$\det A = a_{1j} \begin{vmatrix} a_{11} & \cdots & 1 & \cdots & a_{1n} \\ a_{21} & \cdots & 0 & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & 0 & \cdots & a_{nn} \end{vmatrix} + \cdots + a_{nj} \begin{vmatrix} a_{11} & \cdots\cdots & 0 & \cdots\cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{n-1,1} & \cdots & 0 & \cdots & a_{n-1,n} \\ a_{n1} & \cdots\cdots & 1 & \cdots\cdots & a_{nn} \end{vmatrix} .$$

Alle Determinanten in dieser Summe können mit Hilfe von Satz 3 so umgeformt werden, daß wir (*) anwenden können. Wir erhalten durch Vertauschen von Zeilen und Spalten

$$k\left\{ \begin{vmatrix} a_{11} & \cdots & 0 & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{k1} & \cdots & 1 & \cdots & a_{kn} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & \underbrace{0}_{j} & \cdots & a_{nn} \end{vmatrix} \right. = (-1)^{n-j}(-1)^{n-k} \det A_{k,j} = (-1)^{k+j} \det A_{k,j}$$

und somit für $j = 1,\ldots,n$: $\det A = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A_{k,j}$.

(b) Die Behauptung kann analog bewiesen werden, folgt aber auch sofort aus Satz 1 und Teil (a). ∎

Beispiele. **(a)** Wir rechnen das obige Beispiel nochmals:

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \underset{\substack{\text{Entw.} \\ 1.\text{Sp.}}}{=} 2 \cdot \left| \begin{matrix} -1 & -1 & -1 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 3 & 1 & 2 \\ 0 & -1 & -1 & 1 \end{matrix} \right| \begin{matrix} \\ \\ \\ \\ \end{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right] -1$$

$$= 2 \cdot \begin{vmatrix} -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & -1 & -1 & 3 \end{vmatrix} \underset{\substack{\text{Entw.} \\ 1.\text{Ze.}}}{=} (-2) \cdot \begin{vmatrix} 2 & 1 & 0 \\ 3 & 1 & 0 \\ -1 & -1 & 3 \end{vmatrix} \underset{\substack{\text{Entw.} \\ 3.\text{Sp.}}}{=} (-2) \cdot 3 \cdot \begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} = 6 .$$

(b) *Vandermondesche Determinante.* Für alle $x_1,\ldots,x_k \in \mathbb{K}$ gilt:

$$\begin{vmatrix} 1 & 1 & 1 & \cdots\cdots & 1 \\ x_1 & x_2 & x_3 & \cdots\cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots\cdots & x_n^2 \\ \vdots & \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots\cdots & x_n^{n-1} \end{vmatrix} \begin{matrix} \\ \\ \\ \\ \end{matrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right] -x_1 \quad = \quad \begin{vmatrix} 1 & 1 & \cdots\cdots & 1 \\ 0 & (x_2-x_1) & \cdots\cdots & (x_n-x_1) \\ 0 & x_2(x_2-x_1) & \cdots\cdots & x_n(x_n-x_1) \\ \vdots & \vdots & & \vdots \\ 0 & x_2^{n-2}(x_2-x_1) & \cdots\cdots & x_n^{n-2}(x_n-x_1) \end{vmatrix}$$

$$= (x_2 - x_1)\cdots(x_n - x_1) \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_2 & x_3 & \cdots & x_n \\ \vdots & & & \vdots \\ x_2^{n-2} & x_3^{n-2} & \cdots & x_n^{n-2} \end{vmatrix}$$

$$= \cdots = (x_2 - x_1)\cdots(x_n - x_1)(x_3 - x_2)\cdots(x_n - x_2)\cdots(x_n - x_{n-1}) = \prod_{1 \le i < j \le n} (x_j - x_i) .$$

Für Produkte von Matrizen gilt der folgende Determinantenmultiplikationssatz.

Satz 5. *Es seien $A,B \in \mathbb{K}^{n\times n}$. Dann gilt:* $\det(A\,B) = \det A \cdot \det B$.

Beweis. Seien $a_1,\ldots,a_n$ die Spalten von A und $b_1,\ldots,b_n$ die von $B = (\!(b_{ij})\!)$. Dann ist

$$A\,B = (Ab_1 \mid \cdots \mid Ab_n) = (b_{11}a_1 + \cdots + b_{n1}a_n \mid \cdots \mid b_{1n}a_1 + \cdots + b_{nn}a_n),$$

woraus mit Hilfe von Satz 2 (a),(b),(d) folgt:

$$\begin{aligned}
\det(A\,B) &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} b_{i_1 1} \cdots b_{i_n n}\, \Delta(a_{i_1},\ldots,a_{i_n}) \\
&= \sum_{\pi \in S_n} b_{\pi(1),1} \cdots b_{\pi(n),n}\, \Delta(a_{\pi(1)},\ldots,a_{\pi(n)}) \\
&= \sum_{\pi \in S_n} (-1)^{F(\pi)}\, b_{\pi(1),1} \cdots b_{\pi(n),n}\, \Delta(a_1,\ldots,a_n) \\
&= \det A \cdot \det B\,. \qquad \blacksquare
\end{aligned}$$

Korollar 6 (*Kästchenmultiplikationssatz*).

(a) *Ist A von der Form*

$$A = \begin{bmatrix} B & O \\ C & D \end{bmatrix}$$

mit $B \in \mathbb{K}^{m\times m}$, $C \in \mathbb{K}^{(n-m)\times m}$, $O \in \mathbb{K}^{m\times(n-m)}$ und $D \in \mathbb{K}^{(n-m)\times(n-m)}$, so gilt

$$\det A = \det B \cdot \det D.$$

(b) *Ist A von der Form*

$$A = \begin{bmatrix} B & C \\ O & D \end{bmatrix}$$

mit $B \in \mathbb{K}^{m\times m}$, $O \in \mathbb{K}^{(n-m)\times m}$, $C \in \mathbb{K}^{m\times(n-m)}$ und $D \in \mathbb{K}^{(n-m)\times(n-m)}$, so gilt

$$\det A = \det B \cdot \det D.$$

Beweis. (a) Es ist

$$A = \begin{bmatrix} B & O \\ C & E_{n-m} \end{bmatrix} \cdot \begin{bmatrix} E_m & O \\ O & D \end{bmatrix},$$

woraus mit den Sätzen 5 und 4 sofort die Behauptung folgt.

(b) Satz 1 und Teil (a). ∎

Bemerkungen und Definitionen. **(a)** Ist die Matrix $A \in \mathbb{K}^{n\times n}$ regulär, so gilt $\det(A^{-1}) = (\det A)^{-1}$.

(b) Ähnliche Matrizen besitzen dieselbe Determinante.

(c) Sind V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $\Phi \in \mathrm{Hom}(V,V)$, so haben wegen der Aussage (b) alle Abbildungsmatrizen von Φ dieselbe Determinante. Wir setzen daher $\det \Phi := \det A_\Phi$ und nennen dies die *Determinante* der linearen Abbildung Φ.

Wir wollen nun Determinanten zur Lösung eines linearen Gleichungssystems bzw. zur Bestimmung der Inversen einer regulären Matrix heranziehen.

Satz 7 (*Cramersche Regel*). *Es seien $A \in \mathbb{K}^{n\times n}$ eine reguläre Matrix mit den Spalten $a_1,\dots,a_n$ und $b \in \mathbb{K}^n$. Dann ist die eindeutige Lösung des linearen Gleichungssystems $A\,x = b$ gegeben durch*

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \textit{mit} \quad x_i = \frac{\det\,(a_1 \,|\cdots|\, a_{i-1} \,|\, b \,|\, a_{i+1} \,|\cdots|\, a_n)}{\det A}.$$

Beweis. Nach Satz 3 gilt $x_i \cdot \det A = \det(a_1 \,|\cdots|\, a_{i-1} \,|\, x_i\, a_i \,|\, a_{i+1} \,|\cdots|\, a_n)$

$= \det(a_1 \,|\cdots|\, a_{i-1} \,|\, \sum_{j=1}^{n} x_j\, a_j \,|\, a_{i+1} \,|\cdots|\, a_n) = \det(a_1 \,|\cdots|\, a_{i-1} \,|\, b \,|\, a_{i+1} \,|\cdots|\, a_n)$. ∎

Satz 8. *Es sei $A \in \mathbb{K}^{n\times n}$ eine reguläre Matrix mit der Inversen $A^{-1} = (\!(b_{ij})\!)$. Dann gilt für alle $i,j = 1,\dots,n$:*

$$b_{ij} = (-1)^{i+j} (\det A)^{-1} \cdot \det A_{j,i}.$$

Beweis. Für $A^{-1} = (b_1 \,|\cdots|\, b_n)$ gilt $A\cdot(b_1 \,|\cdots|\, b_n) = E_n = (e_1 \,|\cdots|\, e_n)$, also $A\,b_j = e_j$ für $j = 1,\dots,n$. Mit den Sätzen 7 und 4 erhalten wir daher

$$b_{ij} = (\det A)^{-1} \cdot \det (a_1 \mid \cdots \mid a_{i-1} \mid e_j \mid a_{i+1} \mid \cdots \mid a_n) = (-1)^{i+j} (\det A)^{-1} \cdot \det A_{j,i} \,. \blacksquare$$

Neben dem hier gewählten direkten Zugang zur Determinante findet man in vielen Büchern auch den Zugang über eine Funktion (*Determinantenfunktion, Volumenfunktion*) auf $\mathbb{K}^{n\times n}$ bzw. $(\mathbb{K}^n)^n$ bzw. V^n mit gewissen Eigenschaften. Es wird dann gezeigt, daß es bis auf Normierung genau eine derartige Funktion Δ gibt und daß sie auf $(\mathbb{K}^n)^n$ die Form hat, die wir auf S. 166 als Definition benutzt haben. Wir wollen deshalb zum Schluß die entsprechende Aussage anführen.

Definition. Jede n–fach multilineare Abbildung $\Delta : V^n \longrightarrow \mathbb{K}$ eines n–dimensionalen $\mathbb{K}$–Vektorraums V, die für beliebige Vektoren $x_1,\ldots,x_n \in V$

$$\Delta(x_1,\ldots,x_n) = 0 \iff x_1,\ldots,x_n \text{ sind linear abhängig}$$

erfüllt, heißt eine *Determinantenform* auf V.

Bemerkungen. **(a)** Sind Δ_1 und Δ_2 Determinantenformen auf V, so existiert ein $c \in \mathbb{K}$ mit $\Delta_1 = c\,\Delta_2$. Insbesondere gibt es auf V genau eine Determinantenform, die auf einer gegebenen Basis den Wert 1 besitzt.

(b) Die Funktion Δ aus Satz 2 ist eine (normierte) Determinantenform auf $\mathbb{K}^n$. Sie ist durch die Eigenschaften (a), (c), (d) in Satz 2 eindeutig festgelegt. Die Eigenschaft alternierend zu sein folgt aus (a) und (d). Umgekehrt kann man (d) für alternierende Multilinearformen nur dann folgern, wenn die Charakteristik des Körpers $\mathbb{K}$ von 2 verschieden ist.

Zum Schluß wollen wir noch an die Bemerkung am Ende des Abschnittes über Matrizen in § I.4 anknüpfen. Die Definition der Determinante einer Matrix ist auch für Matrizen $A = (\!(a_{ij})\!)$ über einem kommutativen Ring mit 1 sinnvoll. Sind die Ringelemente a_{ij} speziell Polynome, $a_{ij} \in \mathbb{K}[X]$, so ist det A ebenfalls ein Polynom. Alle Ergebnisse dieses Paragraphen bleiben weiterhin richtig, solange sie nicht von der Division Gebrauch machen, insbesondere gelten die Sätze 4, 5 und 6 auch in diesem Fall.

§ 2 Eigenwerte und Diagonalisierbarkeit

Bei der geometrischen Beschreibung affiner Abbildungen in § 3.4 spielten die Fixpunkte und Fixrichtungen eine wesentliche Rolle. Die Bestimmung solcher Fixelemente führte auf Eigenwerte und Eigenvektoren der zugehörigen linearen Abbildung Φ. Wir werden nun allgemeiner untersuchen, welche Informationen über lineare Abbildungen Φ mit den Eigenwerten von Φ verbunden sind. In bestimmten Fällen lassen sich lineare Abbildungen vollständig über ihre Eigenwerte beschreiben.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum und $\Phi : V \longrightarrow V$ eine lineare Abbildung. Der Skalar $c \in \mathbb{K}$ heißt *Eigenwert* von Φ, falls ein Vektor $v \in V$, $v \neq o$, existiert mit

$$\Phi(v) = c\,v.$$

Der Vektor v heißt *Eigenvektor* von Φ zum Eigenwert c.

Bemerkungen und Bezeichnungen. **(a)** Die Menge aller Eigenvektoren von Φ zu einem festen Eigenwert c bildet zusammen mit dem Nullvektor einen Untervektorraum von V. Er heißt der zu c gehörige *Eigenraum* von Φ und wird mit E_c bezeichnet. Offensichtlich ist $E_c = \mathrm{Kern}(\Phi - c\cdot \mathrm{id}_V)$. Die Menge aller Eigenwerte von Φ heißt das *Spektrum* von Φ.

(b) Es sei $A \in \mathbb{K}^{n\times n}$. Die Matrix A legt in eindeutiger Weise die lineare Abbildung $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^n$, $x \longmapsto A\cdot x$, fest. Wir können deshalb auch von *Eigenwerten, Eigenvektoren* und *Eigenräumen* quadratischer Matrizen reden.

(c) Ähnliche Matrizen besitzen dieselben Eigenwerte.

Satz 9. *Es seien V ein $\mathbb{K}$–Vektorraum, $\Phi : V \longrightarrow V$ eine lineare Abbildung und $c_1,\ldots,c_k$ paarweise verschiedene Eigenwerte von Φ mit zugehörigen Eigenvektoren $v_1,\ldots,v_k$. Dann sind die Vektoren $v_1,\ldots,v_k$ linear unabhängig.*

Beweis. Wir führen den Beweis durch vollständige Induktion nach k:

$k = 1$: Definition des Eigenvektors.

$k-1 \longrightarrow k$: Sei $a_1 v_1 + \cdots + a_k v_k = o$. Wir wenden Φ darauf an und erhalten

$$o = \Phi(o) = a_1 c_1 v_1 + \cdots + a_k c_k v_k .$$

Andererseits gilt

$$o = c_1 o = a_1 c_1 v_1 + \cdots + a_k c_1 v_k ,$$

woraus $(c_2 - c_1) a_2 v_2 + \cdots + (c_k - c_1) a_k v_k = o$ folgt und somit nach Induktionsvoraussetzung $(c_2 - c_1) a_2 = \cdots = (c_k - c_1) a_k = 0$. Damit ergibt sich der Reihe nach $a_2 = \cdots = a_k = 0$, $a_1 v_1 = o$ und schließlich $a_1 = 0$. ∎

Korollar 10. *Jeder Endomorphismus eines n–dimensionalen $\mathbb{K}$–Vektorraums besitzt höchstens n Eigenwerte. Jede Matrix $A \in \mathbb{K}^{n\times n}$ besitzt höchstens n Eigenwerte.*

Wir wollen uns nun mit der Frage beschäftigen, wie man bei einem Endomorphismus Φ das Spektrum und die Eigenräume bestimmen kann. Im allgemeinen ist dies ein schwieriges Problem, im endlich dimensionalen Fall läßt es sich jedoch mit Hilfe des im vorigen Paragraphen erklärten Begriffs der Determinante vereinfachen und in vielen Fällen auch lösen.

Seien also V ein n–dimensionaler Vektorraum, $\Phi \in \mathrm{Hom}\,(V,V)$, A_Φ die Abbildungsmatrix von Φ bezüglich einer (geordneten) Basis B und $\hat{v}$ die Koordinatendarstellung von $v \in V$ bezüglich derselben Basis B. Dann ist $\Phi(v) = c\,v$, $c \in \mathbb{K}$, gleichwertig mit $A_\Phi \hat{v} = c \hat{v}$. Es genügt daher, Eigenwerte von Matrizen zu betrachten. Nun gilt:

$$\begin{gathered} c \text{ ist Eigenwert von } A \in \mathbb{K}^{n\times n} \\ \Longleftrightarrow \text{ es existiert } v \in \mathbb{K}^n,\ v \neq o, \text{ mit } A v = c v \text{ bzw. mit } (A - c E_n) v = o \\ \Longleftrightarrow (A - c E_n) v = o \text{ ist nichttrivial lösbar} \\ \Longleftrightarrow \mathrm{Rang}\,(A - c E_n) < n \Longleftrightarrow \det (A - c E_n) = 0 . \end{gathered}$$

Nach Definition der Determinante erhalten wir für $A = (\!(a_{ij})\!)$

$$\det (A - c E_n) = \sum_{\pi \in S_n} (-1)^{F(\pi)} (a_{\pi(1),1} - c\,\delta_{\pi(1),1}) \cdots (a_{\pi(n),n} - c\,\delta_{\pi(n),n}) .$$

Gemäß der Bemerkung von S. 174 können wir das Polynom

$$p = \det(A - X E_n) = \sum_{\pi \in S_n} (-1)^{F(\pi)} (a_{\pi(1),1} - X \delta_{\pi(1),1}) \cdots (a_{\pi(n),n} - X \delta_{\pi(n),n})$$

definieren. Es heißt *charakteristisches Polynom* (oder *Hauptpolynom*) von A.

Wir ordnen nach den Potenzen von X und erhalten

$$p = a_0 + a_1 X + a_2 X^2 + \cdots + a_{n-1} X^{n-1} + (-1)^n X^n$$

mit $a_0, \ldots, a_n \in \mathbb{K}$. Dabei ist speziell

$$a_0 = \det A ,$$

$$a_{n-1} = (-1)^{n-1}(a_{11} + \cdots + a_{nn}) = (-1)^{n-1} \text{ Spur } A ,$$

wobei Spur A definiert ist als Summe der Diagonalelemente von A.

Wir fassen zusammen:

Satz 11. *Genau dann ist $c \in \mathbb{K}$ ein Eigenwert der Matrix $A \in \mathbb{K}^{n \times n}$, wenn c Nullstelle des charakteristischen Polynoms p von A ist.*

Ist c ein Eigenwert von A, so ist der Eigenraum E_c gleich dem Lösungsraum des homogenen linearen Gleichungssystems $(A - c E_n)\, x = o$.

Bemerkung. Ähnliche Matrizen besitzen dasselbe charakteristische Polynom.

Beweis. Aus $A, B \in \mathbb{K}^{n \times n}$, $B = S^{-1} A S$, wobei $S \in \mathbb{K}^{n \times n}$ regulär ist, folgt $\det (S^{-1} A S - X E_n) = \det S^{-1} \cdot \det (A - X E_n) \cdot \det S = \det (A - X E_n)$. ∎

Damit können wir auch vom *charakteristischen Polynom* eines Endomorphismus Φ reden, und Satz 11 gilt entsprechend.

Beispiele. **(a)** Sei $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, $\mathbb{K} = \mathbb{R}$.

Es ist $p = \begin{vmatrix} 1 - X & 0 \\ 1 & 1 - X \end{vmatrix} = (1 - X)^2$, also ist $c = 1$ einziger Eigenwert

von A. Für den Eigenraum E_1 erhalten wir $E_1 = [\begin{pmatrix} 0 \\ 1 \end{pmatrix}]$.

(b) Sei $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{K}^{2\times 2}$.

Dann ist $p = \begin{vmatrix} -X & 1 \\ -1 & -X \end{vmatrix} = X^2 + 1$.

An diesem einfachen Beispiel sehen wir, daß die Existenz von Eigenwerten von dem zugrunde gelegten Körper abhängt.

Für $\mathbb{K} = \mathbb{R}$ existiert keine Nullstelle von p, also auch kein Eigenwert von A.

Für $\mathbb{K} = \mathbb{C}$ ergeben sich die Eigenwerte $c_1 = i$ und $c_2 = -i$. Die zugehörigen Eigenräume sind

$$E_{c_1} = [\begin{pmatrix} 1 \\ i \end{pmatrix}] \text{ und } E_{c_2} = [\begin{pmatrix} 1 \\ -i \end{pmatrix}].$$

Für $\mathbb{K} = \mathbb{F}_2$ ist $c = 1$ einziger Eigenwert mit zugehörigem Eigenraum $[\begin{pmatrix} 1 \\ 1 \end{pmatrix}]$.

(c) Gegeben sei die reelle Matrix

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

Es gilt:

$$p = \begin{vmatrix} -X & -1 & 1 & 1 \\ -1 & 1-X & -2 & 3 \\ 2 & -1 & -X & 0 \\ 1 & -1 & 1 & -X \end{vmatrix} = \begin{vmatrix} -X & -1 & 1 & 1 \\ 0 & -X & -1 & 3-X \\ 0 & 1 & -2-X & 2X \\ 1 & -1 & 1 & -X \end{vmatrix}$$

$$= \begin{vmatrix} -X & -1-X & 0 & 1 \\ 0 & -X & -1-X & 3-X \\ 0 & 1 & -1-X & 2X \\ 1 & 0 & 0 & -X \end{vmatrix} = (-X)(-X)\begin{vmatrix} -X & -1-X \\ 1 & -1-X \end{vmatrix} - \begin{vmatrix} -1-X & 0 & 1 \\ -X & -1-X & 3-X \\ 1 & -1-X & 2X \end{vmatrix}$$

$$= X^2(1+X)^2 + (1+X)[2X(-1-X) - (3-X)(-1-X)] - (X+1)^2$$

$$= (1 + X)^2 (X^2 - 3X + 2) = (1 + X)^2 (1 - X)(2 - X).$$

Also sind $c_1 = -1$, $c_2 = 1$ und $c_3 = 2$ die Eigenwerte von A. Die zugehörigen Eigenräume sind

$$E_{c_1} = [\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}] \ , \ E_{c_2} = [\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}] \ , \ E_{c_3} = [\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}] .$$

Zum Schluß dieses Paragraphen wollen wir diejenigen Endomorphismen eines endlich dimensionalen Vektorraums untersuchen, die bezüglich einer geeigneten Basis eine Abbildungsmatrix besitzen, die Diagonalgestalt hat, bzw. solche quadratischen Matrizen, die ähnlich zu einer Diagonalmatrix sind.

Definition. Eine Matrix $A \in \mathbb{K}^{n \times n}$ heißt *diagonalisierbar*, wenn sie zu einer Diagonalmatrix

$$\begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix}$$

ähnlich ist. Ein Endomorphismus Φ heißt *diagonalisierbar*, wenn es eine Abbildungsmatrix von Φ gibt, die Diagonalgestalt hat.

Bemerkung. Da alle Abbildungsmatrizen eines Endomorphismus Φ ähnlich sind, ist im Falle der Diagonalisierbarkeit von Φ jede Abbildungsmatrix von Φ diagonalisierbar und außerdem zur *gleichen* Diagonalmatrix ähnlich. Es kann aber mehrere ähnliche Diagonalmatrizen geben

Im folgenden Satz stellen wir einfache äquivalente Bedingungen für Diagonalisierbarkeit zusammen.

Satz 12. *Für einen Endomorphismus Φ eines $n-$dimensionalen $\mathbb{K}-$Vektorraumes V sind folgende Aussagen äquivalent :*

(a) *Φ ist diagonalisierbar.*

(b) *In V gibt es eine Basis aus Eigenvektoren von Φ.*

(c) *V ist die direkte Summe der Eigenräume von Φ.*

(d) *Die Summe der Dimensionen der Eigenräume von Φ ist n.*

Beweis. (a) $\Rightarrow$ (b): Nach Definition der Diagonalisierbarkeit gibt es eine Basis $B = (v_1,...,v_n)$ von V, bezüglich der die Abbildungsmatrix von Φ Diagonalgestalt

$$\begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix}$$

(mit nicht notwendig verschiedenen $c_1,...,c_n$) hat. Also gilt $\Phi(v_i) = c_i\, v_i$ für $i = 1,...,n$, d.h. die Vektoren v_i sind Eigenvektoren von Φ.

(b) $\Rightarrow$ (c): Es seien $c_1,...,c_k$ die verschiedenen Eigenwerte von Φ. Dann ist nach Satz 9 die Summe der Eigenräume von Φ direkt, und es gilt $E_{c_1} \oplus \cdots \oplus E_{c_k} \subset V$. Sei nun x ein beliebiger Vektor aus V. Bezüglich der Basis $(v_1,...,v_n)$ aus Eigenvektoren von Φ gilt dann $x = a_1\, v_1 + \cdots + a_n\, v_n$. Fassen wir alle Eigenvektoren zum gleichen Eigenwert c_i zusammen, so erhalten wir $x = \tilde{v}_1 + \cdots + \tilde{v}_k$ mit $\tilde{v}_i \in E_{c_i}$, $i = 1,...,k$.

(c) $\Rightarrow$ (d) ist trivial.

(d) $\Rightarrow$ (a): Es seien $E_{c_1},...,E_{c_k}$ die Eigenräume von Φ und $\dim E_{c_i} = n_i$ für $i = 1,...,k$. In jedem Eigenraum E_{c_i} wählen wir eine Basis B_i. Dann ist $B = B_1 \cup \cdots \cup B_k$ nach Satz 9 linear unabhängig, und wegen $n_1 + \cdots + n_k = n$ ist B sogar Basis von V. Bezüglich dieser Basis hat die Abbildungsmatrix von Φ Diagonalgestalt. ∎

Bemerkungen. **(a)** Satz 12 gilt analog, wenn wir Φ durch $A \in \mathbb{K}^{n\times n}$ und V durch $\mathbb{K}^n$ ersetzen.

(b) Ist $A \in \mathbb{K}^{n\times n}$ diagonalisierbar, und ist $(v_1,...,v_n)$ eine Basis von $\mathbb{K}^n$ aus Eigenvektoren von A mit $A\, v_i = c_i\, v_i$ für $i = 1,...,n$, so gilt für die reguläre Matrix $S = (v_1 \mid \cdots \mid v_n)$

$$S^{-1} A\, S \;=\; \begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & c_n \end{bmatrix}$$

Beweis. Die obige Diagonalmatrix ist gerade die Abbildungsmatrix von $x \longmapsto A\,x$ bezüglich der neuen Basis $(v_1,\ldots,v_n)$. ∎

Korollar 13. *Ein Endomorphismus Φ eines n–dimensionalen $\mathbb{K}$–Vektorraumes bzw. eine (n,n)–Matrix A mit n verschiedenen Eigenwerten ist diagonalisierbar.*

Es stellt sich nun die Frage, ob sich die Diagonalisierbarkeit von Φ bzw. A auch in anderen Fällen am charakteristischen Polynom p ablesen läßt. Wir zeigen dazu das folgende, für die Anwendung wichtigste Diagonalisierbarkeitskriterium.

Satz 14. *Es seien V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $\Phi : V \longrightarrow V$ eine lineare Abbildung. Dann gilt:*

Φ ist genau dann diagonalisierbar, wenn sein charakteristisches Polynom p in der Form

$$(*) \qquad p \;=\; (-1)^n\,(X-c_1)^{r_1} \cdots (X-c_k)^{r_k}$$

darstellbar ist mit $r_i \in \mathbb{N}$ und paarweise verschiedenen $c_i \in \mathbb{K}$ und wenn für $i = 1,\ldots,k$ gilt:

$$\dim \text{Bild}\,(\Phi - c_i \cdot \mathrm{id}_V) \;=\; n - r_i\,.$$

Bezeichnung. Besitzt ein Polynom p die Darstellung (*), so sagen wir auch, daß p in *Linearfaktoren zerfällt* und nennen r_i die *Vielfachheit* der Nullstelle c_i.

Bemerkungen. **(a)** Die zweite Forderung in Satz 14 besagt, daß die Dimension des Eigenraumes E_{c_i} mit der Vielfachheit r_i übereinstimmen muß $(i = 1,\ldots,k)$.

(b) Satz 14 gilt entsprechend für $A \in \mathbb{K}^{n\times n}$, wenn wir dim Bild $(\Phi - c_i \cdot \mathrm{id}_V)$ durch Rang $(A - c_i\,E_n)$ ersetzen.

Beweis von Satz 14. Es sei Φ diagonalisierbar. Dann besitzt Φ eine Abbildungsmatrix A_Φ der Form

$$A_\Phi = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & c_1 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & c_k \end{bmatrix} \begin{matrix} \} r_1 \\ \\ \\ \\ \} r_k \end{matrix}$$

wobei die c_i paarweise verschieden sind. Somit gilt

$$\det(A_\Phi - X E_n) = (c_1 - X)^{r_1} \cdots (c_k - X)^{r_k} = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k}$$

und

$$\dim \text{Bild}(\Phi - c_i \,\text{id}_V) = \text{Rang}(A_\Phi - c_i E_n)$$

$$= \text{Rang} \begin{bmatrix} c_1 - c_i & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & c_1 - c_i & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & c_k - c_i \end{bmatrix} = n - r_i .$$

Ist umgekehrt $p = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k}$, so sind $c_1, \ldots c_k$ gerade die Eigenwerte von Φ. Wegen $\dim \text{Bild}(\Phi - c_i \cdot \text{id}_V) = n - r_i$ gilt $\dim E_{c_i} = r_i$ und somit $\dim E_{c_1} + \cdots + \dim E_{c_k} = r_1 + \cdots + r_k = n$. Nach Satz 12 ist daher der Endomorphismus Φ diagonalisierbar. ∎

Die folgenden Beispiele setzen die Beispiele aus Abschnitt 4.2. fort.

Beispiele. **(a)** Wir betrachten die reelle Matrix

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

mit dem charakteristischen Polynom $p = (1-X)^2$. Wegen Rang $(A - E_2) = 1 \neq n-2 = 0$ ist A nicht diagonalisierbar.

(b) Die Matrix

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

ist über $\mathbb{R}$ nicht diagonalisierbar. Über $\mathbb{C}$ gilt dagegen $p = (i-X)(-i-X)$, also hat A zwei verschiedene Eigenwerte und ist somit diagonalisierbar. Über $\mathbb{K} = \mathbb{F}_2$ ist A nicht diagonalisierbar, da 1 einziger Eigenwert von A ist und dim $E_1 = 1$ gilt.

(c) Die reelle Matrix

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

hat das charakteristische Polynom $p = (1+X)^2\,(1-X)\,(2-X)$ und die Eigenwerte $c_1 = -1$, $c_2 = 1$ und $c_3 = 2$ mit den Vielfachheiten $r_1 = 2$, $r_2 = 1$ und $r_3 = 1$. Wie wir schon wissen, ist dim $E_{c_1} = 1$. Wegen $1 \neq r_1$ ist A nicht diagonalisierbar.

(d) Die reelle Matrix

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ 0 & 0 & 2 \end{bmatrix}$$

hat das charakteristische Polynom $p = (2-X)^2\,(7-X)$. Wegen

$$c_1 = 2,\ r_1 = 2,\ \text{Rang}\,(A - c_1 E_3) = 1 = 3-2\,,$$
$$c_2 = 7,\ r_2 = 1,\ \text{Rang}\,(A - c_2 E_3) = 2 = 3-1\,,$$

ist A diagonalisierbar. Ferner ist

$$E_{c_1} = \left[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}\right]\ ,\quad E_{c_2} = \left[\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right]\,.$$

Wir bilden aus den Basisvektoren von E_{c_1} und E_{c_2} die Transformationsmatrix

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

und erhalten damit

$$S^{-1} A\, S = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

§ 3 Der Satz von Cayley – Hamilton

Wie wir im vorigen Paragraphen gesehen haben, gibt es zu jedem diagonalisierbaren Endomorphismus Φ eines endlich dimensionalen $\mathbb{K}$–Vektorraums V eine Darstellung von V als direkte Summe der Eigenräume $E_{c_1},\ldots,E_{c_k}$ von Φ.

Wegen $E_{c_i} = \operatorname{Kern}(\Phi - c_i \cdot \mathrm{id}_V)$ gilt also

$$V = \operatorname{Kern}(\Phi - c_1 \cdot \mathrm{id}_V) \oplus \cdots \oplus \operatorname{Kern}(\Phi - c_k \cdot \mathrm{id}_V).$$

Wir betrachten nun den speziellen Endomorphismus

$$\Psi = (\Phi - c_1 \cdot \mathrm{id}_V) \circ \cdots \circ (\Phi - c_k \cdot \mathrm{id}_V).$$

Die einzelnen Abbildungen, aus denen Ψ zusammengesetzt ist, sind miteinander vertauschbar, denn es ist

$$(\Phi - c_i \cdot \mathrm{id}_V) \circ (\Phi - c_j \cdot \mathrm{id}_V) = \Phi^2 - (c_i + c_j)\,\Phi + c_i\, c_j \cdot \mathrm{id}_V$$

$$= (\Phi - c_j \cdot \mathrm{id}_V) \circ (\Phi - c_i \cdot \mathrm{id}_V).$$

Es sei nun $x \in V$ ein beliebiger Vektor. Dieser besitzt dann eine Darstellung $x = x_1 + \cdots + x_k$ mit $x_i \in E_{c_i}$. Wenden wir Ψ darauf an, so erhalten wir

$$\Psi(x) = \Psi(x_1) + \cdots + \Psi(x_k)$$

$$= (\Phi - c_2 \cdot \mathrm{id}_V) \circ \cdots \circ (\Phi - c_k \cdot \mathrm{id}_V) \circ (\Phi - c_1 \cdot \mathrm{id}_V)(x_1)$$

$$+ \cdots + (\Phi - c_1 \cdot \mathrm{id}_V) \circ \cdots \circ (\Phi - c_{k-1} \cdot \mathrm{id}_V) \circ (\Phi - c_k \cdot \mathrm{id}_V)(x_k)$$

$$= o + \cdots + o = o,$$

d.h. Ψ ist die Nullabbildung auf V.

Den Endomorphismus Ψ können wir uns auch so entstanden denken, daß wir in das Polynom $m = (X - c_1) \cdots (X - c_k)$ den Endomorphismus Φ eingesetzt haben.

Das Polynom m ist ein *annullierendes Polynom*, weil $m(\Phi)$ die Nullabbildung in V ist. Wie wir später sehen werden, ist in diesem speziellen Fall m das normierte Polynom kleinsten Grades, für das $m(\Phi) = O$ gilt. Es wird deshalb auch *Minimalpolynom* von Φ genannt.

Es stellt sich die Frage, ob es zu jedem Endomorphismus Φ von V bzw. zu jeder Matrix $A \in \mathbb{K}^{n\times n}$ ein solches Polynom gibt. Bevor wir diese Frage beantworten, wollen wir noch einmal präzisieren, was unter dem Einsetzen eines Endomorphismus bzw. einer quadratischen Matrix in ein Polynom verstanden werden soll.

Definition. Es seien V ein $\mathbb{K}$–Vektorraum, Φ ein Endomorphismus von V sowie

$$q = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n \in \mathbb{K}[X]$$

ein Polynom. Dann sei $q(\Phi)$ der Endomorphismus

$$q(\Phi) := a_0 \cdot \mathrm{id}_V + a_1 \cdot \Phi + a_2 \cdot \Phi^2 + \ldots + a_n \cdot \Phi^n,$$

wobei $\Phi^k = \underbrace{\Phi \circ \Phi \circ \cdots \circ \Phi}_{k \text{ Faktoren}}$ gilt.

Analog gilt für $A \in \mathbb{K}^{n\times n}$

$$q(A) := a_0 \cdot E_n + a_1 \cdot A + a_2 \cdot A^2 + \ldots + a_n \cdot A^n$$

mit $A^k = \underbrace{A \cdot A \cdots A}_{k \text{ Faktoren}}$.

Beispiel. Wir setzen eine Matrix A bzw. einen Endomorphismus Φ in das Polynom $q = 3 + X + X^4$ ein und erhalten $q(A) = 3E_n + A + A^4$ bzw. $q(\Phi) = 3\,\mathrm{id}_V + \Phi + \Phi^4$.

Bemerkung. Sei $\Phi \in \mathrm{Hom}(V,V)$ fest gewählt. Dann ist die Abbildung

$$f_\Phi : \mathbb{K}[X] \longrightarrow \mathrm{Hom}(V,V)$$

$$q \longmapsto q(\Phi)$$

ein Homomorphismus bezüglich der Vektorraumstruktur und der Ringstruktur, der sogenannte *Einsetzungshomomorphismus*, denn es gilt für alle $p,q \in \mathbb{K}[X]$ und alle $a \in \mathbb{K}$

$$(p+q)(\Phi) = p(\Phi) + q(\Phi),$$

$$(a \cdot p)(\Phi) = a \cdot p(\Phi),$$

$$(p \cdot q)(\Phi) = p(\Phi) \circ q(\Phi).$$

Die ersten beiden Eigenschaften sind trivial; f_Φ ist also linear. Es genügt deswegen, die letzte Eigenschaft auf der Basis $\{1, X, X^2, \ldots\}$ von $\mathbb{K}[X]$ nachzuprüfen. Dort ist sie aber unmittelbar klar.

Entsprechend wird bei festem $A \in \mathbb{K}^{n\times n}$ der Einsetzungshomomorphismus von $\mathbb{K}[X]$ in $\mathbb{K}^{n\times n}$ erklärt.

Man beachte, daß $f_\Phi(\mathbb{K}[X]) \subset \mathrm{Hom}(V,V)$ als Bild des kommutativen Rings $\mathbb{K}[X]$ ebenfalls kommutativ ist, der Ring $\mathrm{Hom}(V,V)$ dagegen nicht.

Wir wollen nun den Endomorphismus Φ in ein spezielles Polynom einsetzen, nämlich in sein charakteristisches Polynom.

Satz 15 (Cayley–Hamilton). *Es seien V ein n–dimensionaler $\mathbb{K}$–Vektorraum, Φ ein Endomorphismus von V und p das charakteristische Polynom von Φ. Dann ist $p(\Phi) = O$.*

Bemerkung. Man beachte, daß in der Behauptung rechts vom Gleichheitszeichen die Nullabbildung steht. Es ist also schon daher unsinnig, einen "Beweis" folgender Art führen zu wollen: $p(\Phi) = \det(\Phi - \Phi \circ \mathrm{id}_V) = \det(O) = 0$.

Beweis. Es ist zu zeigen, daß $p(\Phi)(v) = o$ für alle $v \in V$ gilt. Für $v = o$ ist dies klar. Sei also $v \neq o$. Wir betrachten nun für jedes $m \in \mathbb{N}_0$ die Vektoren

$$v, \Phi(v), \Phi^2(v), \ldots, \Phi^m(v),$$

wobei wir die Konvention $\Phi^0 := \mathrm{id}_V$ benutzen. Für $m = 0$ sind diese Vektoren linear

unabhängig, für $m \geq n$ linear abhängig. Also existiert ein kleinstes $m \in \mathbb{N}$, für das die Vektoren linear abhängig sind. Dann ist

$$\tilde{B} = \{v, \Phi(v), \Phi^2(v), \dots, \Phi^{m-1}(v)\}$$

linear unabhängig und

$$\tilde{B} \cup \{\Phi^m(v)\}$$

linear abhängig. Nach Kapitel 2 existieren also Skalare $a_0, \dots, a_{m-1} \in \mathbb{K}$ mit

$$\Phi^m(v) = a_0 v + a_1 \Phi(v) + \dots + a_{m-1} \Phi^{m-1}(v) .$$

Sei $q = a_0 + a_1 X + \dots + a_{m-1} X^{m-1} + (-1) X^m \in \mathbb{K}[X]$. Dann ist $q(\Phi) \in \operatorname{Hom}(V, V)$, und es gilt $q(\Phi)(v) = o$.

Wir wollen jetzt zeigen, daß auch $p(\Phi)(v) = o$ gilt. Dazu setzen wir $U := [\tilde{B}]$. Wegen $\Phi(\tilde{B}) \subset U$ ist $\Phi(U) \subset U$. Somit ist $\tilde{\Phi} := \Phi|_U$ ein Endomorphismus von U. Bezüglich der (geordneten) Basis $\tilde{B}$ besitzt $\tilde{\Phi}$ die Abbildungsmatrix

$$\tilde{A} = \begin{bmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & a_{m-2} \\ 0 & \cdots & 0 & 1 & a_{m-1} \end{bmatrix} \in \mathbb{K}^{m \times m} .$$

Wir bestimmen das charakteristische Polynom $\tilde{p}$ von $\tilde{\Phi}$.

$$\det(\tilde{A} - X E_m) = \begin{vmatrix} -X & 0 & \cdots & 0 & a_0 \\ 1 & -X & \cdots & 0 & a_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & -X & a_{m-2} \\ 0 & \cdots & 0 & 1 & a_{m-1} - X \end{vmatrix}$$

$$= \cdots = \begin{vmatrix} 0 & \cdots\cdots & 0 & a_0 + a_1 X + \cdots + a_{m-1} X^{m-1} - X^m \\ 1 & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{vmatrix} \begin{matrix} \\ \\ \\ \\ a_{m-2} + a_{m-1} X - X^2 \\ a_{m-1} - X \end{matrix} = (-1)^{m+1} q .$$

Also ist $\tilde{p} = (-1)^{m+1} q$. Wir ergänzen nun $\tilde{B}$ zu einer Basis B von V. Der Endomorphismus Φ hat bezüglich B eine Abbildungsmatrix A_Φ der Form

$$A_\Phi = \begin{bmatrix} \tilde{A} & C \\ O & D \end{bmatrix}$$

mit geeigneten Matrizen C und D. Für das charakteristische Polynom p von Φ folgt nach dem Kästchenmultiplikationssatz für Determinanten $p = \tilde{p} \cdot \overline{p}$, wobei $\overline{p}$ das charakteristische Polynom von D ist. Setzen wir $r = (-1)^{m+1} \overline{p}$, so gilt $p = q \cdot r = r \cdot q$ und $p(\Phi) = r(\Phi) \circ q(\Phi)$. Daraus folgt $p(\Phi)(v) = r(\Phi)(q(\Phi)(v)) = o$. Da v beliebig war, gilt $p(\Phi) = O$. ∎

Bemerkungen und Definition. **(a)** Satz 15 gilt analog für $A \in \mathbb{K}^{n \times n}$.

(b) Die Bedeutung dieses Satzes liegt nicht darin, daß es überhaupt ein Polynom $q \neq o$ gibt mit $q(\Phi) = O$, sondern daß das charakteristische Polynom von Φ diese Eigenschaft hat. Ein Polynom $q \neq o$ mit $q(\Phi) = O$ findet man nämlich leicht:

Wegen dim Hom $(V, V) = n^2$, sind die $n^2 + 1$ Vektoren id_V, Φ, $\Phi^2, \ldots, \Phi^{n^2}$ aus Hom (V, V) linear abhängig. Also gibt es Skalare $a_0, a_1, \ldots, a_{n^2}$, die nicht alle Null sind, mit

$$a_0 \,\mathrm{id}_V + a_1 \Phi + \cdots + a_{n^2} \Phi^{n^2} = O .$$

Für das Polynom $q = a_0 + a_1 X + \cdots + a_{n^2} X^{n^2}$ gilt dann $q \neq o$ und $q(\Phi) = O$.

(c) Das normierte Polynom $m \in \mathbb{K}[X]$ vom kleinsten Grad, das $m(\Phi) = O$ erfüllt, heißt *Minimalpolynom* von Φ. Es gibt ein und nur ein derartiges Polynom:

Seien nämlich m und $\tilde{m}$ normierte Polynome vom kleinsten Grad r mit $m(\Phi) = \tilde{m}(\Phi) = O$. Ist $m \neq \tilde{m}$, so ist $m - \tilde{m}$ vom Nullpolynom verschieden und hat einen Grad, der kleiner als r ist. Normieren wir $m - \tilde{m}$ noch, so erhalten wir ein annullierendes Polynom, dessen Grad kleiner als r ist. Dies ist ein Widerspruch zur Wahl von r. Also ist $m = \tilde{m}$. ∎

Analog wird das *Minimalpolynom* einer Matrix $A \in \mathbb{K}^{n\times n}$ erklärt.

Wegen $\Phi^0 = \mathrm{id}_V \neq O$ ist der Grad von m immer mindestens eins.

(d) Ähnliche Matrizen haben das gleiche Minimalpolynom, denn sie können als Abbildungsmatrizen eines festen Endomorphismus aufgefaßt werden.

Zwischen dem Minimalpolynom eines Endomorphismus Φ und den Polynomen $q \in \mathbb{K}[X]$ mit $q(\Phi) = O$ besteht folgender Zusammenhang:

Satz 16. *Das Minimalpolynom m des Endomorphismus Φ teilt jedes Polynom $q \in \mathbb{K}[X]$, für das $q(\Phi) = O$ gilt.*

Beweis. Wegen $m \neq o$ können wir nach Satz 1.14 q durch m dividieren: $q = s \cdot m + r$ mit $s, r \in \mathbb{K}[X]$. Dabei ist der Grad von r kleiner als der von m. Wir setzen Φ ein und erhalten $r(\Phi) = q(\Phi) - s(\Phi) \circ m(\Phi) = O$. Also muß r das Nullpolynom sein. ∎

Korollar 17. **(a)** *m teilt das charakteristische Polynom p von Φ.*

(b) *Die Nullstellen von m sind genau die Nullstellen von p, also gerade die Eigenwerte von Φ.*

Beweis von (b). Aus $m(c) = 0$ folgt wegen (a) $p(c) = 0$, und c ist Eigenwert von Φ. Ist umgekehrt c Nullstelle von p, also Eigenwert von Φ, so existiert ein Vektor $v \neq o$ mit $\Phi(v) = c\,v$. Für $i = 1,2,\ldots$ gilt dann $\Phi^i(v) = c^i\,v$, woraus $m(\Phi)(v) = m(c)\,v$ folgt. Also ist $m(c)\,v = o$ und wegen $v \neq o$ somit $m(c) = 0$. ∎

Bemerkungen. **(a)** Ist p von der Form

$$p = (-1)^n (X-c_1)^{r_1} \cdots (X-c_k)^{r_k}, \quad r_i \geq 1,$$

mit paarweise verschiedenen $c_1,\ldots,c_k \in \mathbb{K}$, so kann m nur die Form

$$m = (X-c_1)^{s_1} \cdots (X-c_k)^{s_k}$$

haben mit $1 \leq s_i \leq r_i$.

(b) Ist Φ diagonalisierbar und sind $c_1,\ldots,c_k$ die Eigenwerte von Φ, so ist

$$m = (X-c_1)\cdots(X-c_k)$$

ein annullierendes Polynom und somit wegen Korollar 17 das Minimalpolynom von Φ. Die Umkehrung gilt auch, wie wir im nächsten Abschnitt sehen werden.

Beispiele. **(a)** Wir betrachten den Endomorphismus $\Phi : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, $x \longmapsto A\,x$, mit

$$A = \begin{bmatrix} 7 & -6 & 11 \\ 0 & 1 & -1 \\ -4 & 4 & -7 \end{bmatrix}$$

Es ist

$$p = \begin{vmatrix} 7-X & -6 & 11 \\ 0 & 1-X & -1 \\ -4 & 4 & -7-X \end{vmatrix} = (7-X)\begin{vmatrix} 1-X & -1 \\ 4 & -7-X \end{vmatrix} - 4\begin{vmatrix} -6 & 11 \\ 1-X & -1 \end{vmatrix}$$

$$= -(X-1)^2(X+1).$$

Für das Minimalpolynom m gibt es nach Korollar 17 somit nur die Möglichkeiten $m = (X-1)(X+1)$ oder $m = (X-1)^2(X+1)$. Wegen

$$(A-E_3)(A+E_3) = \begin{bmatrix} 6 & -6 & 11 \\ 0 & 0 & -1 \\ -4 & 4 & -8 \end{bmatrix}\begin{bmatrix} 8 & -6 & 11 \\ 0 & 2 & -1 \\ -4 & 4 & -6 \end{bmatrix} \neq O$$

ist $m = (X-1)^2\,(X+1) = -p$ das Minimalpolynom von Φ.

Im nächsten Paragraphen werden wir für den Fall, daß das charakteristische Polynom in Linearfaktoren zerfällt ein allgemeines Verfahren zur Bestimmung des Minimalpolynoms herleiten.

(b) Der Satz von Cayley–Hamilton ist auch bei dem Problem nützlich, für ein Polynom q und eine quadratische Matrix A die Matrix $q(A)$ zu bestimmen. Ist z. B.

$$q = -1 + 2\,X + 6\,X^2 - 4\,X^3 - X^4$$

und A die Matrix aus Beispiel (a), so verfahren wir folgendermaßen:

Wir dividieren q durch das charakteristische Polynom $p = -(X-1)^2\,(X+1)$ von A und erhalten

$$q = (X+5)\,p - 2\,X + 4.$$

Also gilt

$$q(A) = (A + 5\,E_3)\,p(A) - 2\,A + 4\,E_3 = -2\,A + 4\,E_3$$

und daher

$$q(A) = \begin{bmatrix} -10 & 12 & -22 \\ 0 & 2 & 2 \\ 8 & -8 & 18 \end{bmatrix}.$$

Ist das Minimalpolynom m bekannt und Grad $m <$ Grad p, so arbeitet man vorteilhafter mit m.

§ 4 Jordansche Normalform

Wir haben in § 2 gesehen, daß die diagonalisierbaren Endomorphismen bei geeigneter Basiswahl eine Diagonalmatrix als Abbildungsmatrix besitzen. In diesem Abschnitt wollen wir versuchen, auch für nichtdiagonalisierbare Endomorphismen eine möglichst einfache Abbildungsmatrix zu erhalten.

Wie wir wissen, läßt sich bei einem diagonalisierbaren Endomorphismus Φ der Vektorraum V in eine direkte Summe von Untervektorräumen zerlegen:

$$V = E_{c_1} \oplus \cdots \oplus E_{c_k}.$$

Es gilt $\Phi(E_{c_i}) \subset E_{c_i}$. Untervektorräume $U \subset V$, die $\Phi(U) \subset U$ erfüllen, heißen *Φ-invariant.*

Unser nächstes Ziel ist es, auch für nichtdiagonalisierbares $\Phi \in \mathrm{End}(V)$ Φ-invariante Untervektorräume $V_1,\ldots,V_k$ zu finden, so daß

$$V = V_1 \oplus \cdots \oplus V_k$$

gilt und die Einschränkungen $\Phi|_{V_1},\ldots,\Phi|_{V_k}$ möglichst einfache Abbildungsmatrizen besitzen. Analog zu den Eigenräumen E_{c_i} versuchen wir, die Untervektorräume V_i in der Form $V_i = \mathrm{Kern}\, q_i(\Phi)$ mit geeigneten Polynomen $q_i \in \mathbb{K}[X]$ darzustellen. Dazu betrachten wir zunächst die speziellen Untervektorräume $\mathrm{Kern}\, q(\Phi)$ mit $q \in \mathbb{K}[X]$.

Bemerkungen. **(a)** Für jedes Polynom $q \in \mathbb{K}[X]$ ist $\mathrm{Kern}\, q(\Phi)$ ein Φ-invarianter Untervektorraum von V.

Beweis. Aus $v \in \mathrm{Kern}\, q(\Phi)$ folgt $q(\Phi)(v) = o$ und somit $q(\Phi)(\Phi(v)) = (\Phi \circ q(\Phi))(v) = o$. Also ist $\Phi(v) \in \mathrm{Kern}\, q(\Phi)$. ∎

(b) Ist das Polynom $q \in \mathbb{K}[X]$ ein Teiler von $r \in \mathbb{K}[X]$, so folgt $\mathrm{Kern}\, q(\Phi) \subset \mathrm{Kern}\, r(\Phi)$.

Beweis. Es ist $r = s \cdot q$ mit $s \in \mathbb{K}[X]$. Aus $v \in \mathrm{Kern}\, q(\Phi)$ folgt dann $r(\Phi)(v) = s(\Phi)(q(\Phi)(v)) = o$ und somit $v \in \mathrm{Kern}\, r(\Phi)$. ∎

(c) Sind $q,r \in \mathbb{K}[X]$ teilerfremde Polynome, so gilt

$$\text{Kern } (q\cdot r)\,(\Phi) \;=\; \text{Kern } q\,(\Phi) \oplus \text{Kern } r\,(\Phi).$$

Beweis. Sind q und r teilerfremd, so gibt es nach Satz 1.16 Polynome s und t mit $s\cdot q + t\cdot r = 1$. Setzen wir Φ ein, so erhalten wir $s\,(\Phi) \circ q\,(\Phi) + t\,(\Phi) \circ r\,(\Phi) = \mathrm{id}_V$ und somit für $v \in V$

$$v \;=\; \underbrace{s(\Phi)(q(\Phi)(v))}_{y} \;+\; \underbrace{t(\Phi)(r(\Phi)(v))}_{z}.$$

Aus $v \in \text{Kern } (q\cdot r)(\Phi)$ folgt

$$r\,(\Phi)(y) = (r\,(\Phi) \circ s\,(\Phi) \circ q\,(\Phi))(v) = (s\,(\Phi) \circ q\,(\Phi) \circ r\,(\Phi))(v) = s\,(\Phi)(o) = o\,,$$

also $y \in \text{Kern } r\,(\Phi)$. Ebenso erhalten wir $z \in \text{Kern } q\,(\Phi)$. Mit (b) ergibt sich daher

$$\text{Kern } (q\cdot r)\,(\Phi) \;=\; \text{Kern } q\,(\Phi) + \text{Kern } r\,(\Phi).$$

Schließlich folgt aus $v \in \text{Kern } r\,(\Phi) \cap \text{Kern } q\,(\Phi)$ direkt $z = y = o$ und somit $v = o$. Also ist die Summe direkt. ∎

(d) Sind $q_1,\ldots,q_k$ paarweise teilerfremde Polynome, so gilt

$$\text{Kern } (q_1 \cdots q_k)(\Phi) \;=\; \text{Kern } q_1\,(\Phi) \oplus \cdots \oplus \text{Kern } q_k\,(\Phi).$$

Beweis. Nach Bemerkung (a) von S. 67 sind die Polynome $q_1 \cdots q_{k-1}$ und q_k teilerfremd. Durch vollständige Induktion und mit (c) folgt die Behauptung. ∎

Die Bemerkungen (a),(b),(c) und (d) lassen sich insbesondere auf Polynome $q \in \mathbb{K}[X]$ anwenden, für die $q(\Phi) = O$ gilt. Wir erhalten so die folgende Aussage:

Satz 18. *Es seien V ein $\mathbb{K}$-Vektorraum,* dim $V = n$ *und $\Phi : V \longrightarrow V$ eine lineare Abbildung. Weiterhin sei $q \in \mathbb{K}[X]$ ein Polynom mit $q(\Phi) = O$ und $q = q_1 \cdots q_k$ eine Zerlegung von q in paarweise teilerfremde Faktoren $q_i \in \mathbb{K}[X]$, $i = 1,\ldots,k$. Dann sind die Untervektorräume $V_i = \text{Kern } q_i\,(\Phi)$ Φ-invariant und es gilt:*

$$V \;=\; V_1 \oplus \cdots \oplus V_k\,.$$

Bei geeigneter Basiswahl hat die zugehörige Abbildungsmatrix A von Φ die Form

$$A = \begin{bmatrix} \boxed{A_1} & & O \\ & \ddots & \\ O & & \boxed{A_k} \end{bmatrix},$$

wobei A_i die Abbildungsmatrix von $\Phi|_{V_i}$ ist.

Bemerkung. Speziell für das Minimalpolynom m von Φ gilt über Satz 18 hinaus:

Ist $m = m_1 \cdots m_k$ eine Zerlegung von m in teilerfremde normierte Faktoren $m_i \in \mathbb{K}[X]$, $i = 1,\ldots,k$, so ist m_i das Minimalpolynom von $\Phi|_{\text{Kern } m_i(\Phi)}$.

Beweis. Nach Satz 18 läßt sich jeder Vektor $v \in V$ in der Form $v = x_1 + \cdots + x_k$ mit $x_i \in \text{Kern } m_i(\Phi)$ darstellen. Wäre nun $\tilde{m}_i$ Minimalpolynom von $\Phi|_{\text{Kern } m_i(\Phi)}$ und $\tilde{m}_i \neq m_i$, so wäre Grad $\tilde{m}_i <$ Grad m_i und für das Polynom

$$\tilde{m} = m_1 \cdots m_{i-1} \cdot \tilde{m}_i \cdot m_{i+1} \cdots m_k$$

würde gelten:

$$\tilde{m}(\Phi)(v) = \tilde{m}(\Phi)(x_1) + \cdots + \tilde{m}(\Phi)(x_k) = m_2(\Phi) \circ \ldots \circ m_k(\Phi) \circ m_1(\Phi)(x_1) + \cdots +$$

$$m_1(\Phi) \circ \ldots \circ m_{i-1}(\Phi) \circ m_{i+1}(\Phi) \circ \ldots \circ m_k(\Phi) \circ \tilde{m}_i(\Phi)(x_i) + \cdots +$$

$$m_1(\Phi) \circ \ldots \circ m_k(\Phi)(x_k) = o$$

Also wäre $\tilde{m}(\Phi) = O$ und Grad $\tilde{m} <$ Grad m. Das ist ein Widerspruch, und somit ist m_i Minimalpolynom von $\Phi|_{\text{Kern } m_i(\Phi)}$. ∎

Als Folgerung aus Satz 18 erhalten wir ein weiteres Diagonalisierbarkeitskriterium.

Satz 19. *Es seien V ein $\mathbb{K}$-Vektorraum,* dim $V = n$ *und* $\Phi \in \text{Hom}(V,V)$. *Dann gilt:*

Φ ist genau dann diagonalisierbar, wenn das Minimalpolynom m in einfache

Linearfaktoren zerfällt, d.h. die Form

$$m = (X - c_1) \cdots (X - c_k)$$

hat mit paarweise verschiedenen $c_1, \ldots, c_k \in \mathbb{K}$.

Beweis. Wegen Bemerkung (b) von S. 191 müssen wir nur noch eine Richtung beweisen. Hat m die obige Form, so folgt nach Satz 18, daß $V = V_1 \oplus \cdots \oplus V_k$ ist mit $V_i = \operatorname{Kern}(\Phi - c_i \operatorname{id}_V) = E_{c_i}$. Nach Satz 12 ist Φ dann diagonalisierbar. ∎

Bemerkung. Ist $A \in \mathbb{K}^{n \times n}$, so gelten für die zugehörige lineare Abbildung $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^n$, $\Phi(x) = A\,x$, die Sätze 18 und 19 sowie die Bemerkung nach Satz 18.

Insbesondere ist also A genau dann zu einer Diagonalmatrix ähnlich, wenn sein Minimalpolynom in einfache Linearfaktoren zerfällt.

Es sei nun c ein Eigenwert des Endomorphismus Φ mit der Vielfachheit r. Für das charakteristische Polynom von Φ gilt also

$$p = (X - c)^r \cdot \overline{p}$$

mit $\overline{p}(c) \neq 0$, und für das Minimalpolynom m von Φ gilt dann wegen Korollar 17

$$m = (X - c)^s \cdot \overline{m}$$

mit $1 \leq s \leq r$, $\overline{m}(c) \neq 0$ und $\overline{m} \mid \overline{p}$. Nach Satz 18 ist dann

$$V = \operatorname{Kern}(\Phi - c \operatorname{id}_V)^r \oplus \operatorname{Kern} \overline{p}(\Phi) = \operatorname{Kern}(\Phi - c \operatorname{id}_V)^s \oplus \operatorname{Kern} \overline{m}(\Phi),$$

woraus wegen

$$\operatorname{Kern}(\Phi - c \operatorname{id}_V)^s \subset \operatorname{Kern}(\Phi - c \operatorname{id}_V)^r$$

und

$$\operatorname{Kern} \overline{m}(\Phi) \subset \operatorname{Kern} \overline{p}(\Phi).$$

aus Dimensionsgründen

$$\operatorname{Kern}(\Phi - c \operatorname{id}_V)^s = \operatorname{Kern}(\Phi - c \operatorname{id}_V)^r$$

und

$$\text{Kern } \overline{m}\,(\Phi) \;=\; \text{Kern } \overline{p}\,(\Phi)$$

folgen.

Definition. Der Untervektorraum $H_c := \text{Kern } (\Phi - c\,\text{id}_V)^r$ heißt *Hauptraum* zum Eigenwert c von Φ, die Vielfachheit s der Nullstelle c des Minimalpolynoms m von Φ heißt *Index* des Hauptraums H_c.

Wir werden später zeigen, daß $\dim H_c = r$ gilt und daß $(-1)^r\,(X-c)^r$ gerade das charakteristische Polynom von $\Phi\,|\,_{H_c}$ ist.

Der Index des Hauptraumes H_c kann folgendermaßen berechnet werden.

Satz 20. *Es seien V ein $\mathbb{K}-$Vektorraum, $\dim V = n$, $\Phi \in \text{Hom}\,(V,V)$ und c ein Eigenwert von Φ. Dann ist der Index des Hauptraums H_c zum Eigenwert c die kleinste Zahl $s \in \mathbb{N}$, für die*

$$\text{Kern } (\Phi - c\,\text{id}_V)^s \;=\; \text{Kern } (\Phi - c\,\text{id}_V)^{s+1}$$

gilt.

Beweis. Für den Index s von H_c und alle $k \in \mathbb{N}$ gilt

$$\text{Kern } (\Phi - c\,\text{id}_V)^s \;=\; \text{Kern } (\Phi - c\,\text{id}_V)^{s+k},$$

denn für das Polynom $\tilde{m} = (X-c)^{s+k}\,\overline{m}$ folgt wegen $\tilde{m}(\Phi) = (\Phi - c\,\text{id}_V)^k \circ m(\Phi) = O$ nach Satz 18

$$V \;=\; \text{Kern } (\Phi - c\,\text{id}_V)^{s+k} \oplus \text{Kern } \overline{m}\,(\Phi),$$

woraus wegen

$$\text{Kern } (\Phi - c\,\text{id}_V)^s \;\subset\; \text{Kern } (\Phi - c\,\text{id}_V)^{s+k}$$

aus Dimensionsgründen die Gleichheit folgt.

Ist nun $\overline{s}$ die kleinste natürliche Zahl mit

$$\text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}} \;=\; \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}+1},$$

so gilt $\overline{s} \leq s$, und durch vollständige Induktion erhalten wir für alle $k \in \mathbb{N}$

$$\text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}} = \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}+k}.$$

Sei nämlich $v \in \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}+k}$. Dann ist

$$(\Phi - c\,\text{id}_V)(v) \in \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}+k-1} = \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}},$$

also

$$v \in \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}+1} = \text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}}.$$

Aus $\overline{s} < s$ würde nun

$$\text{Kern } (\Phi - c\,\text{id}_V)^{\overline{s}} = \text{Kern } (\Phi - c\,\text{id}_V)^{s} = H_c$$

folgen, und somit wäre $(X-c)^{\overline{s}}\,\overline{m}$ annullierendes Polynom von Φ im Widerspruch dazu, daß $m = (X-c)^s\,\overline{m}$ das Minimalpolynom von Φ ist. Also folgt $\overline{s} = s$. ∎

Bemerkungen. **(a)** Für den Index s des Hauptraums H_c gibt es folgende weitere äquivalente Charakterisierungen:

$s \in \mathbb{N}$ ist die kleinste Zahl mit $\dim \text{Kern } (\Phi - c\,\text{id}_V)^s = \dim \text{Kern } (\Phi - c\,\text{id}_V)^{s+1}$,

$s \in \mathbb{N}$ ist die kleinste Zahl mit $\dim \text{Bild } (\Phi - c\,\text{id}_V)^s = \dim \text{Bild } (\Phi - c\,\text{id}_V)^{s+1}$,

$s \in \mathbb{N}$ ist die kleinste Zahl mit $\text{Rang } (A_\Phi - c\,E_n)^s = \text{Rang } (A_\Phi - c\,E_n)^{s+1}$.

(b) Für Matrizen $A \in \mathbb{K}^{n\times n}$ werden der Hauptraum H_c zum Eigenwert c und der Index von H_c entsprechend erklärt. Es ist

$$H_c = \{v \in \mathbb{K}^n \mid (A - c\,E_n)^r\, v = o\},$$

wobei r die Vielfachheit der Nullstelle c im charakteristischen Polynom von A ist, und entsprechend zu Satz 20 folgt dann für den Index des Hauptraumes H_c, daß er die kleinste natürliche Zahl s ist mit

$$\text{Rang } (A - c\,E_n)^s = \text{Rang } (A - c\,E_n)^{s+1}.$$

(c) Falls das charakteristische Polynom p von Φ bzw. von A vollständig in Linearfaktoren zerfällt,

$$p = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k},$$

mit paarweise verschiedenen $c_1,\ldots,c_k \in \mathbb{K}$, ergibt sich somit ein effektives Verfahren zur Bestimmung des Minimalpolynoms m, das ja dann die Form

$$m = (X - c_1)^{s_1} \cdots (X - c_k)^{s_k}$$

hat mit $1 \leq s_i \leq r_i$ für $i = 1,\ldots,k$.

Im Spezialfall

$$p = (-1)^n (X - c)^n,$$

wenn also $H_c = V$ bzw. $H_c = \mathbb{K}^n$ gilt (und nur dann!), ist der Index s die kleinste natürliche Zahl, für die $(\Phi - c\,\mathrm{id}_V)^s$ die Nullabbildung bzw. $(A - c\,E_n)^s$ die Nullmatrix ist.

Wir wenden dieses Verfahren in den folgenden Beispielen an.

Beispiele. **(a)** Gegeben sei die Matrix

$$A = \begin{bmatrix} -4 & 1 & 0 & 1 \\ -1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

mit dem charakteristischen Polynom $p = (X - 1)\,(X + 3)^3$ und den Eigenwerten $c_1 = 1$ und $c_2 = -3$. Wir wollen die zugehörigen Haupträume angeben. Es ist

$$m = (X - 1)\,(X + 3)^s$$

mit $1 \leq s \leq 3$. Wir bestimmen s:

$$\mathrm{Rang}\,(A + 3\,E_4) = \mathrm{Rang} \begin{bmatrix} -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = 2,$$

$$\mathrm{Rang}\,(A + 3\,E_4)^2 = \mathrm{Rang} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = 1,$$

$$\text{Rang}\,(A+3\,E_4)^3 \;=\; \text{Rang}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;=\; 1\,.$$

Also ist $s = 2$ und $m = (X-1)\,(X+3)^2$. Die zugehörigen Haupträume sind

$$H_{c_1} \;=\; E_{c_1} \;=\; \left[\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right] \;,\; H_{c_2} \;=\; \left[\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right] \supset E_{c_2} \;=\; \left[\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right].$$

(b) Die Matrix

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

besitzt das charakteristische Polynom $p = (X-2)^4$. Wir wollen das Minimalpolynom m und den Hauptraum H_2 bestimmen. Es ist $m = (X-2)^s$ mit $1 \le s \le 4$. Nun berechnen wir den Index s:

$$\text{Rang}\,(A-2\,E_4) \;=\; \text{Rang}\begin{bmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{bmatrix} \;=\; 1\,,$$

$$\text{Rang}\,(A-2\,E_4)^2 \;=\; \text{Rang}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;=\; 0\,.$$

Damit ist auch Rang $(A-2\,E_4)^3 = 0$, also $s = 2$ und $m = (X-2)^2$. Der Hauptraum ist hier trivialerweise der $\mathbb{R}^4$.

Wir kehren jetzt wieder zurück zur Zerlegung

$$p \;=\; (X-c)^r \cdot \overline{p}\,, \qquad \overline{p}\,(c) \neq 0\,,$$

mit der zugehörigen Darstellung $V = H_c \oplus \text{Kern}\;\overline{p}\,(\Phi)$ und der zu dieser Zerlegung

gehörenden Abbildungsmatrix

$$A_\Phi = \begin{bmatrix} \boxed{A_c} & O \\ O & \boxed{\overline{A}} \end{bmatrix}$$

Unser nächstes Ziel ist, im Hauptraum H_c eine geeignete Basis zu finden, bezüglich der die Abbildungsmatrix A_c von $\Phi|_{H_c}$ eine besonders einfache Form hat. Hierzu betrachten wir die aufsteigende Kette

$$\{o\} \underset{\neq}{\subset} E_c = \operatorname{Kern}(\Phi - c\,\mathrm{id}_V) \underset{\neq}{\subset} \cdots \underset{\neq}{\subset} \operatorname{Kern}(\Phi - c\,\mathrm{id}_V)^s = H_c .$$

Zur Abkürzung setzen wir $U_j := \operatorname{Kern}(\Phi - c\,\mathrm{id}_V)^j$ für $j = 1,\ldots,s$ und $U_0 := \{o\}$.

Wir stellen H_c nun als direkte Summe von s Untervektorräumen dar:

$$\begin{aligned} H_c &= U_{s-1} \oplus W_1 & \dim W_1 &=: q_1 \geq 1 \\ U_{s-1} &= U_{s-2} \oplus W_2 & \dim W_2 &=: q_2 \geq 1 \\ &\;\;\vdots & &\;\;\vdots \\ U_2 &= E_c \oplus W_{s-1} & \dim W_{s-1} &=: q_{s-1} \geq 1 \\ & & \dim E_c &=: q_s =: q \end{aligned}$$

Dann gilt

$$H_c = W_1 \oplus W_2 \oplus \cdots \oplus W_{s-1} \oplus E_c .$$

Die Untervektorräume W_i werden dabei nach dem folgenden Prinzip bestimmt:

Der erste Ergänzungsraum W_1 kann noch beliebig gewählt werden. Es sei $B_1 = (x_1^{(1)},\ldots, x_{q_1}^{(1)})$ eine geordnete Basis von W_1. Wir wenden auf diese q_1 Vektoren die Abbildung $(\Phi - c\,\mathrm{id}_V)$ an.

Behauptung. Für $x_i^{(2)} := (\Phi - c\,\mathrm{id}_V)(x_i^{(1)})$, $i = 1,\ldots,q_1$, gilt dann $x_i^{(2)} \in U_{s-1}$, $[x_1^{(2)},\ldots,x_{q_1}^{(2)}] \cap U_{s-2} = \{o\}$, und die Vektoren $x_1^{(2)},\ldots,x_{q_1}^{(2)}$ sind linear unabhängig.

Beweis. Es gilt $(\Phi - c\,\mathrm{id}_V)^{s-1}\,(x_i^{(2)}) = (\Phi - c\,\mathrm{id}_V)^{s}\,(x_i^{(1)}) = o$, also $x_i^{(2)} \in U_{s-1}$. Ist $v = a_1\,x_1^{(2)} + \cdots + a_{q_1}\,x_{q_1}^{(2)} \in U_{s-2}$, so ist $(\Phi - c\,\mathrm{id}_V)^{s-1}\,(a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)}) = (\Phi - c\,\mathrm{id}_V)^{s-2}\,(v) = o$, also $a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)} \in U_{s-1} \cap W_1$. Daraus folgt $a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)} = o$ und somit auch $v = o$.

Sei nun $a_1\,x_1^{(2)} + \cdots + a_{q_1}\,x_{q_1}^{(2)} = o$. Dann ist $(\Phi - c\,\mathrm{id}_V)\,(a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)}) = o$ und $a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)} \in U_{s-1} \cap W_1$. Daraus folgt $a_1\,x_1^{(1)} + \cdots + a_{q_1}\,x_{q_1}^{(1)} = o$ und daher $a_1 = \cdots = a_{q_1} = 0$. Also sind die Vektoren $x_1^{(2)},\ldots,x_{q_1}^{(2)}$ linear unabhängig.■

Der Komplementärraum W_2 wird nun so gewählt, daß er die linear unabhängigen Vektoren $x_1^{(2)},\ldots,x_{q_1}^{(2)}$ enthält. Diese werden zu einer Basis

$$B_2 = (x_1^{(2)},\ldots,x_{q_1}^{(2)}, x_{q_1+1}^{(2)},\ldots,x_{q_2}^{(2)})$$

von W_2 ergänzt. Auch hier gilt analog, daß die Vektoren

$$x_i^{(3)} := (\Phi - c\,\mathrm{id}_V)\,(x_i^{(2)}), \qquad i = 1,\ldots,q_2,$$

in U_{s-2} liegen, linear unabhängig sind, und daß $[x_1^{(3)},\ldots,x_{q_2}^{(3)}] \cap U_{s-3} = \{o\}$ gilt. Wir wählen W_3 derart, daß W_3 diese Vektoren enthält und ergänzen sie zu einer Basis B_3 von W_3

$$B_3 = (x_1^{(3)},\ldots,x_{q_1}^{(3)},\ldots,x_{q_2}^{(3)},\ldots,x_{q_3}^{(3)}).$$

Fahren wir so fort, so erhalten wir schließlich eine Basis von W_{s-1}:

$$B_{s-1} = (x_1^{(s-1)},\ldots,x_{q_1}^{(s-1)},\ldots,x_{q_2}^{(s-1)},\ldots,\ldots,x_{q_{s-1}}^{(s-1)}).$$

Ergänzen wir jetzt noch die Vektoren

$$x_i^{(s)} := (\Phi - c\,\mathrm{id}_V)\,(x_i^{(s-1)}), \qquad i = 1,\ldots,q_{s-1},$$

die ja nach den obigen Überlegungen in $U_1 = E_c$ liegen, zu einer Basis B_s von E_c,

$$B_s = (x_1^{(s)},\ldots,x_{q_1}^{(s)},\ldots,x_{q_2}^{(s)},\ldots,\ldots,x_{q}^{(s)}),$$

so sind wir fertig.

Wir schreiben die Basis $B_1 \cup \ldots \cup B_s$ des Hauptraums H_c als Schema hin und ordnen dann um:

$$
\begin{aligned}
B_1 &: \quad x_1^{(1)},\ldots,x_{q_1}^{(1)} \\
B_2 &: \quad x_1^{(2)},\ldots,x_{q_1}^{(2)} \;,\; x_{q_1+1}^{(2)},\ldots,x_{q_2}^{(2)} \\
B_3 &: \quad x_1^{(3)},\ldots,x_{q_1}^{(3)} \;,\; x_{q_1+1}^{(3)},\ldots,x_{q_2}^{(3)} \;,\; x_{q_2+1}^{(3)},\ldots,x_{q_3}^{(3)} \\
&\;\;\vdots \\
B_s &: \quad x_1^{(s)},\ldots,x_{q_1}^{(s)} \;,\; x_{q_1+1}^{(s)},\ldots,x_{q_2}^{(s)} \;,\; x_{q_2+1}^{(s)},\ldots,x_{q_3}^{(s)} \;,\ldots,\ldots,\; x_{q_{s-1}+1}^{(s)},\ldots,x_{q}^{(s)}
\end{aligned}
$$

Die geordnete Basis sei nun

$$B = (x_1^{(1)},\ldots,x_1^{(s)},x_2^{(1)},\ldots,x_2^{(s)},\ldots,x_{q_1}^{(1)},\ldots,x_{q_1}^{(s)},x_{q_1+1}^{(2)},\ldots,x_{q_1+1}^{(s)},\ldots,x_{q_2}^{(2)},\ldots,x_{q_2}^{(s)},x_{q_2+1}^{(3)},\ldots) .$$

Hierbei gilt für $j = 1,\ldots,s-1$ und entsprechende i

$$(\Phi - c\,\mathrm{id}_V)\,(x_i^{(j)}) \;=\; x_i^{(j+1)} ,$$

also

$$\Phi(x_i^{(j)}) \;=\; c\,x_i^{(j)} + x_i^{(j+1)} ,$$

und für $j = s$, $i = 1,\ldots,q$ gilt

$$\Phi(x_i^{(s)}) \;=\; c\,x_i^{(s)} .$$

Die ersten s Spalten der Abbildungsmatrix A_c bezüglich B haben demnach die Gestalt

$$
\left.
\begin{array}{ccccccc}
c & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\
1 & c & \cdot & & & & \vdots \\
0 & 1 & c & \cdot & & & \vdots \\
\vdots & \cdot & \cdot & \cdot & \cdot & \cdot & \vdots \\
\vdots & & \cdot & \cdot & \cdot & c & 0 \\
0 & \cdots & & 0 & 1 & c &
\end{array}
\right\} s
$$

$$
\begin{array}{ccccccc}
0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
\vdots & & & & & & \vdots \\
0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0
\end{array}
$$

Insgesamt hat A_c die Form

$$A_c = \begin{bmatrix} \boxed{A_1} & & & O \\ & \boxed{A_2} & & \\ & & \ddots & \\ O & & & \boxed{A_q} \end{bmatrix}$$

wobei die A_i wiederum von der Form

$$A_i = \begin{bmatrix} c & 0 & \cdots & \cdots & 0 \\ 1 & c & \ddots & & \vdots \\ 0 & 1 & c & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & c \end{bmatrix}$$

sind.

Wir nennen jedes solche A_i ein *Jordan–Kästchen* und die gesamte Matrix A_c den *Jordan–Block* zum Eigenwert c.

Ist dim $H_c = t$, so hat der Jordan–Block A_c genau t Zeilen. Weil das charakteristische Polynom von A_c offensichtlich $(c-X)^t$ ist, gilt $p = (X-c)^r \cdot \overline{p} = (c-X)^t \cdot \tilde{p}$, wobei $\tilde{p}$ das charakteristische Polynom von $\overline{A}$ ist. Es ist $t \leq r$. Wäre $t < r$, so wäre c Nullstelle von $\tilde{p}$ und es gäbe somit zum Eigenwert c einen Eigenvektor $v \in$ Kern $\overline{p}(\Phi)$, im Widerspruch zu $H_c \cap$ Kern $\overline{p}\,(\Phi) = \{o\}$. Also gilt $t = r$ und $(-1)^r\,\overline{p} = \tilde{p}$. Somit ist dim $H_c = r$ und $(-1)^r\,\overline{p}$ ist das charakteristische Polynom von $\Phi|_{\text{Kern}\,\overline{p}\,(\Phi)}$.

Die Jordan–Kästchen A_i haben eine Zeilenzahl (Länge) zwischen 1 und dem Index s. Dabei treten

q_1	Kästchen der Länge s	(mindestens eins !)
$q_2 - q_1$	Kästchen der Länge $s-1$	
$q_3 - q_2$	Kästchen der Länge $s-2$	
$\vdots$	$\vdots$	
$q - q_{s-1}$	Kästchen der Länge 1	

auf.

Insgesamt gibt es im Jordan–Block A_c genau $q = \dim E_c$ Jordan–Kästchen A_i. Die Anzahl der Jordan–Kästchen, welche die Länge l besitzen, ist

$$q_{s-l+1} - q_{s-l} = \dim W_{s-l+1} - \dim W_{s-l}$$

$$= (\dim U_{s-(s-l+1)+1} - \dim U_{s-(s-l+1)}) - (\dim U_{s-(s-l)+1} - \dim U_{s-(s-l)})$$

$$= \dim U_l - \dim U_{l-1} - \dim U_{l+1} + \dim U_l$$

$$= 2 \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^l - \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^{l+1} - \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^{l-1}.$$

$(l = 1,\ldots,s)$

Bemerkung. Durch Änderung der Reihenfolge in der Basis B kann man erreichen, daß die Einsen, die ja nur bei Jordan–Kästchen der Länge $l > 1$ auftreten, oberhalb der Diagonalen stehen, statt unterhalb.

Wir haben jetzt für den Hauptraum H_c eine Basis gefunden, für die die zugehörige Matrix A_c eine einfache Form annimmt. Wenn das charakteristische Polynom p von Φ in Linearfaktoren zerfällt,

$$p = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k},$$

mit paarweise verschiedenen $c_1,\ldots,c_k \in \mathbb{K}$, also $V = H_{c_1} \oplus \cdots \oplus H_{c_k}$ gilt, können wir die Jordan–Blöcke $A_{c_1},\ldots,A_{c_k}$ zur Jordanschen Normalform A_Φ von Φ zusammensetzen.

Satz 21 (Jordansche Normalform). *Es seien V ein n–dimensionaler $\mathbb{K}$–Vektorraum und $\Phi : V \longrightarrow V$ eine lineare Abbildung mit dem charakteristischen Polynom*

$$p = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k},$$

wo $c_1,\ldots,c_k \in \mathbb{K}$ paarweise verschieden sind, und dem Minimalpolynom

$$m = (X - c_1)^{s_1} \cdots (X - c_k)^{s_k}.$$

Dann gibt es eine geordnete Basis B von V, bezüglich der die Abbildungsmatrix A_Φ die Form

$$A_\Phi = \begin{bmatrix} \boxed{A_{c_1}} & & \\ & \ddots & \\ & & \boxed{A_{c_k}} \end{bmatrix}$$

besitzt mit Jordan–Blöcken A_{c_i} zu den Eigenwerten c_i,

$$A_{c_i} = \begin{bmatrix} \boxed{\begin{matrix} c_i & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & c_i \end{matrix}} & & \\ & \ddots & \\ & & \boxed{c_i} \end{bmatrix}$$

Der Jordan–Block A_{c_i} hat die Länge r_i, $i = 1,\ldots,k$. Innerhalb des Jordan–Blockes A_{c_i} zum Eigenwert c_i gibt es

$$2 \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^l - \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^{l+1} - \dim \text{Kern } (\Phi - c_i \,\mathrm{id}_V)^{l-1}$$

Jordan–Kästchen der Länge l, $l = 1,\ldots,s_i$.

Im Jordan–Block A_{c_i} treten insgesamt $\dim E_{c_i}$ Jordan–Kästchen auf. Es gibt mindestens ein Kästchen der Maximallänge s_i.

Bemerkungen und Bezeichnungen. **(a)** A_Φ heißt die *Jordansche Normalform* von Φ, und B heißt eine *Jordan–Basis* von Φ. Die Jordansche Normalform ist, bis auf die Reihenfolge der Kästchen, eindeutig.

(b) Ist $A \in \mathbb{K}^{n\times n}$ und zerfällt das charakteristische Polynom p von A in Linearfaktoren, $p = (-1)^n (X - c_1)^{r_1} \cdots (X - c_k)^{r_k}$, $c_1,\ldots,c_k \in \mathbb{K}$ paarweise verschieden, so gilt für die zugehörige lineare Abbildung $\Phi : \mathbb{K}^n \longrightarrow \mathbb{K}^n$, $\Phi(x) = A\,x$, der eben

bewiesene Satz 21.

A ist dann ähnlich zu der Jordanschen Normalform von Φ. Diese Matrix $\tilde{A}$ nennen wir die *Jordansche Normalform* der Matrix A.

Existieren die Jordanschen Normalformen der Matrizen $A,B \in \mathbb{K}^{n\times n}$, so sind diese Normalformen genau dann gleich, wenn A und B ähnlich sind.

(c) Die Forderung, daß das charakteristische Polynom bzw. das Minimalpolynom in Linearfaktoren zerfällt, kann entfallen, wenn $\mathbb{K} = \mathbb{C}$ ist.

Beispiele. **(a)** Der Endomorphismus $\Phi : V \longrightarrow V$ sei bezüglich der Basis (v_1,v_2,v_3) durch

$$\begin{aligned} \Phi(v_1) &= 3\,v_1 - v_2 + v_3 \\ \Phi(v_2) &= 4\,v_1 + 2\,v_3 \\ \Phi(v_3) &= 3\,v_1 - v_2 + 3\,v_3 \end{aligned}$$

gegeben. Die zugehörige Abbildungsmatrix A hat dann die Form

$$A = \begin{bmatrix} 3 & 4 & 3 \\ -1 & 0 & -1 \\ 1 & 2 & 3 \end{bmatrix}$$

Damit folgt $p = -(X-2)^3$ sowie $m = (X-2)^s$ mit $s \leq 3$. Nun gilt:

$$\text{Rang}\,(A - 2\,E_3) = \text{Rang} \begin{bmatrix} 1 & 4 & 3 \\ -1 & -2 & -1 \\ 1 & 2 & 1 \end{bmatrix} = 2\,.$$

Also ist dim $E_c = 1$ und es gibt nur ein Jordan–Kästchen. Dieses muß dann zwangsläufig die Länge 3 besitzen, woraus für den Index $s = 3$ folgt. Die Jordansche Normalform ist also

$$A_\Phi = \begin{bmatrix} \boxed{\begin{matrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{matrix}} \end{bmatrix}$$

Wir wollen nun auch eine Basis B bestimmen, bezüglich der diese Form angenommen wird. Es gilt

$$V = H_c = \text{Kern}\,(\Phi - 2\,\text{id}_V)^3\,,$$

$$U_2 = \text{Kern}\,(\Phi - 2\,\text{id}_V)^2\,,$$

$$U_1 = \text{Kern}\,(\Phi - 2\,\text{id}_V) = E_c\,.$$

Zur Bestimmung von U_2 lösen wir das LGS $(A - 2\,E_3)^2\,\hat{v} = o$:

$$\begin{bmatrix} 0 & 2 & 2 \\ 0 & -2 & -2 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Es ergibt sich $U_2 = \{v \in V \mid \hat{v} = (a_1, a_2, -a_2)\,,\ a_1, a_2 \in \mathbb{K}\} = [v_1, v_2 - v_3]$.

Zur Bestimmung von U_1 lösen wir das LGS $(A - 2\,E_3)\,\hat{v} = o$:

$$\begin{bmatrix} 1 & 4 & 3 \\ -1 & -2 & -1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Longleftrightarrow \begin{bmatrix} 1 & 4 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Es ergibt sich $U_1 = \{v \in V \mid \hat{v} = (a_3, -a_3, a_3)\,,\ a_3 \in \mathbb{K}\} = [v_1 - v_2 + v_3]$.

Nun wählen wir $x_1 \in H_c \setminus U_2$ beliebig, etwa $x_1 = v_2$; dann ist

$$x_2 := (\Phi - 2\,\text{id}_V)\,x_1 \quad \text{und} \quad x_2 \in U_2 \setminus U_1\,,$$

$$x_3 := (\Phi - 2\,\text{id}_V)\,x_2 = (\Phi - 2\,\text{id}_V)^2\,x_1 \quad \text{und} \quad x_3 \in U_1 = E_c\,.$$

Die zugehörigen Koordinatendarstellungen sind dann

$$\hat{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\,,\quad \hat{x}_2 = \begin{bmatrix} 1 & 4 & 3 \\ -1 & -2 & -1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}\,,\quad \hat{x}_3 = \begin{bmatrix} 0 & 2 & 2 \\ 0 & -2 & -2 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}.$$

Die gesuchte Jordan–Basis ist also $B = (x_1, x_2, x_3)$ mit

$$\begin{aligned} x_1 &= v_2 \\ x_2 &= 4\,v_1 - 2\,v_2 + 2\,v_3 \\ x_3 &= 2\,v_1 - 2\,v_2 + 2\,v_3 \end{aligned}$$

(b) (vgl. S.199) Wir betrachten die reelle Matrix

$$A = \begin{bmatrix} -4 & 1 & 0 & 1 \\ -1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}.$$

Es ist

$$p = (X-1)(X+3)^3\,, \text{ also } c_1 = 1,\, r_1 = 1 \text{ und } c_2 = -3,\, r_2 = 3\,,$$

$$m = (X-1)\,(X+3)^2\,, \text{ also } s_1 = 1 \text{ und } s_2 = 2\,.$$

Es gibt daher einen Jordan–Block der Länge $r_1 = 1$ zum Eigenwert $c_1 = 1$ und einen solchen der Länge $r_2 = 3$ zum Eigenwert $c_2 = -3$. In diesem Block existiert mindestens ein Jordan–Kästchen der maximalen Länge $s_2 = 2$. Also lautet die Jordansche Normalform von A

$$\tilde{A} = \begin{bmatrix} \boxed{\begin{matrix} -3 & 0 & 0 \\ 1 & -3 & 0 \\ 0 & 0 & -3 \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \boxed{1} \end{bmatrix}$$

Wir wollen eine Transformationsmatrix S bestimmen, für die $S^{-1} A\, S = \tilde{A}$ gilt. Dazu bestimmen wir eine Jordan–Basis der zugehörigen linearen Abbildung Φ :

Für den Hauptraum zum Eigenwert $c_1 = 1$ erhalten wir

$$H_{c_1} = \{v \in \mathbb{R}^4 \mid (A - E_4)\, v = o\} = \left[\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right].$$

Für den Hauptraum zum Eigenwert $c_2 = -3$ erhalten wir

$$H_{c_2} = \{v \in \mathbb{R}^4 \mid (A + 3\,E_4)^2\, v = o\} = U_2 = [\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\end{bmatrix}]$$

und für den Eigenraum E_{c_2}

$$E_{c_2} = \{v \in \mathbb{R}^4 \mid (A + 3\,E_4)\, v = o\} = U_1 = [\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}]\,.$$

Dabei sind

$$(A + 3\,E_4) = \begin{bmatrix}-1 & 1 & 0 & 1\\ -1 & 1 & 0 & 1\\ 0 & 0 & 4 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \quad \text{und} \quad (A + 3\,E_4)^2 = \begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 16 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

Wir wählen $x_1 \in U_2 \setminus U_1$ beliebig; der Basisvektor x_2 ist dann festgelegt:

$$x_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}\,, \quad x_2 = (A + 3\,E_4)\,x_1 = \begin{bmatrix}-1\\-1\\0\\0\end{bmatrix} \in U_1\,.$$

Nun ergänzen wir x_2 mit dem Vektor

$$x_3 = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} \in U_1$$

zu einer Basis von U_1. Bezüglich x_1, x_2, x_3 ergibt sich dann der Jordan–Block zum Eigenwert $c_2 = -3$. Insgesamt erhalten wir als Transformationsmatrix

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

Kontrolle: $S^{-1}AS = \tilde{A} \iff AS = S\tilde{A}$,

$$AS = \begin{bmatrix} -4 & 1 & 0 & 1 \\ -1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -4 & 3 & -3 & 0 \\ -1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & 0 \end{bmatrix}$$

$$S\tilde{A} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 & 0 \\ 1 & -3 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -4 & 3 & -3 & 0 \\ -1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & 0 \end{bmatrix}$$

(c) (vgl. S.183) Sei

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

Hier ist

$$p = (X+1)^2 (X-1)(X-2) \ , \ c_1 = -1, \ c_2 = 1, \ c_3 = 2 \ ,$$

und $1 \leq s_1 \leq r_1 = 2$, $r_2 = s_2 = 1$, $r_3 = s_3 = 1$. Weil A nicht diagonalisierbar ist, muß $s_1 > 1$ gelten. Also ist $m = p$.

Die Haupträume H_{c_2} und H_{c_3} sind also eindimensional, der Hauptraum H_{c_1} hat die Dimension 2. Im zugehörigen Jordan–Block A_{c_1} gibt es ein Kästchen der maximalen Länge 2.

Damit ergibt sich als Jordansche Normalform für A die Matrix

$$\tilde{A} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Wir erhalten wieder eine Transformationsmatrix S mit $S^{-1}AS = \tilde{A}$, indem wir in den einzelnen Haupträumen jeweils Jordan–Basen bestimmen:

Für den Hauptraum H_{c_1} ergibt sich

$$H_{c_1} = \{v \in \mathbb{R}^4 \mid (A + E_4)^2 v = o\} = [\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}],$$

denn es gilt

$$\begin{bmatrix} 5 & -5 & 5 & -1 \\ -4 & 4 & -4 & 8 \\ 5 & -5 & 5 & -1 \\ 5 & -5 & 5 & -1 \end{bmatrix} v = o \iff v = \begin{bmatrix} a_1 \\ a_1 + a_3 \\ a_3 \\ 0 \end{bmatrix}, \quad a_1, a_3 \in \mathbb{K}.$$

Für den Untervektorraum $U_1 \subset H_{c_1}$ erhalten wir

$$U_1 = \{v \in \mathbb{R}^4 \mid (A + E_4) v = o\} = [\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}],$$

denn es gilt

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ -1 & 2 & -2 & 3 \\ 2 & -1 & 1 & 0 \\ 1 & -1 & 1 & 1 \end{bmatrix} v = o \iff v = \begin{bmatrix} 0 \\ a \\ a \\ 0 \end{bmatrix}, \quad a \in \mathbb{K}.$$

Wir wählen $x_1 \in H_{c_1} \setminus U_1$ beliebig; dann ist x_2 festgelegt:

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad x_2 = (A + E_4)\, x_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Für die Haupträume H_{c_2} und H_{c_3} ergibt sich

$$H_{c_2} = E_{c_2} = [\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}], \quad H_{c_3} = E_{c_3} = [\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}].$$

Damit erhalten wir als Transformationsmatrix

$$S = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Kontrolle:

$$A\,S = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$S\,\tilde{A} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

(d) (abstrakt) Es sei $A \in \mathbb{R}^{12\times 12}$ mit $p = (2-X)^{12}$, d.h. $c = 2$ ist 12–facher Eigenwert und der Hauptraum H_c ist der Gesamtraum $\mathbb{R}^{12}$. Weiterhin gelte

$$\text{Rang}\,(A - 2\,E_{12}) = 7 \;,\; \text{Rang}\,(A - 2\,E_{12})^2 = 4 \;,$$

$$\text{Rang}\,(A - 2\,E_{12})^3 = 2 \;,\; \text{Rang}\,(A - 2\,E_{12})^4 = 0.$$

Damit ist $(A - 2\,E_{12})^4 = O$, also $s = 4$ und $m = (X-2)^4$. Für die Untervektorräume U_i gilt:

$$E_c = U_1 \underset{\neq}{\subset} U_2 \underset{\neq}{\subset} U_3 \underset{\neq}{\subset} U_4 = H_c\,.$$

Die zugehörigen Dimensionen sind:

$$\dim U_1 = 12 - 7 = 5 \;,\; \dim U_2 = 12 - 4 = 8 \;,\; \dim U_3 = 12 - 2 = 10 \;,\; \dim U_4 = 12.$$

Wegen $\dim E_c = 5$ gibt es 5 Jordan–Kästchen und zwar

$2\dim U_4 - \dim U_5 - \dim U_3 = 24 - 12 - 10 = 2$ Kästchen der Länge 4 ,

$2\dim U_3 - \dim U_4 - \dim U_2 = 20 - 12 - 8 = 0$ Kästchen der Länge 3 ,

$2\dim U_2 - \dim U_3 - \dim U_1 = 16 - 10 - 5 = 1$ Kästchen der Länge 2 ,

$2\dim U_1 - \dim U_2 - \dim U_0 = 10 - 8 - 0 = 2$ Kästchen der Länge 1 .

Damit ergibt sich folgende Jordansche Normalform von A (die freien Stellen sind mit Nullen aufzufüllen),

$$\tilde{A} = \begin{bmatrix} 2 & 0 & 0 & 0 & & & & & & \\ 1 & 2 & 0 & 0 & & & & & & \\ 0 & 1 & 2 & 0 & & & & & & \\ 0 & 0 & 1 & 2 & & & & & & \\ & & & & 2 & 0 & 0 & 0 & & \\ & & & & 1 & 2 & 0 & 0 & & \\ & & & & 0 & 1 & 2 & 0 & & \\ & & & & 0 & 0 & 1 & 2 & & \\ & & & & & & & & 2 & 0 \\ & & & & & & & & 1 & 2 \\ & & & & & & & & & & 2 & \\ & & & & & & & & & & & 2 \end{bmatrix}$$

Zum Schluß dieses Paragraphen wollen wir noch den Fall ansprechen, daß das charakteristische Polynom p **nicht** in Linearfaktoren zerfällt, also die Form

$$p = (X - c_1)^{r_1} \cdots (X - c_k)^{r_k} \cdot \overline{p}$$

hat, wobei $\overline{p} \in \mathbb{K}[X]$ keine Nullstelle besitzt und Grad $\overline{p} > 1$ ist. Hier ergibt sich die Normalform

$$A_\Phi = \begin{bmatrix} \boxed{A_{c_1}} & & & \\ & \ddots & & \\ & & \boxed{A_{c_k}} & \\ & & & \boxed{\overline{A}} \end{bmatrix}$$

die der Zerlegung $V = H_{c_1} \oplus \cdots \oplus H_{c_k} \oplus \overline{V}$ entspricht. Neben den Jordan–Blöcken A_{c_i}, $i = 1,\ldots,k$, tritt also noch der Block $\overline{A}$ auf, der zu dem invarianten Unterraum $\overline{V} = \text{Kern } \overline{p}\,(\Phi)$ gehört. Es gilt

$$\overline{V} = \bigcap_{i=1}^{k} \text{Bild } (\Phi - c_i \,\text{id}_V)^{r_i}$$

(Übungsaufgabe).

Im Fall $\mathbb{K} = \mathbb{R}$ können wir auch für $\overline{A}$ eine einfachere Form angeben. Diese *reelle Normalform* wird im nächsten Abschnitt hergeleitet werden.

§ 5 Reelle Jordansche Normalform

Zur Herleitung der reellen Jordanschen Normalform benutzen wir folgende Bemerkung über die Zerlegung reeller Polynome.

Bemerkung. Jedes normierte Polynom $p \in \mathbb{R}[X]$ besitzt eine Zerlegung

$$p = (X-c_1)^{r_1} \cdots (X-c_k)^{r_k} (X^2 + a_1 X + b_1)^{t_1} \cdots (X^2 + a_m X + b_m)^{t_m}$$

mit c_i , a_j , $b_j \in \mathbb{R}$, wobei die Polynome $(X^2 + a_j X + b_j)^{t_j}$, $j = 1,\ldots,m$, keine reelle Nullstelle haben.

Beweis. Das reelle Polynom $p = d_0 + d_1 X + \cdots + d_n X^n$, $d_i \in \mathbb{R}$, $d_n = 1$, läßt sich über $\mathbb{C}$ in Linearfaktoren zerlegen, $p = (X - c_1) \cdot \ldots \cdot (X - c_n)$, $c_i \in \mathbb{C}$. Für das Polynom $\overline{p} := (X - \overline{c}_1) \cdot \ldots \cdot (X - \overline{c}_n)$ gilt dann $\overline{p} = \overline{d}_0 + \overline{d}_1 X + \cdots + \overline{d}_n X^n = p$. Also ist mit jeder komplexen Nullstelle c von p der Vielfachheit t auch die konjugiert komplexe Zahl $\overline{c}$ eine Nullstelle der Vielfachheit t.

Seien nun $c_1,\ldots,c_k$ die reellen Nullstellen von p und $c_1', \overline{c_1'}, \ldots, c_m', \overline{c_m'}$ die verschiedenen komplexen Nullstellen. Dann folgt

$$\begin{aligned} p &= (X-c_1)^{r_1} \cdots (X-c_k)^{r_k} (X-c_1')^{t_1} (X-\overline{c_1'})^{t_1} \cdots (X-c_m')^{t_m} (X-\overline{c_m'})^{t_m} \\ &= (X-c_1)^{r_1} \cdots (X-c_k)^{r_k} (X^2 - \underbrace{(c_1' + \overline{c_1'})}_{-a_1} X + \underbrace{c_1' \, \overline{c_1'}}_{b_1})^{t_1} \\ &\quad \cdots (X^2 - \underbrace{(c_m' + \overline{c_m'})}_{-a_m} X + \underbrace{c_m' \, \overline{c_m'}}_{b_m})^{t_m} \end{aligned}$$

∎

Wir betrachten jetzt die Matrix $A \in \mathbb{R}^{n \times n}$. Seien p das charakteristische Polynom von A und $c, \overline{c}$ ein Paar konjugiert komplexer Nullstellen von p. Bei der Zerlegung $c = a + i\,b$ in Real- und Imaginärteil können wir o.B.d.A. $b > 0$ voraussetzen. Wegen

$$(A - c\,E_n)^j\, z \;=\; o \iff (A - \overline{c}\,E_n)^j\, \overline{z} \;=\; o$$

für $z \in \mathbb{C}^n$, ist

$$\text{Rang}\,(A - c\,E_n)^j \;=\; \text{Rang}\,(A - \overline{c}\,E_n)^j\,.$$

Dies bedeutet, daß in der komplexen Jordanschen Normalform von $A = A_\Phi$ die Jordan–Kästchen immer in Paaren gleicher Länge auftreten,

$$\begin{bmatrix} c & 0 & \cdots & 0 \\ 1 & c & \ddots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & c \end{bmatrix} \qquad \begin{bmatrix} \overline{c} & 0 & \cdots & 0 \\ 1 & \overline{c} & \ddots & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \overline{c} \end{bmatrix}$$

wobei zum 1. Kästchen die Basis $z_1,\ldots,z_q$, und zum 2. Kästchen die Basis $\overline{z}_1,\ldots,\overline{z}_q$ gehört, d.h. auch die Basen können konjugiert komplex gewählt werden. Setzen wir für $j = 1,\ldots,q$

$$x_j \;=\; \tfrac{1}{2}\,(z_j + \overline{z}_j)\;,\;\; y_j \;=\; -\tfrac{i}{2}\,(z_j - \overline{z}_j)\,,$$

so gilt im Vektorraum $\mathbb{C}^n$

$$[z_1,\ldots,z_q,\overline{z}_1,\ldots,\overline{z}_q] \;=\; [x_1,y_1,\ldots,x_q,y_q]\;,$$

also sind $x_1\,,\; y_1\,,\ldots,\; x_q\,,\; y_q$ linear unabhängig. Nach Konstruktion ist $x_j\,,\; y_j \in \mathbb{R}^n$ und für alle $j = 1,\ldots,q-1$ gilt:

$$A\,x_j \;=\; \tfrac{1}{2}\,(A\,z_j + A\,\overline{z}_j) \;=\; \tfrac{1}{2}\,(c\,z_j + z_{j+1} + \overline{c}\,\overline{z}_j + \overline{z}_{j+1})$$

$$= a\,\frac{z_j + \overline{z}_j}{2} - y\,\frac{(-i)(z_j - \overline{z}_j)}{2} + \frac{z_{j+1} + \overline{z}_{j+1}}{2} \;=\; a\,x_j - b\,y_j + x_{j+1}$$

und

$$A\,y_j \;=\; -\tfrac{i}{2}\,(A\,z_j - A\,\overline{z}_j) \;=\; a\,y_j + b\,x_j + y_{j+1}\,.$$

Für $j = q$ erhalten wir

$$A\,x_q \;=\; \tfrac{1}{2}\,(c\,z_q + \overline{c}\,\overline{z}_q) \;=\; a\,x_q - b\,y_q$$

und

$$A\,y_q \;=\; a\,y_q + b\,x_q\,.$$

Also ergibt sich für diesen Teil der Abbildungsmatrix bezüglich der neuen Basis $x_1\,,\,y_1\,,\ldots,\,x_q\,,\,y_q$

$$\begin{bmatrix} \begin{matrix} a & b \\ -b & a \end{matrix} & & & \\ & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} & \begin{matrix} a & b \\ -b & a \end{matrix} & \\ & & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} & \ddots \\ & & & \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \;\; \begin{matrix} a & b \\ -b & a \end{matrix} \end{bmatrix}$$

Wenden wir das Verfahren auf jedes Paar von Jordan–Kästchen der konjugiert komplexen Eigenwerte von A an, so erhalten wir das gewünschte Ergebnis.

Satz 22 (Reelle Normalform). *Es sei $A \in \mathbb{R}^{n\times n}$ mit charakteristischem Polynom*

$$p \;=\; (-1)^n\,(X-c_1)^{r_1}\cdots(X-c_k)^{r_k}\,(X^2 + a_1'\,X + b_1')^{t_1}\cdots(X^2 + a_m'\,X + b_m')^{t_m}\,,$$

wo $c_1,\ldots,c_k$ die paarweise verschiedenen Eigenwerte von A und $X^2 + a_1'\,X + b_1'$,..., $X^2 + a_m'\,X + b_m'$ paarweise verschiedene reelle Polynome ohne Nullstellen sind. Dann gibt es eine zu A ähnliche Matrix $\tilde{A}$ der Form

$$\tilde{A} \;=\; \begin{bmatrix} A_{c_1} & & & & \\ & \ddots & & O & \\ & & A_{c_k} & & \\ & & & B_1 & \\ & O & & & \ddots \\ & & & & B_m \end{bmatrix}$$

Dabei sind $A_{c_1},\ldots,A_{c_k}$ *die Jordan–Blöcke zu den Eigenwerten* $c_1,\ldots,c_k$ *und die Blöcke* B_l *haben die Form*

$$B_l = \begin{bmatrix} B_l^{(1)} & & O \\ & \ddots & \\ O & & B_l^{(n_l)} \end{bmatrix}, \quad l = 1,\ldots,m\,.$$

Hierbei hat das Kästchen $B_l^{(j)}$ *die Gestalt*

$$B_l^{(j)} = \begin{bmatrix} \begin{array}{cc} a_l & b_l \\ -b_l & a_l \end{array} & & & \\ & \begin{array}{cc|cc} 1 & 0 & a_l & b_l \\ 0 & 1 & -b_l & a_l \end{array} & & \\ & & \ddots & \\ & & & \begin{array}{cc|cc} 1 & 0 & a_l & b_l \\ 0 & 1 & -b_l & a_l \end{array} \end{bmatrix}, \quad j = 1,\ldots,n_l\,,$$

wo $a_l \pm i\, b_l$, $b_l > 0$, *die komplexen Nullstellen von* $X^2 + a_l' X + b_l'$ *sind.*

Bemerkungen. **(a)** Die Zahlen n_l, $l = 1,\ldots,m$, und die Größe der Kästchen $B_l^{(j)}$ ergeben sich aus der komplexen Jordanschen Normalform der Matrix A, wenn man diese als Element aus $\mathbb{C}^{n\times n}$ auffaßt.

n_l ist gerade die Zahl der Jordan–Kästchen zum Eigenwert $a_l + i\, b_l$ und zu jedem Jordan–Kästchen der Länge q zum Eigenwert $a_l + i\, b_l$ gibt es ein Kästchen $B_l^{(j)}$ der Länge $2\,q$.

(b) Die Basis, bezüglich der man die Normalform $\tilde{A}$ von A erhält, läßt sich mit dem oben beschriebenen Verfahren aus den konjugiert komplexen Basisvektoren in der Jordan–Basis der komplexen Matrix A bestimmen.

Beipiel. Gegeben sei die Matrix

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

mit dem charakteristischen Polynom

$$p = (X^2 - 2\,X + 2)^2 = (X - (1+i))^2\,(X - (1-i))^2\,.$$

Wir bestimmen zunächst die komplexe Jordansche Normalform A' von A. Die Nullstellen von $X^2 - 2\,X + 2$ sind $c_1 = 1 + i$ und $c_2 = 1 - i$. Der Eigenraum zu c_1 ergibt sich als Lösungsmenge des LGS $(A - (1+i)\,E_4)\,z = o$. Es ist

$$E_{1+i} = [\begin{bmatrix} 1 & -i \\ 1 & \\ -i & \\ 1 & -i \end{bmatrix}]\,.$$

Der Hauptraum zu c_1 ist Lösungsmenge des LGS $(A - (1+i)\,E_4)^2\,z = o$. Es ist

$$H_{1+i} = [\begin{bmatrix} i \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ i \\ 1 \\ 0 \end{bmatrix}]\,.$$

Im Hauptraum H_{1+i} bestimmen wir eine Jordan–Basis:

$$z_1 = \begin{bmatrix} i \\ 0 \\ 0 \\ 1 \end{bmatrix} \in H_{1+i} \setminus E_{1+i} \quad \text{und} \quad z_2 = (A - (1+i)\,E_4)\,z_1 = \begin{bmatrix} 1 & -i \\ 1 & \\ -i & \\ 1 & -i \end{bmatrix}.$$

Damit ergibt sich für den Eigenraum und den Hauptraum zur konjugiert komplexen Nullstelle $c_2 = 1 - i$ sowie für die Basis in diesem Hauptraum:

$$E_{1-i} = [\begin{bmatrix} 1+i \\ 1 \\ i \\ 1+i \end{bmatrix}] \quad \text{und} \quad H_{1-i} = [\begin{bmatrix} -i \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -i \\ 1 \\ 0 \end{bmatrix}]\,,$$

$$\overline{z}_1 = \begin{bmatrix} -i \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{und} \quad \overline{z}_2 = \begin{bmatrix} 1+i \\ 1 \\ i \\ 1+i \end{bmatrix}.$$

Die komplexe Jordansche Normalform und eine zugehörige Transformationsmatrix S lauten:

$$A' = \begin{bmatrix} 1+i & 0 & 0 & 0 \\ 1 & 1+i & 0 & 0 \\ 0 & 0 & 1-i & 0 \\ 0 & 0 & 1 & 1-i \end{bmatrix}, \quad S = \begin{bmatrix} i & 1-i & -i & 1+i \\ 0 & 1 & 0 & 1 \\ 0 & -i & 0 & i \\ 1 & 1-i & 1 & 1+i \end{bmatrix}$$

Damit ergeben sich die reelle Jordansche Normalform

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

und die reelle Transformationsmatrix

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

Kontrolle: $\tilde{A} = S^{-1} A S \iff S \tilde{A} = A S$:

$$S\tilde{A} = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & -1 \\ 2 & 0 & 2 & 0 \end{bmatrix}$$

$$AS = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & -2 & 1 \\ -1 & 2 & 2 & 0 \\ 0 & 0 & -2 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & -1 \\ 2 & 0 & 2 & 0 \end{bmatrix}$$

Kapitel 5 Euklidische und unitäre Vektorräume

Die im letzten Kapitel entwickelten Kriterien für die Diagonalisierbarkeit eines Endomorphismus Φ oder für die Herleitung der Jordanschen Normalform der Abbildungsmatrizen von Φ erlauben noch keine einfache geometrische Formulierung. Hierzu fehlen uns bisher die aus dem Anschauungsraum vertrauten Begriffe wie Länge, Abstand und Orthogonalität.

Wir wollen deshalb jetzt Vektorräume betrachten, in denen solche Begriffe definiert werden können. Eine sinnvolle Einführung des Längenbegriffs setzt die Vergleichbarkeit von Längen voraus. Dies ist in allgemeinen $\mathbb{K}$–Vektorräumen nicht möglich, daher wollen wir uns im folgenden zunächst auf reelle, später auf komplexe Vektorräume beschränken.

§ 1 Skalarprodukte

Es sei V ein reeller Vektorraum beliebiger Dimension. Wir betrachten eine Bilinearform β auf V, also eine $2-$fach multilineare Abbildung $\beta : V \times V \longrightarrow \mathbb{R}$.

β heißt *symmetrisch*, wenn $\beta(x,y) = \beta(y,x)$ für alle $x,y \in V$ gilt.

β heißt *positiv definit*, wenn $\beta(x,x) > 0$ für $x \neq o$ gilt ($\beta(o,o) = 0$ folgt aus der Bilinearität).

Definition. Eine positiv definite, symmetrische Bilinearform $\beta : V \times V \longrightarrow \mathbb{R}$ heißt *Skalarprodukt* oder *inneres Produkt* auf V. Für Skalarprodukte schreiben wir $<x,y>$ statt $\beta(x,y)$.

Das Paar $(V, <\cdot,\cdot>)$ oder kurz V heißt *euklidischer Vektorraum* oder auch *(reeller) Vektorraum mit Skalarprodukt.*

Beispiele. **(a)** $V = \mathbb{R}^n$, $<x,y> := \sum_{i=1}^{n} x_i\, y_i = x^T\, y$ ist Skalarprodukt. Es heißt das

Standardskalarprodukt . Im $\mathbb{R}^n$ werden wir mit $<\cdot,\cdot>$ immer dieses Standardskalarprodukt bezeichnen.

(b) $V = \mathbb{R}^2$, $\beta(x,y) := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2 x_2 y_2$ ist Skalarprodukt und offensichtlich verschieden vom Standardskalarprodukt $<\cdot,\cdot>$.

(c) Es sei $V = C([a,b])$ der Vektorraum der stetigen, reellen Funktionen auf dem Intervall $[a,b]$. Weiter sei s eine feste, positive, stetige Funktion auf $[a,b]$. Dann wird durch

$$<g,h> := \int_a^b g(t)\, h(t)\, s(t)\, dt$$

ein Skalarprodukt erklärt.

In einem euklidischen Vektorraum V können wir nun die gewünschten Begriffe Länge, Abstand und Orthogonalität einführen.

Definitionen und Bemerkungen. Es sei $(V,<\cdot,\cdot>)$ ein euklidischer Vektorraum.

(a) $\| x \| := \sqrt{<x,x>}$ heißt die *Länge* oder *Norm* von $x \in V$. Die Abbildung

$$\|\cdot\| : V \longrightarrow \mathbb{R}$$
$$x \longmapsto \| x \|$$

heißt *Norm* . Die Norm $\|\cdot\|$ hat, wie man unmittelbar aus der Definition erkennt, folgende Eigenschaften:

(i) $\| x \| \geq 0$ für alle $x \in V$ und $\| x \| = 0 \iff x = o$,

(ii) $\| c\, x \| = | c | \, \| x \|$ für alle $x \in V$ und alle $c \in \mathbb{R}$.

Eine dritte Eigenschaft, die Minkowski–Ungleichung, werden wir in dem folgenden Satz 1 beweisen.

(b) $d(x,y) := \| x - y \|$ heißt *Abstand* oder *Distanz* von $x,y \in V$. Die Abbildung

$$d : V \times V \longrightarrow \mathbb{R}$$
$$(x,y) \longmapsto d(x,y)$$

heißt *Metrik*. Für die Metrik d gilt:

(i) $d(x,y) \geq 0$ für alle $x,y \in V$ und $d(x,y) = 0 \iff x = y$,

(ii) $d(x,y) = d(y,x)$ für alle $x,y \in V$.

Weiterhin erfüllt die Metrik d die sogenannte Dreiecksungleichung, die wir ebenfalls in Satz 1 beweisen werden.

(c) x und y heißen *orthogonal*, wenn $\langle x,y\rangle = 0$ gilt; Schreibweise : $x \perp y$.

Sind $A,B \subset V$, so nennen wir A und B *orthogonal*, bzw. A *orthogonal* zu B bzw. B *orthogonal* zu A, wenn $x \perp y$ für alle $x \in A$, $y \in B$ gilt; Schreibweise : $A \perp B$. Statt $\{x\} \perp B$ schreiben wir auch kürzer $x \perp B$.

Für $A \subset V$ sei

$$A^{\perp} := \{x \in V \mid x \perp A\}.$$

$A^{\perp}$ heißt *orthogonales Komplement* von A.

Für alle Teilmengen $A \subset V$ gilt:

(i) $A^{\perp}$ ist ein Untervektorraum von V,

(ii) $A^{\perp} = [A]^{\perp}$,

(iii) $A^{\perp} \cap [A] = \{o\}$,

(iv) $[A] \subset (A^{\perp})^{\perp}$.

Satz 1. *Es seien V ein euklidischer Vektorraum und $x, y, z \in V$. Dann gilt:*

(a) $|\langle x,y\rangle| \leq \| x \| \cdot \| y \|$. *("Cauchy–Schwarzsche Ungleichung")*

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

(b) $\| x + y \| \leq \| x \| + \| y \|$. *("Minkowskische Ungleichung")*

(c) $d(x,y) \leq d(x,z) + d(z,y)$. *("Dreiecksungleichung")*

(d) $\| x + y \|^2 + \| x - y \|^2 = 2 \| x \|^2 + 2 \| y \|^2$. *("Parallelogrammidentität")*

(e) $4 \langle x,y\rangle = \| x + y \|^2 - \| x - y \|^2$.

(f) $x \perp y \iff \| x+y \|^2 = \| x \|^2 + \| y \|^2$. *("Satz von Pythagoras")*

Beweis. (a) Für $y = o$ gilt Gleichheit und x und y sind linear abhängig. Sei nun $y \neq o$. Für alle $t \in \mathbb{R}$ und alle $x, y \in V$ gilt

$$(*) \qquad 0 \leq \langle x + t\,y,\, x + t\,y \rangle = \langle x,x \rangle + 2\,t \langle x,y \rangle + t^2 \langle y,y \rangle .$$

Setzen wir speziell $t = -\frac{\langle x,y \rangle}{\langle y,y \rangle}$, so erhalten wir

$$0 \leq \langle x,x \rangle - \frac{\langle x,y \rangle^2}{\langle y,y \rangle} ,$$

woraus die behauptete Ungleichung folgt.

Gilt in der Cauchy–Schwarzschen Ungleichung das Gleichheitszeichen, so gilt in (*) für $t = -\frac{\langle x,y \rangle}{\langle y,y \rangle}$ das Gleichheitszeichen, woraus $x = -t\,y$ folgt. Die Umkehrung ist trivial.

(b) Mit (a) folgt

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

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

(c) $d(x,y) = \| x-y \| = \| x-z+z-y \| \leq \| x-z \| + \| z-y \| = d(x,z) + d(z,y)$.

(d) $$\| x+y \|^2 + \| x-y \|^2 = \langle x+y, x+y \rangle + \langle x-y, x-y \rangle$$

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

(e) $$\| x+y \|^2 - \| x-y \|^2 = 4 \langle x,y \rangle \quad (\text{wie in (d)}).$$

(f) $$x \perp y \iff \langle x,y \rangle = 0 \iff \| x+y \|^2 = \| x \|^2 + \| y \|^2$$ ∎

Bemerkungen. **(a)** Man kann auf einem reellen Vektorraum V eine *Norm* $\| \cdot \|$ einführen als Abbildung $\| \cdot \| : V \longrightarrow \mathbb{R}$, die folgende Bedingungen erfüllt:

(i) Für alle $x \in V$ gilt: $\| x \| \geq 0$ und $\| x \| = 0 \iff x = o$.

(ii) Für alle $x \in V$ und alle $c \in \mathbb{R}$ gilt: $\| c\,x \| = | c | \, \| x \|$.

(iii) Für alle $x, y \in V$ gilt: $\| x+y \| \leq \| x \| + \| y \|$.

$(V,\|\cdot\|)$ heißt dann *normierter Raum* .

Es gibt normierte Räume, bei denen die Norm nicht von einem Skalarprodukt $<\cdot,\cdot>$ induziert wird. Eine Norm wird genau dann von einem Skalarprodukt induziert, d.h. es gilt für alle $x \in V$, daß $\| x \| = \sqrt{<x,x>}$, wenn die Parallelogrammidentität (d) erfüllt ist. (Übungsaufgabe). Die Aussage (e) zeigt, daß in diesem Fall das Skalarprodukt eindeutig durch die Norm $\|\cdot\|$ bestimmt ist.

(b) Man kann auf einem Vektorraum V oder allgemeiner auf einer beliebigen Menge $V \neq \emptyset$ eine *Metrik* d einführen als Abbildung $d : V \times V \longrightarrow \mathbb{R}$, die den folgenden drei Bedingungen genügt:

(i) Für alle $x,y \in V$ gilt: $d(x,y) \geq 0$ und $d(x,y) = 0 \iff x = y$.

(ii) Für alle $x,y \in V$ gilt: $d(x,y) = d(y,x)$.

(iii) Für alle $x,y,z \in V$ gilt: $d(x,y) \leq d(x,z) + d(z,y)$.

(V,d) heißt dann *metrischer Raum.*

Es gibt metrische Räume, die keine Vektorräume sind, und selbst wenn V ein reeller Vektorraum ist, muß die Metrik d nicht von einer Norm und damit erst recht nicht von einem Skalarprodukt erzeugt werden.

(c) Die Cauchy–Schwarzsche Ungleichung gibt uns jetzt die Möglichkeit, in euklidischen Vektorräumen den Winkel ω zwischen zwei Vektoren x und y zu erklären. Dabei setzen wir sinnvollerweise $x \neq o$ und $y \neq o$ voraus. Dann ist

$$-1 \leq \frac{<x,y>}{\|x\| \, \|y\|} \leq 1 .$$

Damit existiert genau ein $\omega \in [0,\pi]$ mit

$$\cos \omega = \frac{<x,y>}{\|x\| \, \|y\|} .$$

Diese Zahl ω heißt *Winkel* zwischen x und y.

In der linearen Algebra ist der Winkelbegriff nicht so wichtig wie der Begriff der Orthogonalität. Wir werden deshalb im folgenden hauptsächlich letzteren untersuchen.

Mit Hilfe von Satz 1 wollen wir noch ein weiteres, wichtiges Beispiel für einen euklidischen Vektorraum angeben.

Beispiel. Es sei $V \subset \mathbb{R}^{\mathbb{N}_0}$ die Menge der reellen Folgen $x = (x_0, x_1, \ldots)$, die quadratisch summierbar sind, für die also $\sum_{i=0}^{\infty} x_i^2 < \infty$ gilt.

V ist ein Untervektorraum: Seien $x \in V$, $a \in \mathbb{R}$. Dann ist $a\,x = (a\,x_0, a\,x_1, \ldots)$, und es gilt

$$\sum_{i=0}^{\infty} a^2\, x_i^2 < \infty ,$$

also ist $a\,x \in V$.

Aus $x = (x_0, x_1, \ldots)$, $y = (y_0, y_1, \ldots) \in V$ folgt mit der Minkowskischen Ungleichung für den Standardraum $\mathbb{R}^{m+1}$

$$\Big(\sum_{i=0}^{m} (x_i + y_i)^2\Big)^{\frac{1}{2}} \le \Big(\sum_{i=0}^{m} x_i^2\Big)^{\frac{1}{2}} + \Big(\sum_{i=0}^{m} y_i^2\Big)^{\frac{1}{2}} \le \Big(\sum_{i=0}^{\infty} x_i^2\Big)^{\frac{1}{2}} + \Big(\sum_{i=0}^{\infty} y_i^2\Big)^{\frac{1}{2}} =: c$$

mit $c \in \mathbb{R}$. Damit gilt für alle $m \in \mathbb{N}_0$

$$\sum_{i=0}^{m} (x_i + y_i)^2 \le c^2$$

und somit

$$\sum_{i=0}^{\infty} (x_i + y_i)^2 < \infty ,$$

also $x + y \in V$.

Für $x, y \in V$ setzen wir nun

$$<x,y> := \sum_{i=0}^{\infty} x_i\, y_i\,.$$

Diese Reihe ist absolut konvergent, denn mit der Cauchy–Schwarzschen Ungleichung, angewendet im $\mathbb{R}^{m+1}$ erhalten wir für alle $m \in \mathbb{N}_0$

$$\sum_{i=0}^{m} |x_i\, y_i| \;=\; \sum_{i=0}^{m} |x_i| \cdot |y_i| \;\leq\; \Big(\sum_{i=0}^{m} x_i^2\Big)^{\frac{1}{2}} \cdot \Big(\sum_{i=0}^{m} y_i^2\Big)^{\frac{1}{2}}$$

$$\leq \Big(\sum_{i=0}^{\infty} x_i^2\Big)^{\frac{1}{2}} \cdot \Big(\sum_{i=0}^{\infty} y_i^2\Big)^{\frac{1}{2}} =: c$$

mit $c \in \mathbb{R}$ und somit

$$\sum_{i=0}^{\infty} |x_i\, y_i| \;\leq\; c\,.$$

Also ist die obige Reihe absolut konvergent und daher auch konvergent.

Damit ist $<\cdot,\cdot>$ offensichtlich ein Skalarprodukt auf V. Der Vektorraum V ist ein wichtiger unendlich dimensionaler euklidischer Vektorraum in der Mathematik. Er heißt *Standard–Hilbertraum* und wird üblicherweise mit l^2 bezeichnet. Der Vektorraum $\mathbb{R}[X]$ der reellen Polynome ist ein Untervektorraum von l^2.

Zum Abschluß dieses Paragraphen wollen wir überlegen, wie man Skalarprodukte auf endlich dimensionalen reellen Vektorräumen V bestimmen kann.

Es sei jetzt also $\dim V = n$ und $B = (v_1,\ldots,v_n)$ sei eine festgewählte Basis von V. Ist $\beta : V \times V \longrightarrow \mathbb{R}$ ein Skalarprodukt, so gilt wegen der Bilinearität von β für alle $x = \sum_{i=1}^{n} x_i\, v_i\,,\ x_i \in \mathbb{R}$, und alle $y = \sum_{j=1}^{n} y_j\, v_j\,,\ y_j \in \mathbb{R}$

$$\beta\,(x,y) \;=\; \sum_{i\,,\;j=1}^{n} x_i\, \beta\,(v_i,v_j)\, y_j \;=\; \hat{x}^{\mathrm{T}} A\, \hat{y}\,,$$

wo $\hat{x}$ und $\hat{y}$ die Koordinatenvektoren von x und y sind und $A = (\!(\beta\,(v_i,v_j))\!)$ gesetzt ist.

β ist also eindeutig festgelegt durch die Matrix A . Weil β symmetrisch ist, ist auch A symmetrisch. Weil β positiv definit ist, gilt für alle $\hat{x} \in \mathbb{R}^n$, $\hat{x} \neq o$,

$$(*) \qquad \hat{x}^{\mathsf{T}} A\, \hat{x} > 0\,.$$

Eine symmetrische Matrix A, die (*) erfüllt, wollen wir ebenfalls *positiv definit* nennen.

Es sei nun umgekehrt $A \in \mathbb{R}^{n\times n}$ eine symmetrische, positiv definite Matrix. Dann wird durch

$$\beta\,(x,y) \;=\; \hat{x}^{\mathsf{T}} A\, \hat{y}\,,$$

wo $\hat{x}$ und $\hat{y}$ die Koordinatendarstellungen von x und y bezüglich der festen Basis B sind, offensichtlich ein Skalarprodukt auf V erklärt. Wir haben also den folgenden Zusammenhang gezeigt:

Satz 2. *Es seien V ein reeller Vektorraum und B eine geordnete Basis von V. Dann ist $\beta : V \times V \longrightarrow \mathbb{R}$ genau dann ein Skalarprodukt, wenn eine symmetrische, positiv definite Matrix $A \in \mathbb{R}^{n\times n}$ existiert mit*

$$\beta\,(x,y) \;=\; \hat{x}^{\mathsf{T}} A\, \hat{y}\,, \qquad x,y \in V\,.$$

Dabei sind $\hat{x}$ und $\hat{y}$ die Koordinatenvektoren von x und y bezüglich B.

Es bleibt noch die Frage zu klären, wie man die positive Definitheit einer symmetrischen Matrix A nachprüfen kann. Wir zeigen dazu zwei Kriterien, benötigen aber zuvor eine Diagonalisierbarkeitsaussage für symmetrische Matrizen.

Satz 3. *Es sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Dann gilt:*

(a) *A ist diagonalisierbar.*

(b) *Eigenvektoren von A zu verschiedenen Eigenwerten sind orthogonal (bezüglich des Standardskalarproduktes im $\mathbb{R}^n$).*

Beweis. (a) Wir fassen A zunächst als komplexe Matrix auf. Sei $c = a + i\,b$, $a,b \in \mathbb{R}$,

komplexe Nullstelle des charakteristischen Polynoms p von A und $z = u + i\,v$ ein zugehöriger Eigenvektor aus $\mathbb{C}^n$, $A\,z = c\,z$. Dann ist

$$A\,u = a\,u - b\,v\ ,\ A\,v = b\,u + a\,v\ .$$

Damit folgt

$$(A\,u)^{\mathsf{T}}\,v = a\,u^{\mathsf{T}}v - b\,v^{\mathsf{T}}v$$

und

$$(A\,u)^{\mathsf{T}}\,v = u^{\mathsf{T}}A\,v = b\,u^{\mathsf{T}}u + a\,u^{\mathsf{T}}v\ .$$

Also ist $b \parallel u \parallel^2 = -\,b \parallel v \parallel^2$, woraus wegen $z \neq o$ unmittelbar $b = 0$ folgt. Das charakteristische Polynom p zerfällt also in Linearfaktoren mit lauter reellen Nullstellen $c_1,\ldots,c_k$, woraus $\mathbb{R}^n = H_{c_1} \oplus \cdots \oplus H_{c_k}$ und $m = (X - c_1)^{s_1} \cdots (X - c_k)^{s_k}$, s_i Index von H_{c_i}, folgen.

Wir zeigen jetzt, daß der Index s_i von H_{c_i} eins ist: Sei $(A - c_i\,E_n)^2\,y = o$. Dann gilt

$$0 = y^{\mathsf{T}}(A - c_i\,E_n)^2\,y = [(A - c_i\,E_n)\,y]^{\mathsf{T}}[(A - c_i\,E_n)\,y]\ ,$$

also $(A - c_i\,E_n)\,y = o$ und somit $s_i = 1$. Nach Satz 4.19 ist A daher diagonalisierbar.

(b) Aus $A\,u = c\,u$ und $A\,v = d\,v$ mit $u \neq o$, $v \neq o$ und $c,d \in \mathbb{R}$, $c \neq d$, folgt

$$c\,u^{\mathsf{T}}v = (A\,u)^{\mathsf{T}}\,v = u^{\mathsf{T}}A\,v = d\,u^{\mathsf{T}}v$$

und daraus $\langle u,v\rangle = u^{\mathsf{T}}v = 0$. ∎

Satz 4. *Es sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix. Dann ist A genau dann positiv definit, wenn alle Eigenwerte von A positiv sind.*

Beweis. Es seien A positiv definit, $c \in \mathbb{R}$ Eigenwert und x Eigenvektor von A. Dann gilt wegen $x \neq o$

$$0 < x^{\mathsf{T}}A\,x = x^{\mathsf{T}}(c\,x) = c\,x^{\mathsf{T}}x = c\,(x_1^2 + \cdots + x_n^2)\ ,$$

also folgt $c > 0$.

Seien umgekehrt $c_1,...,c_k$ die Eigenwerte von A und es gelte $c_i > 0$ für alle $i = 1,...,k$. Nach Satz 3 (a) ist A diagonalisierbar, d.h. es gilt $\mathbb{R}^n = E_{c_1} \oplus \cdots \oplus E_{c_k}$. Zu jedem $x \in \mathbb{R}^n$ gibt es also eine eindeutige Zerlegung $x = x_1 + \cdots + x_k$ mit $x_i \in E_{c_i}$. Damit folgt

$$x^T A\, x \;=\; \sum_{i=1}^{k} x_i^T A\, x_i \;+\; \sum_{\substack{i,\, j=1 \\ i \neq j}}^{k} x_i^T A\, x_j \;=\; \sum_{i=1}^{k} c_i \| x_i \|^2 \;+\; \sum_{\substack{i,\, j=1 \\ i \neq j}}^{k} c_j \langle x_i, x_j \rangle \;,$$

also wegen Satz 3 (b)

$$x^T A\, x \;=\; \sum_{i=1}^{k} c_i \| x_i \|^2 \,.$$

Nun ist für $x \neq o$ mindestens ein $x_i \neq o$. Wegen $c_i > 0$ für $i = 1,...,k$ folgt also insgesamt, daß $x^T A\, x > 0$. ∎

Beispiele. **(a)** Gegeben sei die Matrix

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

mit dem charakteristischen Polynom

$$p = -(X-3)\,(X - \tfrac{1}{2}(3+\sqrt{5}))\,(X - \tfrac{1}{2}(3-\sqrt{5}))\,.$$

Alle Eigenwerte sind positiv, also ist A positiv definit.

(b) Die Matrix

$$A \;=\; \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

mit dem charakteristischen Polynom $p = X^2\,(X-2)^2$ besitzt den Eigenwert 0, ist also nicht positiv definit.

Oft ist es schwierig, die Eigenwerte einer Matrix A auszurechnen. Dann ist eventuell das folgende Kriterium nützlicher.

Satz 5. *Für eine symmetrische Matrix* $A = (\!(a_{ij})\!) \in \mathbb{R}^{n\times n}$ *sind folgende Aussagen äquivalent:*

(a) *A ist positiv definit.*

(b) *Es existiert eine reguläre Matrix* $B \in \mathbb{R}^{n\times n}$ *mit* $A = B^{\mathsf{T}} B$.

(c) *Alle Hauptunterdeterminanten von A sind positiv, d.h. für* $k = 1,\ldots,n$ *gilt*

$$\begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} > 0 .$$

Bemerkung. Man nennt die Determinanten in (c) auch die *Hauptminoren* der Matrix A.

Beweis. (a) $\Longrightarrow$ (c): Ist A positiv definit, so ist auch jede der Matrizen

$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix} , \quad k = 1,\ldots,n,$$

positiv definit ($A_n = A$). Ist nämlich $(x_1,\ldots,x_k) \neq (0,\ldots,0)$, so ist $x = (x_1,\ldots,x_k,0,\ldots,0) \neq o$, und es gilt

$$(x_1 \cdots x_k)\, A_k \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = x^{\mathsf{T}} A\, x > 0 .$$

Nach Satz 3 (a) ist jede der Matrizen A_k ähnlich zu einer Diagonalmatrix D_k :

$$A_k = S_k^{-1} D_k S_k .$$

Nach Satz 4 sind die Eigenwerte von A_k alle positiv, also gilt für $k = 1,\ldots,n$

$$\det A_k = \det D_k > 0 .$$

(c) $\Longrightarrow$ (b): Für $k = 2,\ldots,n$ schreiben wir die Matrizen A_k in der Form

$$A_k = \begin{bmatrix} A_{k-1} & y_k \\ y_k^{\mathsf{T}} & a_{kk} \end{bmatrix}$$

mit $y_k^{\mathsf{T}} = (a_{k1} \ \ldots \ a_{k,k-1})$. Weiterhin sei für $k = 2,\ldots,n$

$$T_k = \begin{bmatrix} E_{k-1} & -A_{k-1}^{-1} y_k \\ o^{\mathsf{T}} & 1 \end{bmatrix}$$

Hierbei haben wir benutzt, daß A_{k-1} wegen $\det A_{k-1} > 0$ invertierbar ist. Es folgt

$$T_k^{\mathsf{T}} A_k T_k = \begin{bmatrix} A_{k-1} & o \\ o^{\mathsf{T}} & a_{kk} - y_k^{\mathsf{T}} A_{k-1}^{-1} y_k \end{bmatrix}$$

Sei $b_k := a_{kk} - y_k^{\mathsf{T}} A_{k-1}^{-1} y_k$. Wegen $\det T_k = \det T_k^{\mathsf{T}} = 1$ folgt

$$\det A_k = \det (T_k^{\mathsf{T}} A_k T_k) = \det A_{k-1} \cdot b_k ,$$

also

$$b_k = \frac{\det A_k}{\det A_{k-1}} > 0 , \qquad k = 2,\ldots,n .$$

Für die (n,n)–Matrix

$$T := T_n \begin{bmatrix} T_{n-1} & O \\ O & 1 \end{bmatrix} \begin{bmatrix} T_{n-2} & O \\ O & E_2 \end{bmatrix} \cdots \begin{bmatrix} T_2 & O \\ O & E_{n-2} \end{bmatrix}$$

gilt nun nach Konstruktion, daß sie regulär ist und daß

$$T^{\mathsf{T}} A T = \begin{bmatrix} a_{11} & & & O \\ & b_2 & & \\ & & \ddots & \\ O & & & b_n \end{bmatrix}$$

Setzen wir noch

$$D = \begin{bmatrix} \sqrt{a_{11}} & & & O \\ & \sqrt{b_2} & & \\ & & \ddots & \\ O & & & \sqrt{b_n} \end{bmatrix}$$

und $B = D\,T^{-1}$, so folgt $B^{\mathsf{T}} B = (T^{-1})^{\mathsf{T}} D^2 T^{-1} = A$ und $B \in \mathbb{R}^{n\times n}$ ist regulär.

(b) $\Longrightarrow$ (a): Sei $x \neq o$. Dann ist $B x \neq o$ und es gilt $x^{\mathsf{T}} A x = x^{\mathsf{T}} B^{\mathsf{T}} B x = (B x)^{\mathsf{T}} (B x) > 0$. Also ist A positiv definit. ∎

Bemerkung. Der Beweis zeigt, daß die reguläre Matrix B sogar eine obere Dreiecksmatrix ist. Die Zerlegung der positiv definiten Matrix A in der Form $A = B^{\mathsf{T}}B$ mit einer regulären Dreicksmatrix B heißt *Cholesky– Zerlegung*, sie spielt eine wichtige Rolle in der numerischen Mathematik.

Beispiel. Für welche $a \in \mathbb{R}$ ist

$$A = \begin{bmatrix} 1 & a & -1 \\ a & 9 & 0 \\ -1 & 0 & 4 \end{bmatrix}$$

positiv definit? Es gilt $p = -X^3 + 14X^2 - (48 - a^2)X + 27 - 4a^2$. Hier die Nullstellen in Abhängigkeit von a bestimmen zu wollen, ist sicher nicht einfach. Wir wenden deshalb Satz 5 an und erhalten

$$\det A_1 = 1 > 0\,,\ \det A_2 = 9 - a^2 > 0 \iff |a| < 3\,,$$

$$\det A_3 = \det A = 27 - 4a^2 > 0 \iff |a| < \tfrac{3}{2}\sqrt{3}\,.$$

Somit ist A genau dann positiv definit, wenn $-\frac{3}{2}\sqrt{3} < a < \frac{3}{2}\sqrt{3}$ gilt.

§ 2 Orthonormalbasen und Orthogonalprojektionen

Da wir in euklidischen Vektorräumen die Orthogonalität von Vektoren eingeführt haben, liegt es nahe, nach Basen zu fragen, deren Vektoren paarweise orthogonal sind. Solche "rechtwinkligen Koordinatensysteme" spielen ja auch im Anschauungsraum eine wichtige Rolle bei der Beschreibung geometrischer Phänomene.

Definition. Es sei V ein euklidischer Vektorraum. Eine nichtleere Teilmenge $A \subset V$ heißt *Orthogonalsystem*, wenn gilt:

(a) $o \notin A$,

(b) Je zwei verschiedene Vektoren $x,y \in A$ sind orthogonal.

Gilt außerdem noch

(c) $\| x \| = 1$ für alle $x \in A$,

so heißt A *Orthonormalsystem*. Ist zusätzlich A Basis von V, so heißt A *Orthogonal–* bzw. *Orthonormalbasis* von V. Für letztere schreiben wir kurz ONB.

Bemerkung. Eine endliche Basis $B = (x_1,...,x_n)$ bzw. eine abzählbare Basis $B = (x_1,x_2,...)$ eines euklidischen Vektorraums ist genau dann eine ONB, wenn für alle $i,j \in \{1,...,n\}$ bzw. für alle $i,j \in \mathbb{N}$ gilt: $<x_i,x_j> = \delta_{ij}$. Wie hier werden wir auch in Zukunft bei endlichen bzw. abzählbaren Basen meistens davon ausgehen, daß sie geordnet sind.

Satz 6. *Es seien V ein euklidischer Vektorraum und $A \subset V$ ein Orthogonalsystem. Dann ist A linear unabhängig.*

Beweis. Seien $x_1,...,x_k \in A$ paarweise verschieden und $a_1 x_1 + \cdots + a_k x_k = o$, $a_i \in \mathbb{R}$, $i = 1,...,k$. Dann folgt für $j = 1,...,k$:

$$0 = <a_1 x_1 + \cdots + a_k x_k , x_j> = a_1 <x_1,x_j> + \cdots + a_k <x_k,x_j> = a_j \| x_j \|^2 ,$$

also $a_j = 0$. Somit ist A linear unabhängig. ∎

Ein Orthogonalsystem ist also genau dann eine Orthogonalbasis, wenn es V erzeugt. In unendlich dimensionalen Vektorräumen kann es Orthogonalsysteme geben, die keine Basen sind, aber dennoch maximal sind, d.h. nicht durch Hinzufügen von weiteren Vektoren vergrößert werden können. In solchen euklidischen Vektorräumen gibt es keine Orthonormalbasis. Für endlich dimensionale euklidische Vektorräume kann dieser Fall jedoch nicht eintreten, wie der folgende Satz zeigt.

Satz 7. *Es sei V ein endlich dimensionaler euklidischer Vektorraum. Dann existiert eine Orthonormalbasis B von V.*

Bemerkung. Wir werden nicht nur die Existenz von B nachweisen, sondern auch ein Konstruktionsverfahren angeben, mit dem man aus einer beliebigen geordneten Basis B' von V eine ONB erhalten kann. Das Verfahren heißt *Gram–Schmidtsches Orthogonalisierungsverfahren.*

Beweis. Sei dim $V = n$ und $B' = (x_1,...,x_n)$ eine Basis von V. Wir definieren zunächst induktiv eine Orthogonalbasis $\overline{B} = (y_1,...,y_n)$ von V und normieren die Basisvektoren anschließend. Sei

$$y_1 := x_1 ,$$

$$y_2 := x_2 - \frac{<x_2,y_1>}{\| y_1 \|^2} y_1 ,$$

$$\vdots$$

$$y_{k+1} := x_{k+1} - \sum_{i=1}^{k} \frac{<x_{k+1},y_i>}{\| y_i \|^2} y_i , \qquad k = 1,...,n-1.$$

Diese Definition ist sinnvoll, da alle y_i von o verschieden sind. Offensichtlich ist $y_1 \neq o$. Wäre $y_{k+1} = o$ für ein $k \in \{1,...,n-1\}$, so wäre $x_{k+1} \in [y_1,...,y_k] \subset [x_1,...,x_k]$, ein Widerspruch.

Durch vollständige Induktion nach k zeigen wir nun, daß $(y_1,...,y_{k+1})$ ein Orthogonalsystem ist.

$$k = 1: \quad <y_1,y_2> = <x_1,x_2> - \frac{<x_2,x_1>}{<x_1,x_1>} <x_1,x_1> = 0 .$$

$k \longrightarrow k+1$: Nach Induktionsvoraussetzung sind die Vektoren $y_1,\ldots,y_k$ paarweise orthogonal, also müssen wir nur noch $<y_{k+1},y_j> = 0$ für $j = 1,\ldots,k$ nachweisen:

$$<y_{k+1},y_j> = <x_{k+1},y_j> - \sum_{i=1}^{k} \frac{<x_{k+1},y_i>}{\| y_i \|^2} <y_i,y_j>$$

$$= <x_{k+1},y_j> - \frac{<x_{k+1},y_j>}{\| y_j \|^2} <y_j,y_j> = 0 .$$

Also ist $\overline{B} = (y_1,\ldots,y_n)$ ein Orthogonalsystem und daher nach Satz 5 linear unabhängig. Wegen $x_j \in [y_1,\ldots,y_j]$, $j = 1,\ldots,n$, ist $[\overline{B}] = V$. Daher ist $\overline{B}$ Orthogonalbasis und somit ist

$$B = \left(\frac{y_1}{\|y_1\|},\ldots,\frac{y_n}{\|y_n\|}\right)$$

eine ONB von V. ∎

Bemerkungen. **(a)** Der Beweis zeigt, daß das Orthogonalisierungsverfahren auch dann funktioniert, wenn V unendlich dimensional ist und in V eine abzählbare Basis $B' = (x_1,x_2,\ldots)$ existiert.

(b) Die Aussage von Satz 7 folgt auch aus Satz 5 :

Beweis. Sei o.B.d.A. $V = \mathbb{R}^n$, und sei β ein Skalarprodukt auf $\mathbb{R}^n$. Bezüglich der Standardbasis gilt dann für alle $x,y \in \mathbb{R}^n$

$$\beta(x,y) = x^T A\, y$$

mit einer symmetrischen positiv definiten Matrix A. Nach Satz 5 existiert eine reguläre Matrix $B \in \mathbb{R}^{n\times n}$ mit $A = B^T B$. Somit gilt $\beta(x,y) = (B\,x)^T\,(B\,y)$, also ist $(B^{-1}e_1,\ldots,B^{-1}e_n)$ eine Orthonormalbasis bezüglich β . ∎

Allerdings ergibt sich so kein einfaches Rechenverfahren zur Bestimmung einer Orthonormalbasis.

Beispiele. **(a)** Im $\mathbb{R}^n$ ist die Standardbasis $B = (e_1,\ldots,e_n)$ ONB bezüglich des Standardskalarprodukts $<\cdot,\cdot>$.

(b) In dem Vektorraum $\mathbb{R}[X]$, aufgefaßt als Untervektorraum des Hilbertraumes l^2, ist $A = \{e_i \mid i \in \mathbb{N}_0\}$ mit

$$e_i = (0,\ldots,0,\underbrace{1}_{\text{i-te Stelle}},0,\ldots)$$

eine ONB. In l^2 ist A ein Orthonormalsystem.

(c) Wir wollen im $\mathbb{R}^3$ von der Basis

$$B' = (\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix})$$

ausgehend eine ONB bezüglich des Standardskalarproduktes bestimmen:

$$y_1 = x_1 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \quad \| y_1 \| = \sqrt{2},$$

$$y_2 = x_2 - \frac{<x_2,y_1>}{\| y_1 \|^2} y_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} - \frac{2}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}, \quad \| y_2 \| = 1,$$

$$y_3 = x_3 - \frac{<x_3,y_1>}{\| y_1 \|^2} y_1 - \frac{<x_3,y_2>}{\| y_2 \|^2} y_2 = \begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} - \frac{1}{1}\begin{bmatrix}0\\0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\-1\\0\end{bmatrix},$$

$$\| y_3 \| = \frac{1}{2}\sqrt{2}.$$

Also ist

$$B = (\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix})$$

eine Orthonormalbasis von $\mathbb{R}^3$.

(d) Wir rechnen Beispiel (c) nochmals, wobei jetzt aber der $\mathbb{R}^3$ mit dem durch

$$\beta(x,y) = x^T A\, y,$$

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

definierten Skalarprodukt β versehen ist:

$$y_1 = x_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \| y_1 \|^2 = \beta(y_1, y_1) = 4,$$

$$y_2 = x_2 - \frac{\beta(x_2, y_1)}{\| y_1 \|^2} y_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}, \quad \| y_2 \|^2 = \frac{7}{4},$$

$$y_3 = x_3 - \frac{\beta(x_3, y_1)}{\| y_1 \|^2} y_1 - \frac{\beta(x_3, y_2)}{\| y_2 \|^2} y_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{0}{4} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{7} \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 6 \\ -1 \\ 3 \end{bmatrix},$$

$$\| y_3 \|^2 = \frac{3}{7}$$

Also ist

$$B = \left(\frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \frac{1}{2\sqrt{7}} \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}, \frac{1}{\sqrt{21}} \begin{bmatrix} 6 \\ -1 \\ 3 \end{bmatrix} \right)$$

bezüglich β eine ONB von $\mathbb{R}^3$.

Bemerkungen. **(a)** Ist $B = (x_1, \ldots, x_n)$ eine ONB von V, so gilt für alle $x \in V$

$$x = \sum_{i=1}^{n} \langle x, x_i \rangle \, x_i,$$

denn aus $x = a_1 x_1 + \cdots + a_n x_n$ folgt $\langle x, x_j \rangle = \sum_{i=1}^{n} a_i \langle x_i, x_j \rangle = a_j$ für alle $j = 1, \ldots, n$.

Damit folgt weiter für alle $x, y \in V$

$$\langle x, y \rangle = \sum_{i=1}^{n} \langle x, x_i \rangle \langle y, x_i \rangle,$$

$<x,y>$ stimmt also mit dem Standardskalarprodukt $\hat{x}^T \hat{y}$ der Koordinatenvektoren $\hat{x}$ und $\hat{y}$ überein.

Daraus ergibt sich für alle $x \in V$

$$\| x \|^2 = \sum_{i=1}^{n} <x,x_i>^2 = \| \hat{x} \|^2 .$$

(b) Ist V ein euklidischer Vektorraum beliebiger Dimension und ist $A \subset V$ Orthonormalsystem, so gilt für alle $x \in V$

$$(*) \qquad \| x \|^2 \geq \sum_{y \in A} <x,y>^2 , \qquad (\textit{"Besselsche Ungleichung"})$$

wobei in der Summe jeweils höchstens abzählbar viele Summanden positiv sind (Beweis als Übungsaufgabe).

Statt nach einer ONB, die es im allgemeinen nicht gibt, sucht man dann nach Orthonormalsystemen A, die *vollständig* sind, d.h. für die in (*) Gleichheit gilt für alle $x \in V$ (*"Parsevalsche Gleichung"*).

Jedes vollständige Orthonormalsystem A ist maximal, denn aus $x \perp A$ folgt wegen der Parsevalschen Gleichung $\| x \|^2 = 0$, also $x = o$.

Beispiel. Sei $V = l^2$. Hier ist $A = \{e_i | \ i \in \mathbb{N}_0\}$ mit

$$e_i = (0,\ldots,0,\underbrace{1}_{\text{i-te Stelle}},0,\ldots)$$

ein Orthonormalsystem. Es ist wegen $[A] = \mathbb{R}[X]$ keine Basis von l^2, aber es ist vollständig, denn für alle $x \in l^2$ gilt:

$$\| x \|^2 = \sum_{i=0}^{\infty} x_i^2 = \sum_{i=0}^{\infty} <x,e_i>^2$$

Es seien nun V ein euklidischer Vektorraum, $U \subset V$ ein endlich dimensionaler Untervektorraum von V und $B = (x_1,\ldots,x_k)$ eine ONB von U. Dann gilt für alle $u \in U$

$$u = \sum_{i=1}^{k} <u,x_i> x_i .$$

Setzen wir also für $x \in V$

$$\pi(x) := \sum_{i=1}^{k} <x,x_i> x_i ,$$

so erhalten wir eine lineare Abbildung $\pi : V \longrightarrow V$ mit Bild $\pi = U$ und $\pi|_U = \mathrm{id}_U$, d.h. $\pi^2 = \pi$. Somit ist π eine Projektion auf den Untervektorraum $U = \text{Bild } \pi$. Weiterhin gilt

$$<\pi(x) - x\ ,\ \pi(x)> = <\pi(x),\pi(x)> - <\pi(x),x>$$

$$= \sum_{i=1}^{k} <x,x_i>^2 - \sum_{i=1}^{k} <x,x_i> <x,x_i> = 0,$$

also ist

$$(\pi(x) - x) \perp \pi(x) .$$

Eine solche Abbildung π wollen wir Orthogonalprojektion nennen.

Definition. Es seien V ein euklidischer Vektorraum und $\pi \in \mathrm{Hom}\,(V,V)$. Die Abbildung π heißt *Orthogonalprojektion* (auf $U = \text{Bild } \pi$), wenn π Projektion ist, also $\pi^2 = \pi$ erfüllt, und wenn

$$(\pi(x) - x) \perp \pi(x)$$

für alle $x \in V$ gilt. Der Vektor $\pi(x)$ heißt dann die *Orthogonalprojektion* von x auf U.

Bemerkung. Ist π Orthogonalprojektion auf U, so gilt $(\pi(x) - x) \perp U$ und somit auch Kern $\pi \perp$ Bild π. Umgekehrt ist jede Projektion π, die Kern $\pi \perp$ Bild π erfüllt, eine Orthogonalprojektion.

Beweis. Es sei $u \in U$. Dann gilt $\pi(u - \pi(x) + x) = \pi(u) = u$ und daher $\pi(u - \pi(x) + x) - (u - \pi(x) + x) \perp u$. Also folgt $(\pi(x) - x) \perp u$.

Aus $(\pi(x) - x) \perp U$ erhalten wir unmittelbar Kern $\pi \perp U$, und umgekehrt ist jede Projektion π auf U mit dieser Eigenschaft wegen $\pi(x) - x \in$ Kern π eine Orthogonalprojektion. ∎

Nicht jeder Untervektorraum U besitzt eine Orthogonalprojektion. Wann dies der Fall ist, besagt der folgende Satz.

Satz 8. *Es seien V ein euklidischer Vektorraum und U ein Untervektorraum von V. Genau dann existiert eine Orthogonalprojektion auf U, wenn $V = U \oplus U^{\perp}$ gilt. Ist dies der Fall, so gibt es genau eine Orthogonalprojektion auf U und es gilt $U = (U^{\perp})^{\perp}$.*

Beweis. Sei π eine Orthogonalprojektion auf U. Für $x \in V$ gilt dann

$$x = \pi(x) + (x - \pi(x)) \in U + U^{\perp}.$$

Somit ist $V = U + U^{\perp}$, und wegen $U \cap U^{\perp} = \{o\}$ ist diese Summe direkt.

Sei nun umgekehrt $V = U \oplus U^{\perp}$. Jeder Vektor $x \in V$ besitzt somit eine eindeutige Zerlegung $x = u + v$, $u \in U$, $v \in U^{\perp}$. Die durch $\pi(x) = u$ erklärte Abbildung π ist dann eine Projektion von V auf U und es gilt $<\pi(x) - x, \pi(x)> = -<v,u> = 0$. Also ist π Orthogonalprojektion auf U.

Eindeutigkeit von π : Ist $u = \pi(x)$, und ist $u' = \pi'(x)$ ebenfalls Orthogonalprojektion von x auf U, so folgt

$$\pi(x) - x \perp U \quad \text{und} \quad \pi'(x) - x \perp U$$

also

$$0 = <u - x, u - u'> \quad \text{und} \quad 0 = <u' - x, u - u'> .$$

Damit ergibt sich durch Subtraktion $0 = <u - u', u - u'>$, also $u = u'$.

Seien nun $V = U \oplus U^{\perp}$ und $x \in (U^{\perp})^{\perp}$. Dann gilt $x = u + v$ mit Vektoren $u \in U$ und $v \in U^{\perp}$. Daraus folgt $\langle v,v\rangle = \langle x,v\rangle - \langle u,v\rangle = 0$. Also ist $v = o$ und somit $x \in U$. Da umgekehrt stets $U \subset (U^{\perp})^{\perp}$ gilt, folgt $U = (U^{\perp})^{\perp}$. ∎

Daß die Bedingung $V = U \oplus U^{\perp}$ nicht immer erfüllt ist, wollen wir an einem Beispiel sehen.

Beispiel. Sei $V = l^2$ der Standardhilbertraum und $U = \mathbb{R}[X]$ sei der Untervektorraum der Polynome. Dann ist $U^{\perp} = \{o\}$, also $V \neq U \oplus U^{\perp}$.

Beweis. Sei $x = (x_0,x_1,\ldots) \in U^{\perp}$. Dann ist $\langle x,y\rangle = 0$ für alle $y \in U$ und somit $\langle x,e_i\rangle = 0$ für alle $i \in \mathbb{N}_0$ mit

$$e_i = (0,\ldots,0,\underbrace{1}_{i\text{-te Stelle}},0,\ldots) \in U.$$

Daraus folgt $x_i = 0$ für alle i, also $x = o$. Somit ist $U^{\perp} = \{o\}$. ∎

In dem euklidischen Vektorraum l^2 gibt es daher keine Orthogonalprojektion auf den Untervektorraum U.

Wie wir zu Beginn unserer Überlegungen über Projektionen gesehen haben, existiert π immer, wenn $\dim U < \infty$ ist. Wir wollen dies als Satz festhalten.

Satz 9. *Es seinen V ein euklidischer Vektorraum beliebiger Dimension und $U \subset V$ ein Untervektorraum endlicher Dimension. Dann existiert die Orthogonalprojektion π auf U, und es gilt $V = U \oplus U^{\perp}$ und $(U^{\perp})^{\perp} = U$.*

Beispiel. Sei V der Vektorraum $\mathbb{R}^5$, versehen mit dem Standardskalarprodukt, und seien

$$U = \left[\begin{pmatrix}1\\1\\0\\1\\0\end{pmatrix},\begin{pmatrix}2\\0\\1\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\\1\\0\end{pmatrix}\right] \quad \text{sowie} \quad x = \begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}.$$

Gesucht ist die Orthogonalprojektion von x auf U.

1. Schritt: Wir bestimmen eine ONB in U:

$$y_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \; ; \; \| y_1 \| = \sqrt{3} \, ,$$

$$y_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{2+1}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \; ; \; \| y_2 \| = \sqrt{3} \, ,$$

$$y_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \; ; \; \| y_3 \| = \frac{2}{\sqrt{3}} \, .$$

Damit erhalten wir als ONB in U:

$$x_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \; x_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; x_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} .$$

2. Schritt: Es ist $\pi(x) = \langle x,x_1 \rangle x_1 + \langle x,x_2 \rangle x_2 + \langle x,x_3 \rangle x_3 = \frac{1}{3} \begin{bmatrix} 3 \\ 5 \\ 8 \\ 13 \\ 0 \end{bmatrix}$.

Orthogonalprojektionen lassen sich auch durch eine Minimalitätseigenschaft des Abstandes charakterisieren.

Satz 10 und Definition. *Es seien V ein euklidischer Vektorraum, $U \subset V$ ein Untervektorraum und $\pi \in \mathrm{Hom}\,(V,V)$ mit* Bild $\pi \subset U$. *Dann ist π genau dann Orthogonalprojektion auf U, wenn für alle $x \in V$ und alle $u \in U$ gilt:*

$$\| x - \pi(x) \| \leq \| x - u \| .$$

Ist π Orthogonalprojektion auf U, so heißt $d(x,U) := \| x - \pi(x) \|$ der Abstand von x zu U.

Beweis. Sei π Orthogonalprojektion auf U. Dann gilt für $x \in V$, $u \in U$:

$$x - u = x - \pi(x) + \pi(x) - u .$$

Wegen $\pi(x) - u \in U$ ist $x - \pi(x) \perp \pi(x) - u$. Damit folgt mit dem Satz von Pythagoras

$$\| x - u \|^2 = \| x - \pi(x) \|^2 + \| \pi(x) - u \|^2 \geq \| x - \pi(x) \|^2 .$$

Umgekehrt gelte nun $\| x - \pi(x) \| \leq \| x - u \|$ für alle $x \in V$, $u \in U$. Dann folgt speziell für $x \in U$

$$\| x - \pi(x) \| \leq \| x - x \| = 0,$$

also $x = \pi(x)$. Somit ist wegen Bild $\pi \subset U$ $\pi^2 = \pi$ und Bild $\pi = U$.

Nun müssen wir noch die Orthogonalitätseigenschaft $\pi(x) - x \perp \pi(x)$ nachweisen. Für $\pi(x) = o$ ist dies trivial. Für $\pi(x) \neq o$ betrachten wir $(1 + a)\, \pi(x)$, $a \in \mathbb{R}$. Es gilt

$$\| x - \pi(x) \|^2 \leq \| x - (1 + a)\, \pi(x) \|^2 = \| x - \pi(x) - a\, \pi(x) \|^2$$

$$= \| x - \pi(x) \|^2 - 2\, a \langle x - \pi(x), \pi(x) \rangle + a^2 \| \pi(x) \|^2 ,$$

also

$$2\, a \langle x - \pi(x), \pi(x) \rangle \leq a^2 \| \pi(x) \|^2 .$$

Wählen wir speziell

$$a = \frac{\langle x - \pi(x), \pi(x) \rangle}{\| \pi(x) \|^2} ,$$

so erhalten wir

$$2\, a^2 \| \pi(x) \|^2 \leq a^2 \| \pi(x) \|^2 .$$

Also ist $a = 0$ und somit $\langle x - \pi(x), \pi(x) \rangle = 0$. ∎

Beispiel. In dem Beispiel von S. 243/244 ist $d\,(x,U) = \| \pi(x) - x \| = \frac{2}{3}\sqrt{57}$.

Bemerkungen. **(a)** Orthogonalprojektionen sind kontrahierend, d.h. sie verkleinern die Norm und den Abstand:

(i) $\| \pi(x) \| \leq \| x \|$, für alle $x \in V$

(ii) $\| \pi(x) - \pi(y) \| \leq \| x - y \|$, für alle $x,y \in V$,

Umgekehrt ist jede Projektion π auf U, die (i) oder (ii) erfüllt, Orthogonalprojektion (Übungsaufgabe).

(b) Geometrische Interpretation des Orthogonalisierungsverfahrens von Gram–Schmidt:

Es sei $(x_1,x_2,\ldots)$ ein System linear unabhängiger Vektoren und $(y_1,y_2,\ldots)$ sei das daraus konstruierte Orthogonalsystem. Bezeichnen wir für $k = 1,2,\ldots$ mit π_k die Orthogonalprojektion auf $U_k = [y_1,\ldots,y_k] = [x_1,\ldots,x_k]$, so gilt

$$y_{k+1} = x_{k+1} - \pi_k(x_{k+1}) .$$

Wendet man das Verfahren allgemeiner auf ein beliebiges System $(x_1,x_2,\ldots)$ an mit $x_i \neq o$, und ist $x_{k+1} \in [x_1,\ldots,x_k]$, so folgt wegen $\pi_k(x_{k+1}) = x_{k+1}$, daß $y_{k+1} = o$ ist. Läßt man bei dem modifizierten Verfahren aus dem System $(y_1,y_2,\ldots)$ alle Vektoren $y_i = o$ weg, so erhält man wieder ein Orthogonalsystem.

§ 3 Adjungierte Abbildungen

Wir betrachten zwei euklidische Vektorräume V, W sowie eine lineare Abbildung $\Phi \in \mathrm{Hom}(V,W)$. Mit Hilfe der Skalarprodukte in V und W wollen wir eine weitere lineare Abbildung $\Psi \in \mathrm{Hom}(W,V)$ finden, die eng mit Φ zusammenhängt.

Definition. Es seien V und W euklidische Vektorräume und $\Phi \in \mathrm{Hom}\,(V,W)$. Eine Abbildung $\Psi \in \mathrm{Hom}(W,V)$ heißt *adjungierte Abbildung* von Φ, wenn

$$(*) \qquad <\Phi(v),w> = <v,\Psi(w)>$$

für alle $v \in V$ und alle $w \in W$ gilt.

Bemerkungen. **(a)** Man beachte, daß in obiger Definition links das Skalarprodukt in W, rechts das in V steht.

(b) Es gibt höchstens eine zu Φ adjungierte Abbildung, wir bezeichnen sie mit Φ^*.

Beweis. Aus $<\Phi(v),w> = <v,\Psi_1(w)> = <v,\Psi_2(w)>$ folgt $<v,(\Psi_1 - \Psi_2)(w)> = 0$. Da hier $v \in V$ und $w \in W$ beliebig sind, ist $(\Psi_1 - \Psi_2)(w) = o$ für alle $w \in W$, d.h. $\Psi_1 = \Psi_2$. ∎

(c) Existiert die adjungierte Abbildung Φ^* von Φ, so existiert auch $(\Phi^*)^*$, nämlich $(\Phi^*)^* = \Phi$.

(d) Nicht zu jedem Homomorphismus $\Phi \in \mathrm{Hom}(V,W)$ existiert die adjungierte Abbildung Φ^*.

Beispiel. Sei $V = \mathbb{R}[X]$, $W = l^2$ und $\Phi : \mathbb{R}[X] \longrightarrow l^2$, $v \longmapsto v$, die identische Abbildung. Würde die adjungierte Abbildung $\Phi^* : l^2 \longrightarrow \mathbb{R}[X]$ existieren, so müsste für alle $v \in \mathbb{R}[X]$, $w \in l^2$

$$<v,w> = <\Phi(v),w> = <v,\Phi^*(w)>$$

gelten. Somit wäre $w - \Phi^*(w) \in \mathbb{R}[X]^\perp = \{o\}$ für alle $w \in l^2$, also $\Phi^*(w) = w$ und damit $\mathbb{R}[X] = l^2$. Dies ist ein Widerspruch.

Um festzustellen, ob zu gegebenem Φ die adjungierte Abbildung Φ^* existiert,

wäre es naheliegend, Φ^* durch die Gleichung (*) zu erklären, d.h. zu festem $w \in W$ das Bild $v = \Phi^*(w)$ so zu wählen, daß

$$<\Phi(x),w> \;=\; <x,v>$$

für alle $x \in V$ gilt. Nun ist $x \longmapsto <\Phi(x),w>$ bei festem $w \in W$ eine Linearform auf V, also ein Element x^* von V^* ; ebenso ist $x \longmapsto <x,v>$ eine Linearform. Unsere obige Definition von Φ^* wäre also sinnvoll, wenn zu jedem $x^* \in V^*$ ein $v \in V$ existierte mit

$$x^*(x) \;=\; <x,v> \,.$$

Wir werden das jetzt im Fall dim $V < \infty$ zeigen, im Fall dim $V = \infty$ ist es im allgemeinen falsch, wie sich aus dem obigen Beispiel ergibt.

Satz 11 (Rieszscher Darstellungssatz). *Es seien V ein euklidischer Vektorraum mit* dim $V < \infty$ *und $x^* \in V^*$ eine Linearform. Dann existiert genau ein $v \in V$ mit*

$$x^*(x) \;=\; <x,v>$$

für alle $x \in V$.

Beweis. Seien dim $V = n$ und $B = (x_1,\ldots,x_n)$ eine ONB von V. Wir setzen

$$v \;:=\; \sum_{i=1}^{n} x^*(x_i)\, x_i \,.$$

Dann gilt

$$<x,v> \;=\; \sum_{i=1}^{n} x^*(x_i)\, <x,x_i> \;=\; x^*\Big(\sum_{i=1}^{n} <x,x_i>\, x_i\Big) \;=\; x^*(x)$$

für alle $x \in V$.

Erfüllt $v' \in V$ ebenfalls $x^*(x) = <x,v'>$ für alle $x \in V$, so folgt $<x, v - v'> = 0$ für alle $x \in V$, also $v - v' = o$. ∎

Mit Hilfe von Satz 11 beweisen wir nun die Existenz der zu Φ adjungierten Abbildung Φ^*, falls V und W endlich dimensional sind.

Satz 12. *Es seien V, W endlich dimensionale euklidische Vektorräume und Φ aus* Hom (V, W). *Dann existiert die adjungierte Abbildung* $\Phi^* \in$ Hom (W, V).

Beweis. Wie schon angedeutet, können wir für $w \in W$ den Bildvektor $\Phi^*(w)$ durch

$$<\Phi(x), w> \;=\; <x, \Phi^*(w)> \qquad \text{für alle } x \in V$$

erklären, weil nach Satz 11 ein solches Element $\Phi^*(w) := v$ existiert. Damit ist eine Abbildung $\Phi^*\colon W \longrightarrow V$ erklärt. Es bleibt noch die Linearität von Φ^* nachzuweisen: Für alle $x \in V$ ist

$$<x, \Phi^*(a_1 w_1 + a_2 w_2)> \;=\; <\Phi(x),\, a_1 w_1 + a_2 w_2> \;=\; a_1 <\Phi(x), w_1> + a_2 <\Phi(x), w_2>$$

$$= a_1 <x, \Phi^*(w_1)> + a_2 <x, \Phi^*(w_2)> \;=\; <x, a_1 \Phi^*(w_1) + a_2 \Phi^*(w_2)> .$$

Also folgt für alle $w_1, w_2 \in W$ und alle $a_1, a_2 \in \mathbb{R}$

$$\Phi^*(a_1 w_1 + a_2 w_2) \;=\; a_1 \Phi^*(w_1) + a_2 \Phi^*(w_2) . \qquad \blacksquare$$

Bemerkungen. **(a)** In Kapitel 3, S.144 haben wir gezeigt, daß es für einen endlich dimensionalen Vektorraum V einen natürlichen basisunabhängigen Isomorphismus $\Psi : V \longrightarrow V^{**}$ gibt, während wir für V und V^* keinen basisunabhängigen Isomorphismus angeben konnten. Für einen euklidischen Vektorraum V mit dim $V < \infty$ ist dies anders, denn hier ist die Abbildung

$$\begin{aligned} \Psi : V &\longrightarrow V^* \\ x &\longmapsto <\cdot, x> \end{aligned}$$

ein solcher Isomorphismus. Wir können deshalb in diesem Fall V mit V^* identifizieren und werden dies ab jetzt auch tun.

(b) Wird V mit V^* identifiziert, so fällt eine Basis B von V genau dann mit der dualen Basis B^* in V^* zusammen, wenn B Orthonormalbasis ist.

(c) Setzt man $V = V^*$ und $W = W^*$ für dim $V < \infty$ und dim $W < \infty$ (wie in (a)), so fällt die adjungierte Abbildung $\Phi^* : W \longrightarrow V$ mit der transponierten Abbildung $\Phi^{\mathsf{T}} : W^* \longrightarrow V^*$ zusammen.

(d) Als Folgerung von (b),(c) ergibt sich nun, daß für endlich dimensionale euklidische Vektorräume V, W, für $\Phi \in \mathrm{Hom}\,(V,W)$ und für Orthonormalbasen B von V und C von W die Abbildungsmatrix von Φ^* gerade die Transponierte der Abbildungsmatrix von Φ ist (bzgl. B, C) :

$$A_{\Phi^*} = A_{\Phi}^{\mathsf{T}}.$$

Dies läßt sich auch direkt beweisen: Seien $B = (v_1,\ldots,v_n)$, $C = (w_1,\ldots,w_m)$ Orthonormalbasen von V, W, $A_\Phi = (\!(a_{ij})\!)$, $A_{\Phi^*} = (\!(b_{ij})\!)$. Dann gilt für $j = 1,\ldots,n$ und $i = 1,\ldots,m$

$$\Phi(v_j) = \sum_{i=1}^{m} a_{ij}\, w_i = \sum_{i=1}^{m} \langle \Phi(v_j), w_i \rangle\, w_i \ ,$$

$$\Phi^*(w_i) = \sum_{j=1}^{n} b_{ji}\, v_j = \sum_{j=1}^{n} \langle \Phi^*(w_i), v_j \rangle\, v_j \ .$$

Also gilt

$$a_{ij} = \langle \Phi(v_j), w_i \rangle = \langle v_j, \Phi^*(w_i) \rangle = b_{ji}\,, \text{ d.h. } A_{\Phi^*} = A_{\Phi}^{\mathsf{T}}. \qquad \blacksquare$$

Für Endomorphismen Φ eines endlich dimensionalen euklidischen Vektorraumes V folgt daraus $\det \Phi = \det \Phi^*$.

(e) Für Φ^* gelten die folgenden Rechenregeln, die wir schon von der transponierten Abbildung her kennen. Für endlich dimensionale Vektorräume folgen die Aussagen daher direkt aus (c). Im allgemeinen Fall kann man sie auch leicht aus der Definition von Φ^* herleiten (Übungsaufgabe):

(i) $(\Phi^*)^* = \Phi$,

(ii) $(a\,\Phi)^* = a\,\Phi^*$, $a \in \mathbb{R}$,

(iii) $(\Phi_1 + \Phi_2)^* = \Phi_1^* + \Phi_2^*$,

(iv) $(\Phi \circ \Psi)^* = \Psi^* \circ \Phi^*$.

Über die adjungierte Abbildung erhalten wir eine wichtige Klasse von Abbildungen.

Definition. Es seien V ein euklidischer Vektorraum und $\Phi \in \text{Hom}\,(V,V)$. Die Abbildung Φ heißt *selbstadjungiert*, wenn

$$<\Phi(x),y> \;=\; <x,\Phi(y)>$$

für alle $x,y \in V$ gilt. Φ heißt *antiselbstadjungiert*, wenn

$$<\Phi(x),y> \;=\; -<x,\Phi(y)>$$

für alle $x,y \in V$ gilt.

Bemerkungen. **(a)** Φ ist genau dann selbstadjungiert bzw. antiselbstadjungiert, wenn Φ^* existiert und $\Phi = \Phi^*$ bzw. $\Phi = -\Phi^*$ erfüllt.
(b) Ist dim $V < \infty$ und ist B eine ONB von V, so ist Φ genau dann selbstadjungiert bzw. antiselbstadjungiert, wenn für die Abbildungsmatrix von Φ bezüglich B

$$A_\Phi = A_\Phi{}^\top \text{ bzw. } A_\Phi = -A_\Phi{}^\top$$

gilt. Matrizen A mit $A^\top = -A$ heißen *schiefsymmetrisch.*

Im endlich dimensionalen Fall werden selbstadjungierte Endomorphismen Φ also durch symmetrische Abbildungsmatrizen beschrieben und sind daher nach Satz 3 diagonalisierbar. Der folgende Satz ist eine Verschärfung von Satz 3.

Satz 13. *Es seien V ein endlich dimensionaler euklidischer Vektorraum und Φ ein Endomorphismus von V. Genau dann ist Φ selbstadjungiert, wenn es in V eine Orthonormalbasis aus Eigenvektoren von Φ gibt.*

Beweis. Sei zunächst $B = (x_1,\ldots,x_n)$ eine ONB aus Eigenvektoren von Φ und $c_1,\ldots c_n$ die zugehörigen Eigenwerte. Dann folgt $<\Phi(x_i),x_j> \;=\; c_i\,<x_i,x_j> \;=\; c_i\,\delta_{ij} \;=\; c_j\,\delta_{ij} \;=\; <x_i,\Phi(x_j)>$. Folglich ist $<\Phi(x),y> = <x,\Phi(y)>$ für alle $x,y \in V$, also $\Phi = \Phi^*$.

Umgekehrt sei $\Phi = \Phi^*$ und B sei eine ONB von V. Dann gilt $A_\Phi = A_\Phi{}^\top$. Nach Satz 3 ist A_Φ diagonalisierbar, und die Eigenräume zu verschiedenen Eigenwerten sind zueinander orthogonal. Wählen wir in jedem Eigenraum eine ONB, so erhalten wir insgesamt eine ONB $(\hat{x}_1,\ldots,\hat{x}_n)$ des $\mathbb{R}^n$ aus Eigenvektoren von A_Φ . Die zugehö-

rigen Vektoren $x_1,\ldots,x_n$ sind dann Eigenvektoren von Φ und bilden eine ONB in V. ■

Bemerkung und Definition. Die Matrix S, die die symmetrische Matrix A_Φ auf Diagonalgestalt transformiert, kann also so gewählt werden, daß ihre Spalten eine ONB von $\mathbb{R}^n$ bilden. Dies ist gleichwertig mit $S^\tau S = E_n$.

Eine Matrix $A \in \mathbb{R}^{n\times n}$, die $A^\tau A = E_n$ erfüllt, heißt *orthogonal*. Es gilt:

$$A \text{ orthogonal} \iff A^{-1} = A^\tau \iff A\,A^\tau = E_n\,.$$

Quadratische (n,n)-Matrizen A und A', für die $A' = S^\tau A\,S$ mit einer orthogonalen Matrix S gilt, heißen *orthogonal äquivalent*.

Damit lautet die Aussage von Satz 13 für Matrizen:

Genau dann ist die Matrix $A \in \mathbb{R}^{n\times n}$ symmetrisch, wenn sie zu einer Diagonalmatrix orthogonal äquivalent ist.

Bemerkung. Für schiefsymmetrische Matrizen kann man ebenfalls eine Normalform angeben (siehe S.274).

Beispiele. **(a)** Zu der symmetrischen Matrix

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

wird eine orthogonale Matrix S gesucht, so daß $S^\tau A\,S$ eine Diagonalmatrix ist.

Aus dem charakteristischen Polynom $p = -(2+X)^2\,X$ von A ergeben sich die Eigenwerte $c_1 = 0$ und $c_2 = -2$. Für die zugehörigen Eigenräume E_{c_1} und E_{c_2} erhalten wir:

$$A\,v = o \iff v = \begin{bmatrix} a \\ 0 \\ a \end{bmatrix},\ a \in \mathbb{R} \quad ; \quad (A + 2\,E_3)\,v = o \iff v = \begin{bmatrix} a \\ b \\ -a \end{bmatrix},\ a,b \in \mathbb{R}\,,$$

also

$$E_{c_1} = \left[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right] \text{ und } E_{c_2} = \left[\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right].$$

Nun sind

$$\left\{\begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}\right\} \quad \text{bzw.} \quad \left\{\begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

Orthonormalbasen von E_{c_1} bzw. von E_{c_2}. Also ist

$$S = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}$$

eine orthogonale Matrix mit

$$S^T A\, S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

(b) Orthogonalprojektionen π_U sind selbstadjungiert: Ist V endlich dimensional, so kann man eine ONB von U wählen und diese zu einer ONB von V ergänzen. Die Abbildungsmatrix von π_U bezüglich dieser Basis hat dann die Gestalt

$$A_{\pi_U} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & O & \\ & & 1 & & & \\ & & & 0 & & \\ & O & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

Für unendlich dimensionale Vektorräume überlassen wir den Beweis als Übungsaufgabe.

In unendlich dimensionalen euklidischen Vektorräumen haben selbstadjungierte Abbildungen nicht die gleichen guten Eigenschaften, sie müssen noch nicht einmal Eigenwerte besitzen.

Beispiel. In dem euklidischen Vektorraum $V = C([0,1])$ mit dem Skalarprodukt

$$<f,g> = \int_0^1 f(t)\, g(t)\, \mathrm{d}t$$

betrachten wir die lineare Abbildung

$$\Phi : V \longrightarrow V, f \longmapsto \Phi(f) \text{ mit } \Phi(f)(t) = t \cdot f(t), t \in [0,1].$$

Φ ist selbstadjungiert:

$$<\Phi(f),g> = \int_0^1 t\, f(t)\, g(t)\, \mathrm{d}t = <f,\Phi(g)>.$$

Ist $c \in \mathbb{R}$ Eigenwert von Φ, so existiert ein Eigenvektor f, d.h. eine von der Nullfunktion o verschiedene Funktion $f \in C([0,1])$ mit $\Phi(f) = c\, f$. Daraus folgt $t\, f(t) = c\, f(t)$ für alle $t \in [0,1]$, also $f(t) = 0$ für alle $t \neq c$. Für $c \notin [0,1]$ folgt daher $f = o$, für $c \in [0,1]$ gilt wegen der Stetigkeit von f ebenfalls $f = o$. Dies ist ein Widerspruch, daher besitzt Φ keine Eigenwerte.

Weitere Struktureigenschaften selbstadjungierter Endomorphismen werden bei der Untersuchung einer allgemeineren Klasse linearer Abbildungen in § 5.5 hergeleitet.

§ 4 Isometrien

Es seien V und W euklidische Vektorräume. Die naheliegendste und interessanteste Klasse von linearen Abbildungen $\Phi : V \longrightarrow W$ wird von denjenigen Abbildungen gebildet, die außer mit der Vektorraumstruktur auch mit dem Skalarprodukt in V bzw. W verträglich sind, also

$$<\Phi(x),\Phi(y)> \;=\; <x,y>$$

für alle $x,y \in V$ erfüllen.

Definition. Es seien V, W euklidische Vektorräume. $\Phi \in \mathrm{Hom}\,(V,W)$ heißt *Isometrie* oder *orthogonale Abbildung*, wenn

$$<\Phi(x),\Phi(y)> \;=\; <x,y>$$

für alle $x,y \in V$ gilt.

Bemerkung. Jede Isometrie Φ ist injektiv. Ist Φ außerdem surjektiv, so heißen die zwei euklidischen Vektorräume V und W *isometrisch isomorph* .

Der Name Isometrie wird durch folgenden Satz gerechtfertigt.

Satz 14. *Es seien V, W euklidische Vektorräume und $\Phi \in \mathrm{Hom}\,(V,W)$. Dann sind folgende Aussagen äquivalent:*

(a) *Φ ist Isometrie.*

(b) $\| \Phi(x) \| = \| x \|$ *für alle* $x \in V$.

(c) $d\,(\Phi(x),\Phi(y)) = d\,(x,y)$ *für alle* $x,y \in V$.

Beweis. (a) $\Longrightarrow$ (b): $\| \Phi(x) \| \;=\; \sqrt{<\Phi(x),\Phi(x)>} \;=\; \sqrt{<x,x>} \;=\; \| x \|$.

(b) $\Longrightarrow$ (c): $\| \Phi(x) - \Phi(y) \| \;=\; \| \Phi(x-y) \| \;=\; \| x-y \|$.

(c) $\Longrightarrow$ (a): $<\Phi(x),\Phi(y)> \;=\; \frac{1}{4}\,(\| \Phi(x) + \Phi(y) \|^2 - \| \Phi(x) - \Phi(y) \|^2)$

$= \frac{1}{4}\,(\| \Phi(x) - \Phi(-y) \|^2 - \| \Phi(x) - \Phi(y) \|^2) \;=\; \frac{1}{4}\,(\| x - (-y) \|^2 - \| x - y \|^2)$

$$= \tfrac{1}{4}(\| x+y \|^2 - \| x-y \|^2) = \langle x,y \rangle . \qquad \blacksquare$$

Im Fall dim V = dim $W = n$ lassen sich für Isometrien weitere Kennzeichnungen angeben.

Satz 15. *Es seien V, W euklidische Vektorräume mit* dim V = dim $W = n$ *sowie* $\Phi \in \mathrm{Hom}(V,W)$. *Dann sind äquivalent:*

(a) Φ *ist Isometrie.*

(b) *Für jede ONB* $(x_1,\ldots,x_n)$ *von* V *ist* $(\Phi(x_1),\ldots,\Phi(x_n))$ *ONB von* W.

(c) *Es gibt eine ONB* $(x_1,\ldots,x_n)$ *von* V, *für die* $(\Phi(x_1),\ldots,\Phi(x_n))$ *ONB von* W *ist.*

(d) $\Phi^* \circ \Phi = \mathrm{id}_V$, $\Phi \circ \Phi^* = \mathrm{id}_W$.

(e) *Bezüglich jeder ONB* B *von* V *und jeder ONB* C *von* W *ist* A_Φ *orthogonal.*

(f) *Es gibt eine ONB* B *von* V *und eine ONB* C *von* W, *für die* A_Φ *orthogonal ist.*

Beweis. (a) $\Longrightarrow$ (b): Φ ist injektiv, also wegen dim V = dim $W = n$ ein Isomorphismus. Daher ist $(\Phi(x_1),\ldots,\Phi(x_n))$ eine Basis von W. Wegen $\langle \Phi(x_i),\Phi(x_j) \rangle = \langle x_i,x_j \rangle = \delta_{ij}$ für $i,j = 1,\ldots,n$ ist $(\Phi(x_1),\ldots,\Phi(x_n))$ ONB von W.

(b) $\Longrightarrow$ (c): Nach Satz 7 existiert in V eine ONB $(x_1,\ldots,x_n)$. Dann ist $(\Phi(x_1),\ldots,\Phi(x_n))$ ONB von W.

(c) $\Longrightarrow$ (f): Bezüglich der Orthonormalbasen $(x_1,\ldots,x_n)$ von V und $(\Phi(x_1),\ldots,\Phi(x_n))$ von W ist $A_\Phi = E_n$, also ist A_Φ orthogonal.

(f) $\Longrightarrow$ (d): Bezüglich der Orthonormalbasen B und C gilt $A_{\Phi^* \circ \Phi} = A_{\Phi^*} A_\Phi = A_\Phi{}^{\mathsf{T}} A_\Phi = E_n$, also $\Phi^* \circ \Phi = \mathrm{id}_V$. Analog folgt $\Phi \circ \Phi^* = \mathrm{id}_W$.

(d) $\Longrightarrow$ (a): Es gilt $\langle \Phi(x),\Phi(y) \rangle = \langle x,\Phi^* \circ \Phi(y) \rangle = \langle x,y \rangle$ für alle $x,y \in V$.

(d) $\Longrightarrow$ (e): Bezüglich jeder ONB B von V und jeder ONB C von W gilt $A_{\Phi^*} = A_\Phi^{\mathsf{T}}$, also $A_\Phi^{\mathsf{T}} A_\Phi = A_{\Phi^*} A_\Phi = A_{\Phi^* \circ \Phi} = E_n$. Somit ist A_Φ orthogonal.

(e) $\Longrightarrow$ (f): Nach Satz 7 gibt es in V eine ONB B und in W eine ONB C. $\blacksquare$

Wir betrachten nun speziell eine Isometrie $\Phi : V \longrightarrow V$, dim $V = n$. Wegen $\Phi^* \circ \Phi = \mathrm{id}_V$ und $\det \Phi = \det \Phi^*$ folgt $\det \Phi = \pm 1$.

Ist c Eigenwert von Φ, so erhalten wir aus $\Phi(x) = c\,x$, $x \neq o$, unmittelbar $\| \Phi(x)\| = |c| \, \| x \|$, also $|c| = 1$.

Isometrien $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ heißen auch *Drehungen* des $\mathbb{R}^n$; ist $\det \Phi = 1$, so spricht man von *eigentlichen Drehungen*, bei $\det \Phi = -1$ von *uneigentlichen Drehungen* oder *Drehspiegelungen*.

Für $\dim V < \infty$ bilden die Isometrien $\Phi : V \longrightarrow V$ eine Untergruppe $O(V)$ von $\mathrm{Aut}(V)$. Die zugehörige Matrizengruppe heißt *orthogonale Gruppe* und wird mit $O(n)$ bezeichnet. Die Matrizen A aus dieser Gruppe mit $\det A = 1$ bilden wieder eine Untergruppe, die *spezielle orthogonale Gruppe* $SO(n)$.

Diese Bezeichnungen beinhalten schon eine geometrische Interpretation von Isometrien, die wir nun nachvollziehen wollen. Dazu leiten wir für Isometrien eine Normalform her.

Satz 16. *Es seien V ein n-dimensionaler euklidischer Vektorraum und $\Phi \in \mathrm{Hom}\,(V,V)$. Dann ist Φ genau dann Isometrie, wenn eine geordnete Orthonormalbasis von V existiert, bezüglich der A_Φ die Form*

$$A_\Phi = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & \cos\omega_1 & -\sin\omega_1 & \\ & & & & & & \sin\omega_1 & \cos\omega_1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & \cos\omega_k & -\sin\omega_k \\ & & & & & & & & & \sin\omega_k & \cos\omega_k \end{bmatrix} \qquad (*)$$

besitzt mit $\omega_i \in (0,\pi)$, $i = 1,\ldots,k$.

Beweis. Jede Matrix A_Φ dieser Form erfüllt $A_\Phi{}^{\mathsf{T}} A_\Phi = E_n$, also ist A_Φ orthogonal und Φ somit eine Isometrie.

Die Umkehrung beweisen wir so, daß damit zugleich ein praktisches Verfahren zur Aufstellung der Normalform gewonnen wird.

Sei Φ eine Isometrie. Wir betrachten die Hilfsabbildung

$$\Psi := \Phi + \Phi^* = \Phi + \Phi^{-1} .$$

Ψ ist selbstadjungiert, also existiert nach Satz 13 in V eine ONB aus Eigenvektoren von Ψ. Somit gilt

$$V = E_2 \oplus E_{-2} \oplus E_{c_1} \oplus \cdots \oplus E_{c_k}$$

mit $c_i \neq \pm 2$, und alle Eigenräume von Ψ sind paarweise orthogonal. Dabei tritt der Eigenraum E_2 von Ψ zum Eigenwert 2 genau dann auf, wenn Φ den Eigenwert 1 besitzt, und E_{-2} tritt genau dann auf, wenn Φ den Eigenwert –1 hat. Wir zeigen dies für den ersten Fall, der andere folgt dann analog:

Aus $\Phi(x) = x$ folgt $\Phi^{-1}(x) = x$ und somit $\Psi(x) = 2\,x$. Sei umgekehrt $\Psi(x) = 2\,x$, also $(\Phi + \Phi^*)(x) = 2\,x$. Dann ist $2 <x,x> = <\Phi(x),x> + <\Phi^*(x),x> = 2 <\Phi(x),x>$ und somit $<\Phi(x) - x, \Phi(x) - x> = 0$. Also gilt $\Phi(x) = x$.

Alle Eigenräume E_c von Ψ sind Φ–invariant, denn für $x \in E_c$ folgt

$$\Psi(\Phi(x)) = (\Phi + \Phi^{-1})(\Phi(x)) = \Phi \circ (\Phi + \Phi^{-1})(x) = \Phi(\Psi(x)) = \Phi(c\,x) = c\,\Phi(x) ,$$

also $\Phi(x) \in E_c$. Somit folgt $\Phi(E_c) \subset E_c$. Da Φ Isometrie ist, gilt sogar $\Phi(E_c) = E_c$.

Wir können damit alle Eigenräume von Ψ getrennt behandeln und dort geeignete Orthonormalbasen suchen.

1. Fall: $c = \pm 2$. Wie wir oben schon gesehen haben, ist $E_2(\Psi)$ gerade der Eigenraum von Φ zum Eigenwert 1 und $E_{-2}(\Psi)$ ist der Eigenraum von Φ zum Eigenwert –1. In jedem dieser Eigenräume können wir beliebige Orthonormalbasen wählen.

2. Fall: $c \neq \pm 2$. Dann ist dim E_c gerade und es gibt $l = \frac{1}{2}$ dim E_c paarweise orthogonale, zweidimensionale und Φ–invariante Untervektorräume $U_1,\ldots,U_l$ mit $E_c = U_1 \oplus \cdots \oplus U_l$.

Beweis. Wäre dim E_c ungerade, so hätte das charakteristische Polynom von $\Phi|_{E_c}$ eine Nullstelle. Diese wäre Eigenwert von Φ, also ± 1, und somit gäbe es in E_c einen Eigenvektor von Ψ zum Eigenwert ± 2. Dies ist ein Widerspruch.

Sei nun $x_1 \in E_c$, $\| x_1 \| = 1$, beliebig gewählt. Dann sind $x_1, \Phi(x_1)$ linear unabhängig, da sonst $\Phi(x_1) = \pm x_1$ und somit $\Psi(x_1) = \pm 2\, x_1$ folgen würde. Der Untervektorraum $U_1 = [x_1, \Phi(x_1)] \subset E_c$ ist also zweidimensional.

U_1 ist Φ-invariant: Aus $\Phi \circ (\Phi + \Phi^{-1})(x_1) = \Phi(\Psi(x_1)) = \Phi(c\, x_1) = c\, \Phi(x_1)$ folgt nämlich $\Phi^2(x_1) + x_1 = c\, \Phi(x_1)$, also $\Phi^2(x_1) \in U_1$. Somit gilt $\Phi(U_1) \subset U_1$ und, weil Φ Isometrie ist, sogar $\Phi(U_1) = U_1$.

$U_1^{\perp}$ ist ebenfalls Φ-invariant: Seien $x \in U_1^{\perp}$ und $y \in U_1$ beliebig gewählt. Dann gilt $\Phi^{-1}(y) \in U_1$ und somit $\langle \Phi(x), y \rangle = \langle x, \Phi^*(y) \rangle = \langle x, \Phi^{-1}(y) \rangle = 0$, also $\Phi(x) \in U_1^{\perp}$. Wegen $\Phi(E_c) = E_c$ und $\Phi(U_1^{\perp}) = U_1^{\perp}$ gilt daher auch $\Phi(U_1^{\perp} \cap E_c) = U_1^{\perp} \cap E_c$.

Ist $U_1^{\perp} \cap E_c \neq \{o\}$, so wählen wir einen beliebigen normierten Vektor x_2 aus dieser Menge und bilden den Untervektorraum $U_2 = [x_2, \Phi(x_2)]$. Dieser ist wieder zweidimensional, Φ-invariant und wegen $U_2 \subset U_1^{\perp} \cap E_c$ orthogonal zu U_1. Danach wählen wir x_3 aus $(U_1 \oplus U_2)^{\perp} \cap E_c$, $\| x_3 \| = 1$, usw. Das Verfahren bricht nach $l = \frac{1}{2} \dim E_c$ Schritten ab, und wir erhalten die Darstellung $E_c = U_1 \oplus \cdots \oplus U_l$.

In jedem dieser l Unterräume $U = [x, \Phi(x)] \subset E_c$ konstruieren wir von der Basis $(x, \Phi(x))$ ausgehend eine Orthonormalbasis (x, y) mit

$$y := \frac{\Phi(x) - \langle \Phi(x), x \rangle\, x}{\| \Phi(x) - \langle \Phi(x), x \rangle\, x \|}.$$

Wir stellen den Vektor y in anderer Form dar: Zunächst folgt aus $\Psi(x) = c\, x$ die Gleichung $\langle \Phi(x), x \rangle = \frac{c}{2} \langle x, x \rangle = \frac{c}{2}$. Ist $\omega \in (0, \pi)$ der Winkel (vgl. S. 226) zwischen x und $\Phi(x)$, so gilt

$$\cos \omega = \frac{c}{2}, \quad \sin \omega = \sqrt{1 - c^2/4}.$$

Daraus folgt

$$y = (1/\sin \omega)\, (\Phi(x) - \cos \omega\, x)$$

und somit

$$\Phi(x) = \cos\omega\, x + \sin\omega\, y$$

und wegen $\Phi^2(x) = c\,\Phi(x) - x = 2\cos\omega\,\Phi(x) - x$

$$\Phi(y) = (1/\sin\omega)\,(\Phi^2(x) - \cos\omega\,\Phi(x))$$

$$= (1/\sin\omega)\,(2\cos\omega\,\Phi(x) - x - \cos\omega\,\Phi(x))$$

$$= (1/\sin\omega)\,(\cos^2\omega\, x + \cos\omega\sin\omega\, y - x) = -\sin\omega\, x + \cos\omega\, y\,.$$

Bezüglich dieser Basis gilt also

$$A_{\Phi|_U} = \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix}$$

und damit

$$A_{\Phi|_{E_c}} = \begin{bmatrix} \cos\omega & -\sin\omega & & & \\ \sin\omega & \cos\omega & & O & \\ & & \ddots & & \\ & O & & \cos\omega & -\sin\omega \\ & & & \sin\omega & \cos\omega \end{bmatrix}$$

mit $l(c) = \frac{1}{2}\dim E_c$ Kästchen.

Insgesamt erhalten wir somit die gesuchte Normalform A_Φ. ∎

Bemerkungen. **(a)** In der Literatur wird manchmal auch eine andere Matrix (**) als Normalform bezeichnet. Sie unterscheidet sich von der obigen Form (*) dadurch, daß die Zweierkästchen in (**) die Gestalt

$$\boxed{\begin{matrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{matrix}}$$

besitzen. Jede der Normalformen geht aus der anderen dadurch hervor, daß bei jedem Zweierkästchen jeweils die Reihenfolge der zugehörigen orthonormalen Basisvektoren x,y vertauscht wird oder daß y durch $-y$ ersetzt wird.

(b) Ist $A \in \mathbb{R}^{n\times n}$ orthogonal, so liefert das obige Beweisverfahren, angewendet auf die symmetrische Hilfsmatrix $B = A + A^{\top}$, eine orthogonale Matrix $S \in \mathbb{R}^{n\times n}$, so daß $S^{\top} A\, S$ die Normalform (*) besitzt.

(c) Im nächsten Paragraphen werden wir noch einen anderen Beweis kennenlernen, bei dem A als komplexe Matrix angesehen wird und der die reelle Normalform aus Satz 4.22 benutzt.

Geometrische Interpretation im $\mathbb{R}^n$

Die Abbildungsmatrix

$$A_1 = \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix}$$

beschreibt eine Abbildung $\Phi_1 : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, die folgende Eigenschaften hat. Schreiben wir $x \in \mathbb{R}^2$ in Polarkoordinaten

$$x = \begin{bmatrix} r\cos\alpha \\ r\sin\alpha \end{bmatrix},$$

so ist

$$\Phi_1(x) = \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} r\cos\alpha \\ r\sin\alpha \end{bmatrix} = \begin{bmatrix} r\cos(\alpha+\omega) \\ r\sin(\alpha+\omega) \end{bmatrix}.$$

Also ist Φ_1 eine Drehung im $\mathbb{R}^2$ um den Ursprung mit dem Drehwinkel ω. Damit ist auch klar, daß Φ_1 bezüglich jeder ONB in $\mathbb{R}^2$ eine der Matrizen A_1 oder $A_1^{\top}$ als Abbildungsmatrix besitzt.

Setzen wir nun diese Information über die einzelnen Kästchen von A_Φ zusammen, so erhalten wir die folgende geometrische Beschreibung einer Isometrie $\Phi : V \longrightarrow V$. Es gibt eine Orthogonalzerlegung

$$V = U_1 \oplus U_2 \oplus W_1 \oplus \cdots \oplus W_k,$$

so daß $\Phi|_{U_1} = \mathrm{id}_{U_1}$ ist, also U_1 Fixunterraum ist, so daß $\Phi|_{U_2} : x \longmapsto -x$ gilt, also Φ

die Vektoren aus U_2 an $U_2^{\perp} = U_1 \oplus W_1 \oplus \ldots \oplus W_k$ spiegelt, und so daß $\Phi|_{W_i}$, $i = 1,\ldots k$, Drehungen in den zweidimensionalen Untervektorräumen W_i darstellen.

Nun kann man die speziellen Matrizen

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ und } \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

ebenfalls als Drehmatrizen mit Winkel 0 bzw. π ansehen. Folglich ergeben sich genaue Beschreibungen von Φ in den Dimensionen 2 und 3:

n = 2 : Für det $\Phi = 1$ ist Φ eine Drehung um einen Winkel $\omega \in [0,\pi]$, für det $\Phi = -1$ ist Φ Spiegelung an einer Geraden durch den Ursprung.

n = 3 : det $\Phi = 1$: Es gibt einen Vektor $v \neq o$, der Fixvektor ist, im Orthogonalraum $v^{\perp}$ ist Φ eine Drehung. Φ heißt dann *Drehung* des Vektorraums V mit der *Drehachse* $[v]$ und der *Drehebene* $v^{\perp}$.

det $\Phi = -1$: Es gibt ein $v \neq o$ mit $\Phi(v) = -v$, im Orthogonalraum $v^{\perp}$ ist Φ eine Drehung. Φ ist dann Produkt einer Drehung um die Achse $[v]$ mit einer Spiegelung an der Drehebene $v^{\perp}$ und heißt deswegen eine *Drehspiegelung* von V .

Beispiel. Die Matrix

$$A = \frac{1}{9\sqrt{2}} \begin{bmatrix} 9 & 4 & 8 & 1 \\ -4 & 9 & -1 & 8 \\ -8 & 1 & 9 & -4 \\ -1 & -8 & 4 & 9 \end{bmatrix} \in \mathbb{R}^{4\times 4}$$

ist orthogonal, denn es gilt $A^{\top} A = E_4$. Wir wollen ihre Normalform $\tilde{A}$ bestimmen und betrachten hierzu zunächst die symmmetrische Matrix

$$A + A^{\top} = \frac{1}{9\sqrt{2}} \begin{bmatrix} 18 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 \\ 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 18 \end{bmatrix} = \sqrt{2}\, E_4 .$$

$A + A^T$ hat den 4–fachen Eigenwert $c = \sqrt{2}$. Die Normalform $\tilde{A}$ von A besteht daher aus 2 Kästchen der Form

$$\boxed{\begin{matrix} \cos\ \omega & -\sin\ \omega \\ \sin\ \omega & \cos\ \omega \end{matrix}}$$

mit $\cos\,\omega = \frac{1}{2}\sqrt{2}$. Daraus folgt $\omega = \frac{\pi}{4}$ und $\sin\,\omega = \frac{1}{2}\sqrt{2}$ sowie

$$\tilde{A} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Berechnung der orthogonalen Transformationsmatrix S:

Sei x Eigenvektor von $A + A^T$ zum Eigenwert c . Wegen $E_c = \mathbb{R}^4$ können wir x beliebig wählen, etwa $x = e_1$. Dann ist

$$A\,x = \frac{1}{9\sqrt{2}} \begin{bmatrix} 9 \\ -4 \\ -8 \\ -1 \end{bmatrix}.$$

Orthogonalisieren ergibt

$$x_1 = e_1\ ,\ \tilde{x}_2 = A\,x_1 - <Ax_1, x_1>\,x_1 = \frac{1}{9\sqrt{2}} \begin{bmatrix} 0 \\ -4 \\ -8 \\ -1 \end{bmatrix},$$

also

$$x_2 = \frac{\tilde{x}_2}{\|\tilde{x}_2\|} = \frac{1}{9} \begin{bmatrix} 0 \\ -4 \\ -8 \\ -1 \end{bmatrix}.$$

Nun muß $[x_1,x_2]^\perp$ bestimmt werden:

$$[x_1,x_2]^\perp = \{\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \mid a_1 = 0,\ 4\,a_2 + 8\,a_3 + a_4 = 0\} = [\begin{bmatrix} 0 \\ 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}].$$

In $[x_1,x_2]^\perp$ bestimmen wir ebenfalls eine ONB : Wir wählen $x_3 \in [x_1,x_2]^\perp$, $\| x_3 \| = 1$, beliebig, bilden $A\,x_3$ und orthogonalisieren:

$$x_3 = \frac{1}{3\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ -1 \\ 4 \end{bmatrix}, \quad \tilde{x}_4 = A\,x_3 - <Ax_3,x_3>\,x_3 = \frac{1}{9}\begin{bmatrix} 0 \\ 7 \\ -4 \\ 4 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 0 \\ 1 \\ -1 \\ 4 \end{bmatrix} = \frac{1}{18}\begin{bmatrix} 0 \\ 11 \\ -5 \\ -4 \end{bmatrix},$$

also

$$x_4 = \frac{\tilde{x}_4}{\|\tilde{x}_4\|} = \frac{1}{9\sqrt{2}}\begin{bmatrix} 0 \\ 11 \\ -5 \\ -4 \end{bmatrix}.$$

Damit erhalten wir

$$S = (x_1 \mid x_2 \mid x_3 \mid x_4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -4/9 & 1/3\sqrt{2} & 11/9\sqrt{2} \\ 0 & -8/9 & -1/3\sqrt{2} & -5/9\sqrt{2} \\ 0 & -1/9 & 4/3\sqrt{2} & -4/9\sqrt{2} \end{bmatrix}$$

§ 5 Unitäre Vektorräume

Wir wollen in diesem Paragraphen zunächst die Ergebnisse aus den vorigen Abschnitten 5.1 bis 5.4 soweit wie möglich auf komplexe Vektorräume übertragen. Ein Skalarprodukt β auf V muß, wenn $\sqrt{\beta(x,x)}$ wieder als Länge von x interpretiert werden soll, die gleichen Definitheitseigenschaften haben wie im reellen Fall. Andererseits bewirken Linearitätsforderungen, daß β eine Abbildung nach $\mathbb{C}$ sein muß. Damit kann β aber, im Gegensatz zum reellen Fall, keine symmetrische Bilinearform sein.

Definition. $\beta : V \times V \longrightarrow \mathbb{C}$ heißt *hermitesche Form*, wenn gilt:

(a) $\beta(a\,x + b\,x', y) = a\,\beta(x,y) + b\,\beta(x',y)$ für alle $a,b \in \mathbb{C}$, $x, x', y \in V$,

(b) $\beta(x,y) = \overline{\beta(y,x)}$ für alle $x,y \in V$.

Bemerkungen. **(a)** β ist im ersten Argument linear, im zweiten gilt für alle $a,b \in \mathbb{C}$ und alle $x, y, y' \in V$

$$\beta(x, a\,y + b\,y') = \overline{a}\,\beta(x,y) + \overline{b}\,\beta(x,y').$$

(b) Für alle $x \in V$ ist $\beta(x,x) \in \mathbb{R}$. Damit kann die positive Definitheit wie in 5.1 erklärt werden: β heißt *positiv definit*, wenn $\beta(x,x) > 0$ für alle $x \neq o$ gilt.

Definition. Eine positiv definite hermitesche Form β auf V heißt *Skalarprodukt* ; Schreibweise : $\langle \cdot, \cdot \rangle$. Das Paar $(V, \langle \cdot, \cdot \rangle)$ oder kurz V heißt *unitärer Vektorraum* oder auch (*komplexer*) *Vektorraum mit Skalarprodukt.*

Beispiel. $V = \mathbb{C}^n$, $\langle x,y \rangle := x_1\,\overline{y}_1 + \cdots + x_n\,\overline{y}_n = x^T\,\overline{y}$ ist Skalarprodukt, das *Standardskalarprodukt* auf $\mathbb{C}^n$. Auf der Teilmenge $\mathbb{R}^n$ ($\mathbb{R}^n$ ist kein Untervektorraum von $\mathbb{C}^n$!) geht $\langle \cdot, \cdot \rangle$ in das reelle Standardskalarprodukt auf $\mathbb{R}^n$ über.

Analog wie im Reellen lassen sich in unitären Vektorräumen nun die *Länge*, der *Abstand* und alle mit der *Orthogonalität* zusammenhängenden Begriffe einführen.

Beispiel. In $\mathbb{C}^n$ ist $\| x \|^2 = x_1\,\overline{x}_1 + \cdots + x_n\overline{x}_n = |x_1|^2 + \cdots + |x_n|^2 = \| \overline{x} \|^2$.

Der Satz 5.1 überträgt sich dann mit kleinen Modifikationen in der Formulierung und im Beweis. So lautet (e) jetzt

$$4 \langle x,y\rangle = \| x+y\|^2 - \| x-y\|^2 + i\| x+iy\|^2 - i\| x-iy\|^2 ,$$

denn es gilt

$$\| x+y\|^2 - \| x-y\|^2 = 2\langle x,y\rangle + 2\overline{\langle x,y\rangle}$$

und

$$\| x+iy\|^2 - \| x-iy\|^2 = 2\langle x,iy\rangle + 2\overline{\langle x,iy\rangle} = -2i\langle x,y\rangle + 2i\overline{\langle x,y\rangle} .$$

Im Satz von Pythagoras gilt nur die eine Richtung

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

Die Umkehrung ist falsch, wie das Beispiel $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $y = \begin{bmatrix} i \\ 0 \end{bmatrix}$ zeigt.

Wie in Satz 5.2 lassen sich in endlich dimensionalen unitären Vektorräumen V Skalarprodukte $\beta : V \times V \longrightarrow \mathbb{C}$ nach Auszeichnung einer Basis $B = (v_1,\ldots,v_n)$ durch Matrizen $A \in \mathbb{C}^{n\times n}$ mit

$$\beta(x,y) = \hat{x}^{\mathsf{T}} A \overline{\hat{y}}$$

beschreiben. Dabei sind $\hat{x}$ und $\hat{y}$ wieder die Koordinatendarstellungen von x und y bezüglich B und A ist die Matrix $(\!(\beta(v_i,v_j))\!)$ mit

$$A = \overline{A}^{\mathsf{T}} .$$

Solche Matrizen heißen *hermitesch*. Weiterhin ist A *positiv definit*, d.h. es gilt $x^{\mathsf{T}} A \overline{x} > 0$ für alle $x \in \mathbb{C}^n \setminus \{o\}$.

Die Sätze 6 und 7 (Gram–Schmidtsches Othogonalisierungsverfahren) aus § 5.2 übertragen sich wörtlich auf den komplexen Fall.

Für die Darstellung eines Vektors x bezüglich einer ONB $(x_1,\ldots,x_n)$ von V gilt wieder

$$x = \sum_{i=1}^{n} <x,x_i> x_i ,$$

für das Quadrat der Norm aber

$$\| x \|^2 = \sum_{i=1}^{n} |<x,x_i>|^2 .$$

Orthogonalprojektionen lassen sich wie im euklidischen Fall definieren und die Sätze 8, 9 und 10 gelten entsprechend für unitäre Vektorräume. Im Beweis von Satz 10 muß dabei $a \in \mathbb{C}$ gewählt und die Rechnung entsprechend modifiziert werden. Insbesondere besitzt also jeder endlich dimensionale unitäre Vektorraum eine Orthonormalbasis und Orthogonalprojektionen auf alle Untervektorräume.

Im Unterschied zum reellen Fall kann man einen endlich dimensionalen unitären Vektorraum V nicht wie in § 5.3 mit seinem Dualraum in natürlicher Weise identifizieren. Zwar gilt Satz 5.11 auch im Komplexen, im Beweis muß man nur

$$v = \sum_{i=1}^{n} \overline{x^*(x_i)}\, x_i$$

setzen, aber die Abbildung $\Psi : V \longrightarrow V^*$, $x \longmapsto <\cdot,x>$, ist jetzt nicht mehr linear.

Sind V und W unitäre Vektorräume, so wird die *adjungierte Abbildung* Φ^* einer Abbildung $\Phi \in \mathrm{Hom}\,(V,W)$ wieder über die Bedingung

$$<\Phi(v),w> = <v,\Phi^*(w)>$$

für alle $v \in V$, $w \in W$ erklärt. Der Existenzsatz 5.12 überträgt sich. Bei den Rechenregeln (S. 250) muß man (ii) durch

$$\text{(ii')} \qquad (a\,\Phi)^* = \overline{a}\,\Phi^* \ , \ a \in \mathbb{C} ,$$

ersetzen. Ist $\dim V = \dim W = n$, und ist A_Φ die Abbildungsmatrix von Φ bezüglich zweier Orthonormalbasen in V und W, so gilt

$$A_{\Phi^*} = \overline{A}_\Phi{}^{\mathsf{T}} .$$

Weil V und V^* bzw. W und W^* nicht mehr identifiziert werden können, muß man im komplexen Fall die adjungierte Abbildung Φ^* von der transponierten Abbildung $\Phi^\top$ unterscheiden.

Abbildungen $\Phi \in \mathrm{Hom}\,(V,V)$ mit $\Phi^* = \Phi$ bzw. $\Phi^* = -\Phi$ heißen wieder *selbstadjungiert* bzw. *antiselbstadjungiert.* Im Fall dim $V = n$ gilt dann für die Abbildungsmatrix einer selbstadjungierten Abbildung bezüglich einer Orthonormalbasis

$$A_\Phi = \overline{A}_\Phi{}^\top ,$$

d.h. sie ist hermitesch und für die Abbildungsmatrix einer antiselbstadjungierten Abbildung bezüglich einer Orthonormalbasis gilt

$$A_\Phi = -\overline{A}_\Phi{}^\top .$$

Matrizen mit dieser Eigenschaft heißen *schiefhermitesch.*

Isometrien werden wie im reellen Fall durch

$$<\Phi(x),\Phi(y)> = <x,y>$$

für alle $x,y \in V$ definiert; statt Isometrie ist auch der Name *unitäre Abbildung* üblich. Für Isometrien gilt wieder $\Phi^* = \Phi^{-1}$.

Für dim $V = n$ erfüllt die Abbildungsmatrix A_Φ einer unitären Abbildung $\Phi : V \longrightarrow V$ bezüglich einer ONB von V

$$A_\Phi \cdot \overline{A}_\Phi{}^\top = \overline{A}_\Phi{}^\top \cdot A_\Phi = E_n .$$

Solche Matrizen heißen *unitär.* Die unitären Matrizen bilden eine Untergruppe $U(n)$ der allgemeinen linearen Gruppe $GL(n,\mathbb{C})$ und die unitären Matrizen A mit det $A = 1$ bilden wiederum eine Untergruppe $SU(n)$ von $U(n)$. Die Gruppen $U(n)$ bzw. $SU(n)$ heißen die *unitäre* bzw. die *spezielle unitäre Gruppe.*

Satz 5.14 überträgt sich wörtlich, beim Beweis von (c) $\Rightarrow$ (a) muß man nur die entsprechende komplexe Darstellung von $<x,y>$ durch die Norm nehmen. In Satz 5.15 müssen in den Aussagen (e) und (f) die Abbildungsmatrizen A_Φ unitär sein.

In euklidischen und unitären Vektorräumen besitzt jeder Endomorphismus Φ, der selbstadjungiert oder antiselbstadjungiert oder eine surjektive Isometrie ist, eine adjungierte Abbildung Φ^*, die $\Phi \circ \Phi^* = \Phi^* \circ \Phi$ erfüllt. Wir nennen solche Endomorphismen normal.

Definition. Es seien V ein euklidischer oder unitärer Vektorraum und $\Phi \in \mathrm{End}\,(V)$. Φ heißt *normaler Endomorphismus* oder kurz *normal*, wenn Φ^* existiert und

$$\Phi \circ \Phi^* = \Phi^* \circ \Phi$$

erfüllt. Analog heißt die Matrix $A \in \mathbb{C}^{n\times n}$ *normal*, wenn

$$A \cdot \overline{A}^{\top} = \overline{A}^{\top} \cdot A$$

gilt.

Beispiele. **(a)** Wie schon bemerkt, sind selbstadjungierte und antiselbstadjungierte Endomorphismen sowie Isometrien normal.

(b) Symmetrische, schiefsymmetrische und orthogonale Matrizen $A \in \mathbb{R}^{n\times n}$ sind normal, ebenso hermitesche, schiefhermitesche und unitäre Matrizen $A \in \mathbb{C}^{n\times n}$.

Normale Abbildungen lassen sich auch noch durch andere Eigenschaften kennzeichnen.

Satz 17. *Es seien V ein euklidischer oder unitärer Vektorraum und $\Phi \in \mathrm{Hom}\,(V,V)$. Dann sind folgende Aussagen äquivalent:*

(a) *Φ ist normal.*

(b) *Φ^* existiert und für alle $x,y \in V$ gilt $\langle \Phi(x),\Phi(y)\rangle = \langle \Phi^*(x),\Phi^*(y)\rangle$.*

(c) *Φ^* existiert und für alle $x \in V$ gilt $\| \Phi(x) \| = \| \Phi^*(x) \|$.*

Für $\dim V = n$ sind die obigen Aussagen außerdem äquivalent zu:

(d) *Für jede ONB von V ist die Abbildungsmatrix A_Φ normal.*

(e) *Es gibt eine ONB von V, so daß die Abbildungsmatrix A_Φ normal ist.*

Beweis. (a) $\Longleftrightarrow$ (b): $\Phi^* \circ \Phi = \Phi \circ \Phi^* \iff \langle x,\Phi^* \circ \Phi(y)\rangle = \langle x,\Phi \circ \Phi^*(y)\rangle$ für alle $x,y \in V \iff \langle \Phi(x),\Phi(y)\rangle = \langle \Phi^*(x),\Phi^*(y)\rangle$ für alle $x,y \in V$.

(b) $\Longleftrightarrow$ (c): Die eine Richtung ist offensichtlich, die andere folgt aus der Darstellung des Skalarproduktes mit Hilfe der Norm.

(a) $\Longrightarrow$ (d): Dies gilt, weil die Abbildungsmatrix von Φ^* bezüglich einer beliebigen ONB von V $A_{\Phi^*} = \overline{A}_{\Phi}^{\,\tau}$ erfüllt und somit

$$A_{\Phi} \cdot \overline{A}_{\Phi}^{\,\tau} = A_{\Phi} \cdot A_{\Phi^*} = A_{\Phi \circ \Phi^*} = A_{\Phi^* \circ \Phi} = A_{\Phi^*} \cdot A_{\Phi} = \overline{A}_{\Phi}^{\,\tau} \cdot A_{\Phi}$$

gilt.

(d) $\Longrightarrow$ (e): offensichtlich.

(e) $\Longrightarrow$ (a): Für die Abbildungsmatrix von Φ^* bezüglich der gegebenen ONB von V gilt $A_{\Phi^*} = \overline{A}_{\Phi}^{\,\tau}$ und aus der Normalität von A_{Φ} folgt dann analog zu oben $\Phi \circ \Phi^* = \Phi^* \circ \Phi$. ∎

Bemerkungen. **(a)** Sei Φ normal. Dann gilt Kern Φ = Kern Φ^*.

Dies folgt unmittelbar aus Satz 17(c).

(b) Sei Φ normal. Dann gilt $\Phi(x) = c\,x$ genau dann, wenn $\Phi^*(x) = \overline{c}\,x$ gilt, d.h. die Eigenräume E_c von Φ und $E_{\overline{c}}$ von Φ^* sind gleich.

Beweis. $\Phi - c\,\mathrm{id}_V$ ist normal, denn es ist

$$(\Phi - c\,\mathrm{id}_V) \circ (\Phi - c\,\mathrm{id}_V)^* = (\Phi - c\,\mathrm{id}_V) \circ (\Phi^* - \overline{c}\,\mathrm{id}_V)$$

$$= \Phi \circ \Phi^* - \overline{c}\,\Phi - c\,\Phi^* + c\,\overline{c}\,\mathrm{id}_V = \Phi^* \circ \Phi - c\,\Phi^* - \overline{c}\,\Phi + \overline{c}\,c\,\mathrm{id}_V$$

$$= (\Phi - c\,\mathrm{id}_V)^* \circ (\Phi - c\,\mathrm{id}_V) .$$

Damit folgt mit Bemerkung (a) und den Rechenregeln für adjungierte Abbildungen

$$\mathrm{Kern}\,(\Phi - c\,\mathrm{id}_V) = \mathrm{Kern}\,(\Phi - c\,\mathrm{id}_V)^* = \mathrm{Kern}\,(\Phi^* - \overline{c}\,\mathrm{id}_V)$$

und daraus die Behauptung. ∎

(c) Bei einem normalen Endomorphismus Φ sind die Eigenvektoren zu verschiedenen Eigenwerten stets orthogonal.

Beweis. Seien $\Phi(x) = a\,x$, $\Phi(y) = b\,y$ und $a \neq b$. Dann folgt $a <x,y> = <\Phi(x),y> = <x,\Phi^*(y)> = <x,\overline{b}\,y> = b <x,y>$, also $<x,y> = 0$. ∎

Als Hauptergebnis leiten wir nun auch für normale Abbildungen bzw. Matrizen eine Normalform her. Hierzu müssen wir aber zwischen dem komplexen und dem reellen Fall unterscheiden.

Satz 18. *Es seien V ein n–dimensionaler unitärer Vektorraum und $\Phi \in \mathrm{End}\,(V)$. Genau dann ist Φ normal, wenn es in V eine Orthonormalbasis aus Eigenvektoren von Φ gibt.*

Beweis. Sei $(x_1,\ldots x_n)$ eine ONB aus Eigenvektoren von Φ und $\Phi(x_i) = c_i\, x_i$. Dann gilt $\Phi^* \circ \Phi(x_i) = \Phi^*(c_i\, x_i) = c_i\, \overline{c}_i\, x_i = \Phi(\overline{c}_i\, x_i) = \Phi \circ \Phi^*(x_i)$, $i = 1,\ldots,n$, also $\Phi \circ \Phi^* = \Phi^* \circ \Phi$, d.h. Φ ist normal.

Sei umgekehrt Φ normal. Nach dem Fundamentalsatz der Algebra besitzt Φ wenigstens einen Eigenwert c. Sei x_1 ein zugehöriger Eigenvektor mit $\| x_1 \| = 1$. Wir führen den weiteren Beweis nun mit vollständiger Induktion nach n :

$n = 1$: Hier ist $B = \{x_1\}$ ONB von V.

$n-1 \longrightarrow n$: Sei $U = \{x_1\}^{\perp}$. Dann ist dim $U = n-1$, und es gilt $\Phi(x_1) \perp U$. Wegen $\Phi^*(x_1) = \overline{c}\, x_1$ folgt auch $\Phi^*(x_1) \perp U$. Somit gilt für alle $y \in U$, daß $\langle \Phi(y), x_1 \rangle = \langle y, \Phi^*(x_1) \rangle = 0$ ist, also $\Phi(U) \subset U$. Analog ergibt sich $\Phi^*(U) \subset U$. Damit ist $\Phi|_U$ ein normaler Endomorphismus von U, und nach Induktionsvoraussetzung existiert in U eine ONB $\{x_2,\ldots,x_n\}$ aus Eigenvektoren von $\Phi|_U$, also auch von Φ. Die Basis $\{x_1,\ldots,x_n\}$ ist somit ONB in V. ∎

Im reellen Fall (Satz 13) hatten wir aus der Existenz einer ONB von Eigenvektoren schließen können, daß $\Phi = \Phi^*$ ist. Hier, im komplexen Fall, erhalten wir nur die schwächere Aussage $\Phi \circ \Phi^* = \Phi^* \circ \Phi$. Der Grund ist, daß das analoge Vorgehen zum Beweis von Satz 13 auf die Gleichung $\langle \Phi(x_i), x_j \rangle = \overline{\langle x_i, \Phi(x_j) \rangle}$ führt, woraus sich nicht mehr $\Phi = \Phi^*$ folgern läßt.

Bemerkung. Die zu Satz 18 äquivalente Aussage für Matrizen $A \in \mathbb{C}^{n\times n}$ lautet:

Genau dann ist $A \in \mathbb{C}^{n\times n}$ normal, wenn es eine unitäre Matrix S gibt, so daß $\overline{S}^{\mathsf{T}} A\, S$ Diagonalgestalt hat.

Damit ergeben sich speziell die Normalformen hermitescher bzw. schiefher-

mitescher bzw. unitärer Matrizen. Alle haben Diagonalgestalt, wobei im ersten Fall alle Eigenwerte reell sind, im zweiten Fall haben alle Eigenwerte den Realteil 0 und im dritten Fall haben sie alle Betrag 1 (Beweis als Übungsaufgabe).

Außerdem sei bemerkt, daß sich nun auch die Sätze 3 bis 5 auf hermitesche Matrizen übertragen lassen. Insbesondere gibt es auch für positiv definite hermitesche Matrizen A eine Cholesky–Zerlegung in der Form $A = \overline{B}^{\top} B$ mit einer regulären oberen Dreiecksmatrix $B \in \mathbb{C}^{n\times n}$ (Beweis als Übungsaufgabe).

In euklidischen Vektorräumen ist die Situation anders. Hier gilt:

Satz 19. *Es seien V ein n–dimensionaler euklidischer Vektorraum und $\Phi \in \mathrm{Hom}\,(V,V)$. Genau dann ist Φ normal, wenn es in V eine Orthonormalbasis gibt bezüglich der die Abbildungsmatrix von Φ die Form*

$$(*) \qquad \tilde{A} = \begin{bmatrix} c_1 & & & & & \\ & \ddots & & & & \\ & & c_k & & & \\ & & & \boxed{\begin{matrix} a_1 & b_1 \\ -b_1 & a_1 \end{matrix}} & & \\ & & & & \ddots & \\ & & & & & \boxed{\begin{matrix} a_m & b_m \\ -b_m & a_m \end{matrix}} \end{bmatrix}$$

mit c_i, a_j, $b_j \in \mathbb{R}$ hat. Dabei sind $c_1,\ldots,c_k$ die reellen Nullstellen und $a_j + i\,b_j$, $a_j - i\,b_j$, $b_j > 0$, $j = 1,\ldots,m$, die restlichen komplexen Nullstellen des charakteristischen Polynoms von Φ.

Wir beweisen die zu Satz 19 äquivalente Aussage für reelle Matrizen.

Satz 20. *Genau dann ist $A \in \mathbb{R}^{n\times n}$ normal, wenn es eine orthogonale Matrix $S \in \mathbb{R}^{n\times n}$ gibt, so daß $S^{\top} A\, S$ die Form (*) mit c_i, a_j, $b_j \in \mathbb{R}$ hat. Dabei sind $c_1,\ldots,c_k$ die reellen Nullstellen und $a_j + i\,b_j$, $a_j - i\,b_j$, $b_j > 0$, $j = 1,\ldots,m$, die restlichen komplexen Nullstellen des charakteristischen Polynoms von A.*

Beweis. Gibt es eine orthogonale Matrix S, so daß $\tilde{A} = S^{\top} A\, S$ die Form (*) hat, so folgt $A\, A^{\top} = S\, \tilde{A}\, \tilde{A}^{\top}\, S^{\top} = S\, \tilde{A}^{\top}\, \tilde{A}\, S^{\top} = A^{\top} A$.

Ist umgekehrt A normal, so gibt es nach dem letzten Satz eine ONB von $\mathbb{C}^n$ aus komplexen Eigenvektoren von A und $\mathbb{C}^n$ ist direkte Summe paarweise orthogonaler Eigenräume von A.

Für jeden reellen Eigenwert c konstruieren wir in dem zugehörigen (reellen) Eigenraum E_c eine ONB. Da die komplexen Eigenräume von A orthogonal zueinander sind, sind es auch die reellen. Also erhalten wir ein Orthonormalsystem, das gerade die ersten k Spalten von $\tilde{A}$ liefert.

Für jeden komplexen Eigenwert $c = a + i\, b$, $b \neq 0$, ist auch $\overline{c} = a - i\, b$ ein Eigenwert von A und wegen $c \neq \overline{c}$ gilt $E_c \perp E_{\overline{c}}$. Wir können o.B.d.A. $b > 0$ annehmen. In E_c konstruieren wir eine ONB $(z_1,\ldots,z_q)$. Dann ist $(\overline{z}_1,\ldots,\overline{z}_q)$ eine ONB von $E_{\overline{c}}$ und somit $(z_1, \overline{z}_1,\ldots,z_q, \overline{z}_q)$ eine ONB von $E_c \oplus E_{\overline{c}}$. Setzen wir für $j = 1,\ldots,q$

$$x_j = \operatorname{Re} z_j = \tfrac{1}{2}(z_j + \overline{z}_j),$$

$$y_j = \operatorname{Im} z_j = \tfrac{1}{2i}(z_j - \overline{z}_j),$$

so gilt $[x_j, y_j] = [z_j, \overline{z}_j]$, und wir erhalten die Darstellung

$$E_c \oplus E_{\overline{c}} = [x_1,y_1] \oplus \ldots \oplus [x_q,y_q]$$

mit paarweise orthogonalen zweidimensionalen Unterräumen $[x_j, y_j]$. Ferner gilt

$$A\, x_j = a\, x_j - b\, y_j$$

$$A\, y_j = b\, x_j + a\, y_j$$

für $j = 1,\ldots,q$. Wegen

$$\langle z_j + \overline{z}_j, z_j - \overline{z}_j\rangle = \langle z_j, z_j\rangle + \langle \overline{z}_j, z_j\rangle - \langle z_j, \overline{z}_j\rangle - \langle \overline{z}_j, \overline{z}_j\rangle = \| z_j\|^2 - \| \overline{z}_j\|^2 = 0$$

und

$$\langle z_j + \overline{z}_j, z_j - \overline{z}_j\rangle = \langle 2\, x_j, 2\, i\, y_j\rangle = -4\, i \langle x_j, y_j\rangle$$

folgt $\langle x_j, y_j\rangle = 0$ für $j = 1,\ldots,q$. Damit ergibt sich

$$0 = \langle z_j , \overline{z}_j \rangle = \langle x_j + i\, y_j , x_j - i\, y_j \rangle = \| x_j \|^2 - \| y_j \|^2 ,$$

und somit

$$\| x_j \| = \| y_j \| .$$

Also erhalten wir für jedes $j = 1,\ldots,q$ mit

$$\frac{x_j}{\| x_j \|} , \frac{y_j}{\| y_j \|}$$

eine ONB zu dem Zweier–Kästchen

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

Insgesamt erhalten wir so eine reelle ONB von $\mathbb{C}^n$ bezüglich der A die gewünschte Normalform (*) hat. ∎

Bemerkungen. **(a)** Die Matrix $\tilde{A}$ ist die reelle Jordansche Normalform (Satz 4.22) der normalen Matrix A.

(b) Aus Satz 20 ergibt sich für symmetrische Matrizen nochmals Satz 13, für schiefsymmetrische Matrizen $A \in \mathbb{R}^{n\times n}$ erhalten wir als Normalform

$$\tilde{A} = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & O \\ & & & \begin{bmatrix} 0 & b_1 \\ -b_1 & 0 \end{bmatrix} & \\ & & & & \ddots \\ & O & & & \begin{bmatrix} 0 & b_m \\ -b_m & 0 \end{bmatrix} \end{bmatrix}$$

und für orthogonale Matrizen ergibt sich aus Satz 20 der Satz 16 (vgl. Bem. (c) von

S. 261). Im letzteren Fall gilt ja $c_i = \pm 1$ für $i = 1,...,k$ und $a_j^2 + b_j^2 = 1$, $b_j > 0$, für $j = 1,...,m$. Satz 16 erhält man, indem man sin $\omega_j = b_j$ setzt und für jedes Zweier–Kästchen die zugehörigen Basisvektoren x_j, y_j vertauscht oder y_j durch $-y_j$ ersetzt (vgl. Bem. (a) von S. 260).

Aus dem oben gesagten ergibt sich damit eine weitere Methode, die Normalform und die Transformationsmatrix einer reellen Isometrie bzw. einer orthogonalen Matrix $A \in \mathbb{R}^{n\times n}$ aufzustellen:

Zu den reellen Eigenwerten 1 bzw. –1 bestimmen wir wie üblich reelle Orthonormalbasen in den zugehörigen Eigenräumen. Zu jedem komplexen Eigenwert $c = a + i\,b$, $b > 0$, sei $z_1,...,z_q$ ein Orthonormalsystem aus komplexen Eigenvektoren. Dann ist $\overline{z}_1,...,\overline{z}_q$ ein Orthonormalsystem aus komplexen Eigenvektoren zu $\overline{z}$. Setzen wir sin $\omega = b$ und

$$x_j = \frac{\operatorname{Re} z_j}{\|\operatorname{Re} z_j\|}\,, \quad y_j = \frac{\operatorname{Im} z_j}{\|\operatorname{Im} z_j\|}\,, \quad j = 1,...,q\,,$$

so ist $(y_1\,, x_1\,,...,\, y_q\,, x_q)$ oder $(x_1, -y_1,...,x_q\,, -y_q)$ eine reelle ONB zu den q Zweierkästchen

$$\begin{bmatrix} \cos\,\omega & -\sin\,\omega \\ \sin\,\omega & \cos\,\omega \end{bmatrix}$$

Beispiel. Die Matrix

$$A = \begin{bmatrix} \sqrt{3}/4 + 1/2 & \sqrt{3}/4 - 1/2 & -\sqrt{2}/4 \\ \sqrt{3}/4 - 1/2 & \sqrt{3}/4 + 1/2 & -\sqrt{2}/4 \\ \sqrt{2}/4 & \sqrt{2}/4 & \sqrt{3}/2 \end{bmatrix}$$

ist orthogonal (Übungsaufgabe). Für das charakteristische Polynom p von A gilt

$$p = \det(A - X\,E_3) = \cdots = -X^3 + (1+\sqrt{3})\,X^2 - (1+\sqrt{3})\,X + 1\,.$$

Die Nullstellen sind $c_1 = 1$ sowie $c = \sqrt{3}/2 + i/2$ und $\overline{c} = \sqrt{3}/2 - i/2$. Die Normalform lautet also

$$\tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 \\ 0 & 1/2 & \sqrt{3}/2 \end{bmatrix}$$

(Drehwinkel : $\omega = \pi/6$, Drehachse : E_c, Drehebene : $E_{c_1}^{\perp}$).

Bestimmung der Transformationsmatrix S :

$$v \in E_{c_1} \iff (A - E_3)\, v = o \iff v = \begin{bmatrix} t \\ -t \\ 0 \end{bmatrix},\ t \in \mathbb{R}.$$

Also wählen wir

$$v := 1/\sqrt{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \in E_{c_1}.$$

$$z \in E_c \iff (A - (\sqrt{3}/2 + i/2)\, E_3)\, z = o \iff z = \begin{bmatrix} t \\ t \\ -\sqrt{2}\, i\, t \end{bmatrix},\ t \in \mathbb{R}.$$

Dann ist

$$z = \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2}\, i \end{bmatrix} \in E_c\ ,\ \overline{z} = \begin{bmatrix} 1 \\ 1 \\ \sqrt{2}\, i \end{bmatrix} \in E_{\overline{c}},$$

und das obige Verfahren liefert die orthonormierten Vektoren

$$\frac{\operatorname{Re} z}{\|\operatorname{Re} z\|} = 1/\sqrt{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\ ,\ \frac{\operatorname{Im} z}{\|\operatorname{Im} z\|} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

Damit ergibt sich als orthogonale Transformationsmatrix

$$S = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & -1 & 0 \end{bmatrix} \text{ bzw. } S = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Kapitel 6 Affine und euklidische Geometrie

§ 1 Affine und euklidische Räume

Bisher haben wir die geometrischen Objekte Punkte, Geraden, Ebenen usw. nur als Elemente bzw. Teilmengen eines Vektorraumes V kennengelernt. Die Doppelrolle der Elemente von V als Vektoren und als Punkte bringt aber gelegentlich Schwierigkeiten mit sich und entspricht auch nicht unserer geometrischen Vorstellung in der Zeichenebene oder im Anschauungsraum. Dort unterscheiden wir sehr wohl zwischen den "Punkten" und den zugehörigen "Verbindungsvektoren".

Wir werden daher im folgenden eine neue Struktur betrachten, den affinen Raum, der zur geometrischen Beschreibung des Anschauungsraumes besser geeignet ist, weil in ihm die Begriffe "Punkt" und "Vektor" getrennt behandelt werden. Wenn wir die Beziehungen zwischen affinen Räumen $\mathbb{A}$ und Vektorräumen V geklärt haben, können wir die bisher entwickelten Methoden und Ergebnisse der linearen Algebra auf geometrische Fragestellungen anwenden. Zur Vereinfachung der Darstellung werden wir dabei, nach Auszeichnung eines Ursprungs, in manchen Fällen wieder $\mathbb{A}$ und V identifizieren.

Definition. Gegeben seien eine nichtleere Menge $\mathbb{A}$, deren Elemente *Punkte* heißen, ein $\mathbb{K}$–Vektorraum V, sowie eine Abbildung $f\colon \mathbb{A} \times \mathbb{A} \longrightarrow V$. Das Tripel $(\mathbb{A}, V, f)$ heißt *affiner Punktraum* oder *affiner Raum mit zugehörigem Vektorraum* V, falls folgende Bedingungen erfüllt sind:

(a) Zu jedem $P \in \mathbb{A}$ und jedem $x \in V$ gibt es genau ein $Q \in \mathbb{A}$ mit $f(P,Q) = x$.

(b) Für alle $P, Q, R \in \mathbb{A}$ gilt $f(P,Q) + f(Q,R) = f(P,R)$.

Statt $(\mathbb{A}, V, f)$ schreiben wir kurz $\mathbb{A}$ und sprechen vom affinen Raum $\mathbb{A}$, wenn der zugehörige Vektorraum V und die Abbildung f klar sind.

Unter der *Dimension* von $\mathbb{A}$ verstehen wir die Dimension des zugehörigen Vektorraumes V, Schreibweise: dim $\mathbb{A}$.

Ist $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{K} = \mathbb{C}$, so sprechen wir von einem *reellen* bzw. *komplexen* affinen Raum. Ist der Vektorraum V euklidisch, so heißt $\mathbb{A}$ *euklidischer Raum.*

Die Gesamtheit der Definitionen und Sätze, die sich aus den Axiomen für einen affinen Raum herleiten lassen, heißt *affine Geometrie.* Analog sprechen wir auch von *euklidischer Geometrie.* Jeder Satz der affinen Geometrie ist auch ein Satz der euklidischen Geometrie, umgekehrt lassen sich aber viele Sätze der euklidischen Geometrie affin gar nicht formulieren. Dies gilt zum Beispiel für den aus der Schule bekannten Satz, daß sich in jedem Dreieck die Höhen in einem gemeinsamen Punkt schneiden.

In den folgenden Paragraphen werden wir zunächst immer erst den gemeinsamen affinen Anteil der Geometrie darstellen und danach zusätzliche Aussagen im Rahmen der euklidischen Geometrie beweisen. Wir beginnen mit einigen Bemerkungen zum Axiomensystem eines affinen Raumes.

Bemerkungen. **(a)** Die Existenz der Abbildung f bedeutet, daß je zwei Punkte P und Q genau einen *Verbindungsvektor* $f(P,Q)$ besitzen. Um dies auch in der Schreibweise deutlich zu machen, schreiben wir deshalb statt $f(P,Q)$ meistens $\overrightarrow{PQ}$.

Da f im allgemeinen nicht injektiv ist, gibt es mehrere Punktepaare mit dem gleichen Verbindungsvektor. Die Vektoren spielen in dem affinen Raum gerade die Rolle, die die Äquivalenzklassen der Pfeile im Anschauungsraum spielten.

(b) Axiom (a) besagt, daß der Vektor x im Punkt P "abgetragen" zu einem eindeutig bestimmten Punkt Q führt, und Axiom (b) bedeutet, daß das Abtragen von Vektoren an Punkten mit der Vektoraddition verträglich ist.

(c) Einige einfache Folgerungen sind:

$$\overrightarrow{PQ} = o \iff P = Q\,,$$

$$\overrightarrow{PQ} = -\overrightarrow{QP}\,,$$

$$\overrightarrow{PQ} = \overrightarrow{PR} \Longrightarrow Q = R \;;\; \overrightarrow{QS} = \overrightarrow{RS} \Longrightarrow Q = R\,.$$

Beispiele. **(a)** Wir können den Anschauungsraum als affinen Raum auffassen. Der zugehörige Vektorraum ist dann der auf S.93 betrachtete Vektorraum der Pfeilklassen.

(b) Jeder $\mathbb{K}$–Vektorraum V wird selbst zu einem affinen Raum mit zugehörigem Vektorraum V, wenn wir für beliebige Punkte $x,y \in V$ den Verbindungsvektor durch $\overrightarrow{xy} := y - x$ definieren. Insbesondere ist $\mathbb{K}^n$ ein affiner Raum.

So wie wir jeden Vektorraum als affinen Raum betrachten können, gehört umgekehrt in einem affinen Raum $\mathbb{A}$ nach Auszeichnung eines beliebigen Punktes $O \in \mathbb{A}$ zu jedem Punkt X ein eindeutig bestimmter Vektor $\overrightarrow{OX}$. Es gilt $V = \{\overrightarrow{OX} \mid X \in \mathbb{A}\}$, und die durch $X \longmapsto \overrightarrow{OX}$ erklärte Abbildung von $\mathbb{A}$ auf V ist bijektiv. Wir nennen $\overrightarrow{OX}$ den *Ortsvektor* des Punktes X bezüglich des *Ursprungs* O. Affine Räume und Vektorräume unterscheiden sich also im wesentlichen nur durch die Auszeichnung eines Ursprungs.

Analog wollen wir nun die affinen Unterräume von $\mathbb{A}$ so definieren, daß sie sich nur durch Auszeichnung eines Ursprunges von den Untervektorräumen von V unterscheiden.

Definition. Eine nichtleere Teilmenge L eines affinen Raumes $\mathbb{A}$ heißt *affiner Unterraum* von $\mathbb{A}$, wenn es einen Punkt $P \in L$ gibt, so daß die Menge $\{\overrightarrow{PX} \mid X \in L\}$ ein Untervektorraum von V ist.

Die affinen Unterräume eines euklidischen Raumes heißen *euklidische Unterräume.*

Bemerkungen und Bezeichnungen. **(a)** Ist L affiner Unterraum von $\mathbb{A}$, so ist für jeden Punkt $Q \in L$ die Menge $\{\overrightarrow{QX} \mid X \in L\}$ ein Untervektorraum von V und es gilt

$$\{\overrightarrow{QX} \mid X \in L\} = \{\overrightarrow{XY} \mid X,Y \in L\}.$$

Somit gehört zu L in eindeutiger Weise der Untervektorraum $U_L = \{\overrightarrow{XY} \mid X,Y \in L\}$, der dann *Richtungsraum* oder kurz *Richtung* von L heißt.

Beweis. Nach Definition gibt es einen Punkt $P \in L$, so daß $U = \{\overrightarrow{PX} \mid X \in L\}$ ein Untervektorraum von V ist. Sei $W = \{\overrightarrow{QX} \mid X \in L\}$. Wegen $Q \in L$ ist $\overrightarrow{QP} = -\overrightarrow{PQ} \in U$, also gilt $W = \overrightarrow{QP} + U = U$, und W ist ebenfalls ein Untervektorraum. Daß die Menge $\{\overrightarrow{XY} \mid X, Y \in L\}$ mit dem Untervektorraum W übereinstimmt, ist offensichtlich. ∎

(b) Ist L affiner Unterraum von $\mathbb{A}$ und ist f_L die auf $L \times L$ eingeschränkte Abbildung f, so erfüllt das Tripel (L, U_L, f_L) die Definition eines affinen Raumes. Damit gilt $\dim L = \dim U_L$.

Die nulldimensionalen affinen Unterräume von $\mathbb{A}$ sind genau die Mengen $\{P\}$, $P \in \mathbb{A}$, bestehen also aus einem einzigen Punkt. Man nennt sie deshalb ebenfalls *Punkte*. Die eindimensionalen affinen Unterräume heißen *Geraden*, die zweidimensionalen affinen Unterräume heißen *Ebenen*.

(c) Jeder affine Unterraum $L \subset \mathbb{A}$ läßt sich nach Auszeichnung eines Ursprungs $O \in \mathbb{A}$ und eines Punktes $P \in L$ vektoriell in der Form

$$\{\overrightarrow{OX} \mid X \in L\} = \overrightarrow{OP} + U_L$$

darstellen.

Speziell für $\mathbb{A} = V$ und $O = o$ gilt: Jeder affine Unterraum $L \subset V$ ist von der Form $L = x_0 + U$ mit einen Untervektorraum $U \subset V$ und $x_0 \in L$ (vgl. S. 126).

Wie in § 2.5 ist der Durchschnitt affiner Unterräume entweder leer oder ein affiner Unterraum.

Satz 1. *Es seien $\mathbb{A}$ ein affiner Raum und $\mathcal{M}$ eine nichtleere Menge affiner Unterräume von $\mathbb{A}$. Dann ist der Schnitt*

$$M = \bigcap_{L \in \mathcal{M}} L$$

entweder leer oder ein affiner Unterraum mit Richtungsraum

$$U_M = \bigcap_{L \in \mathcal{M}} U_L .$$

Beweis. Sei $M \neq \emptyset$ und $P \in M$. Dann gilt

$$U_M = \{\overrightarrow{PX} \mid X \in M\} = \bigcap_{L \in \mathcal{M}} \{\overrightarrow{PX} \mid X \in L\} = \bigcap_{L \in \mathcal{M}} U_L . \qquad \blacksquare$$

Definition. Ist $C \subset \mathbb{A}$ nichtleer, so heißt der Durchschnitt aller affinen Unterräume L, die C enthalten, die *affine Hülle* von C.

Ist $C = \{P_1,...,P_k\}$, so nennen wir die affine Hülle von C den *Verbindungsraum* der Punkte $P_1,...,P_k$ und schreiben dafür $P_1 \vee ... \vee P_k$. Allgemeiner heißt die affine Hülle der Vereinigungsmenge endlich vieler affiner Unterräume $L_1,...,L_k$ der *Verbindungsraum* von $L_1,...,L_k$, Schreibweise: $L_1 \vee ... \vee L_k$.

Auch der Verbindungsraum endlich vieler affiner Unterräume läßt sich genauer beschreiben.

Satz 2. *Es seien $L_1,...,L_k \subset \mathbb{A}$ affine Unterräume und $P_i \in L_i$ beliebig gewählt, $i = 1,...,k$. Dann gilt für den Richtungsraum des Verbindungsraumes $L = L_1 \vee ... \vee L_k$*

$$U_L = U_{L_1} + \cdots + U_{L_k} + [\overrightarrow{P_1P_2},...,\overrightarrow{P_1P_k}] .$$

Ist $L_1 \cap ... \cap L_k \neq \emptyset$, so gilt

$$U_L = U_{L_1} + \cdots + U_{L_k} .$$

Beweis. Seien $\tilde{U} = U_{L_1} + \cdots + U_{L_k} + [\overrightarrow{P_1P_2},...,\overrightarrow{P_1P_k}]$, $\tilde{L} = \{X \in \mathbb{A} \mid \overrightarrow{P_1X} \in \tilde{U}\}$. Aus $X \in L_i$ folgt $\overrightarrow{P_1X} = \overrightarrow{P_1P_i} + \overrightarrow{P_iX} \in \tilde{U}$, also $X \in \tilde{L}$, $i = 1,...,k$. Somit gilt $L \subset \tilde{L}$ und $U_L \subset \tilde{U}$. Die umgekehrte Inklusion ist offensichtlich.

Ist $L_1 \cap ... \cap L_k \neq \emptyset$, so wählen wir $P \in L_1 \cap ... \cap L_k$, setzen $P_1 = ... = P_k = P$ und erhalten $U_L = U_{L_1} + \cdots + U_{L_k}$. $\blacksquare$

Beispiele. **(a)** Der Verbindungsraum von $k+1$ Punkten $P_0,...,P_k$ besitzt den Richtungsraum $U = [\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}]$, und es gilt $\dim U \leq k$.

Für $k = 1$ und $P_0 \neq P_1$ gilt $\dim U = 1$, also ist $P_0 \vee P_1$ eine Gerade; wir schreiben dann kürzer P_0P_1 statt $P_0 \vee P_1$.

Für $k = 2$ und drei verschiedene Punkte P_0, P_1, P_2, die nicht auf einer Geraden liegen, gilt dim $U = 2$, also ist der Verbindungsraum $P_0 \vee P_1 \vee P_2$ eine Ebene.

(b) Im $\mathbb{R}^3$ seien der Punkt x und die Gerade L gegeben mit

$$x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ und } L = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \left[\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right].$$

Der Verbindungsraum $x \vee L$ hat die Darstellung

$$x \vee L = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \left(\left[\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right] + \left[\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right] \right) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \left[\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right].$$

(c) Im $\mathbb{R}^5$ seien die Punkte x_1, x_2, x_3, x_4, x_5 gegeben mit

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, x_2 = \begin{bmatrix} 3 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}, x_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, x_4 = \begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \\ -2 \end{bmatrix}, x_5 = \begin{bmatrix} 0 \\ 5 \\ -3 \\ -4 \\ 0 \end{bmatrix}.$$

Für ihren Verbindungsraum gilt

$$x_1 \vee \cdots \vee x_5 = x_1 + [x_2 - x_1, x_3 - x_1, x_4 - x_1, x_5 - x_1]$$

$$= \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \left[\begin{bmatrix} 2 \\ 6 \\ -4 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -6 \\ 4 \\ 6 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -2 \\ -3 \\ 1 \end{bmatrix} \right].$$

Wir wollen im Richtungsraum von $x_1 \vee \cdots \vee x_5$ eine Basis angeben. Dazu verwenden wir wieder das Verfahren von S.114 und erhalten

$$x_1 \vee \cdots \vee x_5 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \left[\begin{bmatrix} -1 \\ 3 \\ 0 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 12 \\ 0 \\ -11 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right].$$

Also ist $x_1 \vee \cdots \vee x_5$ ein dreidimensionaler affiner Unterraum von $\mathbb{R}^5$.

Für den Durchschnitt und den Verbindungsraum zweier affiner Unterräume gilt die folgende Dimensionsformel.

Satz 3. *Es seien $\mathbb{A}$ ein affiner Raum und L_1, L_2 affine Unterräume von $\mathbb{A}$ mit zugehörigen Richtungsräumen U_1, U_2. Dann gilt für $L_1 \cap L_2 \neq \emptyset$*

$$\dim L_1 + \dim L_2 = \dim (L_1 \vee L_2) + \dim (L_1 \cap L_2),$$

und für $L_1 \cap L_2 = \emptyset$ ist

$$\dim L_1 + \dim L_2 = \dim (L_1 \vee L_2) + \dim (U_1 \cap U_2) - 1.$$

Beweis. Nach Satz 2.19 gilt $\dim L_1 + \dim L_2 = \dim U_1 + \dim U_2 = \dim (U_1 \cap U_2) + \dim (U_1 + U_2)$. Im Fall $L_1 \cap L_2 \neq \emptyset$ ist nach Satz 2 $U_1 + U_2$ der Richtungsraum von $L_1 \vee L_2$. Da $U_1 \cap U_2$ dann auch der Richtungsraum von $L_1 \cap L_2$ ist, folgt der 1. Teil der Behauptung.

Ist $L_1 \cap L_2 = \emptyset$, so wählen wir $P_1 \in L_1$, $P_2 \in L_2$ und nach Satz 2 gilt $\dim (L_1 \vee L_2) = \dim (U_1 + U_2 + [\overrightarrow{P_1P_2}])$. Wäre $\overrightarrow{P_1P_2} \in U_1 + U_2$, so gäbe es Punkte $Q_1 \in L_1$, $Q_2 \in L_2$ mit $\overrightarrow{P_1P_2} = \overrightarrow{P_1Q_1} + \overrightarrow{P_2Q_2}$. Also wäre $\overrightarrow{P_2Q_2} = \overrightarrow{Q_1P_2}$ und somit $Q_1 \in L_1 \cap L_2$. Daher gilt $\dim (U_1 + U_2 + [\overrightarrow{P_1P_2}]) = \dim (U_1 + U_2) + 1$, woraus die 2. Behauptung folgt. ∎

Bezeichnung und Bemerkung. Mit Hilfe des Verbindungsraumes läßt sich eine weitere Klasse affiner Unterräume definieren: Es sei L ein von $\mathbb{A}$ verschiedener affiner Unterraum. Gibt es einen Punkt $P \in \mathbb{A}$, so daß der Verbindungsraum von L und P gleich $\mathbb{A}$ ist, so heißt L eine *Hyperebene* von $\mathbb{A}$. Nach Satz 2 kann P nicht in L liegen. Ist $\mathbb{A}$ n–dimensional, so sind wegen Satz 3 die Hyperebenen genau die affinen Unterräume der Dimension $n-1$.

Affine Unterräume können parallel zueinander sein. Wie auf S. 129 nennen wir zwei affine Unterräume L_1, L_2 *parallel*, Schreibweise $L_1 \parallel L_2$, falls für die zugehörigen

Richtungsräume U_1 , U_2 gilt: $U_1 \subset U_2$ oder $U_2 \subset U_1$. Weiterhin heißen zwei Geraden *windschief*, falls sie weder parallel sind noch einen Punkt gemeinsam haben.

Beispiel. In einem n–dimensionalen affinen Raum $\mathbb{A}$ kann eine Gerade g in bezug auf eine Hyperebene L folgende Lagen einnehmen: g ist entweder Teilmenge von L oder g schneidet L in einem Punkt oder g ist parallel zu L und der Schnitt ist leer (Übungsaufgabe).

Als nächstes wollen wir den Begriff der linearen Unabhängigkeit auf affine Räume übertragen.

Definition. Es seien $\mathbb{A}$ ein affiner Raum, $k \in \mathbb{N}_0$ und $P_0,...,P_k \in \mathbb{A}$. Die Punkte $P_0,...,P_k$ heißen *affin unabhängig* oder *in allgemeiner Lage*, wenn dim $(P_0 \vee ... \vee P_k) = k$ ist. Sind die Punkte nicht affin unabhängig, so heißen sie *affin abhängig*.

Eine Teilmenge $C \subset \mathbb{A}$ heißt *affin unabhängig*, wenn für jedes $k \in \mathbb{N}$ alle paarweise verschiedenen Punkte $P_0,...,P_k$ aus C affin unabhängig sind. Ist C nicht affin unabhängig, so heißt C *affin abhängig*.

Beispiele. Jeder Punkt ist affin unabhängig. Zwei Punkte sind genau dann affin unabhängig, wenn sie verschieden sind. Drei Punkte sind genau dann affin unabhängig, wenn sie nicht auf einer gemeinsamen Geraden liegen, d.h. wenn sie nicht kollinear sind, u.s.w.

Satz 4. *Es seien $\mathbb{A}$ ein affiner Raum und $P_0,...,P_k \in \mathbb{A}$. Dann sind folgende Aussagen äquivalent:*

(a) *Die Punkte $P_0,...,P_k$ sind affin unabhängig.*

(b) *Die Vektoren $\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}$ sind linear unabhängig.*

Beweis. Für den Richtungsraum U von $P_0 \vee ... \vee P_k$ gilt $U = [\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}]$. Sind die Punkte $P_0,...,P_k$ affin unabhängig, so gilt dim $U = k$ und die Vektoren $\overrightarrow{P_0P_1},...,$ $\overrightarrow{P_0P_k}$ sind linear unabhängig. Sind umgekehrt die Vektoren $\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}$ linear unabhängig, so ist $U = [\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}]$ ein k–dimensionaler Untervektorraum und

es gilt somit dim $(P_0 \vee \ldots \vee P_k) = k$. ∎

Bemerkungen. **(a)** Satz 4 gilt ganz entsprechend, wenn der Bezugspunkt P_0 durch irgend einen anderen der Punkte P_i ersetzt wird, wenn man also die Vektoren $\overrightarrow{P_iP_0},\ldots,\overrightarrow{P_iP_{i-1}}, \overrightarrow{P_iP_{i+1}},\ldots,\overrightarrow{P_iP_k}$ betrachtet.

(b) Ist $\mathbb{A} = V$, so gilt außerdem: Die Punkte $x_0,\ldots,x_k$ sind genau dann affin unabhängig, wenn aus $a_0\, x_0 + \cdots + a_k\, x_k = o$ und $a_0 + \cdots + a_k = 0$ stets $a_0 = \ldots = a_k = 0$ folgt.

Beweis. Seien $x_0,\ldots,x_k$ affin unabhängig, also $x_1 - x_0,\ldots,x_k - x_0$ linear unabhängig. Dann folgt aus $a_0\, x_0 + \cdots + a_k\, x_k = o$ und $a_0 + \cdots + a_k = 0$ unmittelbar $a_1 (x_1 - x_0) + \cdots + a_k (x_k - x_0) = o$ und daraus $a_1 = \ldots = a_k = 0$. Dann ist auch $a_0 = 0$.

Umgekehrt folgt aus $a_1 (x_1 - x_0) + \cdots + a_k (x_k - x_0) = o$ zunächst $(-a_1 - \cdots - a_k)\, x_0 + a_1\, x_1 + \cdots + a_k\, x_k = o$ und $a_0 + \cdots + a_k = 0$ mit $a_0 := -a_1 - \cdots - a_k$. Also ist $a_1 = \ldots = a_k = 0$, und die Vektoren $x_1 - x_0,\ldots,x_k - x_0$ somit linear unabhängig, woraus die Behauptung folgt. ∎

(c) Sind die Punkte $P_0,\ldots,P_k$ affin unabhängig und ist $\{i_1,\ldots,i_m\} \subset \{0,\ldots,k\}$, so sind auch die Punkte $P_{i_1},\ldots,P_{i_m}$ affin unabhängig.

(d) Sind die Punkte $P_0,\ldots,P_k$ gegeben und sind für $\{i_1,\ldots,i_m\} \subset \{0,..,k\}$ die Punkte $P_{i_1},\ldots,P_{i_m}$ affin abhängig, so sind auch die Punkte $P_0,\ldots,P_k$ affin abhängig.

(e) Sind die Punkte $P_0,\ldots,P_k$ affin abhängig, so gilt dim $(P_0 \vee \cdots \vee P_k) < k$.

(f) Jeder k–dimensionale affine Unterraum L ist affine Hülle von $k + 1$ affin unabhängigen Punkten.

Beweis. Wir wählen in L einen beliebigen Punkt P_0 und im Richtungsraum U_L eine beliebige Basis $\{x_1,\ldots,x_k\}$. Dann existieren Punkte $P_1,\ldots,P_k$ aus L mit $\overrightarrow{P_0P_i} = x_i$, $i = 1,\ldots,k$. Nach Satz 4 sind die Punkte $P_0,\ldots P_k$ affin unabhängig, und ihre affine Hülle ist L. ∎

Mit Hilfe der affinen Unabhängigkeit lassen sich nun auch affine Koordi-

natensysteme definieren.

Definition. Es seien L ein k–dimensionaler affiner Unterraum von $\mathbb{A}$ und $P_0,...,P_k$ $k+1$ affin unabhängige Punkte in L.

Dann heißt $(P_0; \overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k})$ ein *affines Koordinatensystem* von L mit dem *Koordinatenursprung* P_0, den *Koordinatenachsen* P_0P_i und den *Einheitspunkten* P_i auf diesen Achsen, $i = 1,...,k$.

Für jeden Punkt $P \in L$ heißen die Koordinaten $a_1,...,a_k$ des Vektors $\overrightarrow{P_0P}$ bezüglich der Basis $(\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k})$ von U_L die *affinen Koordinaten* des Punktes P bezüglich des affinen Koordinatensystems $(P_0; \overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k})$. Wir schreiben dann auch $P(a_1,...,a_k)$.

In einem n–dimensionalen affinen Raum $\mathbb{A}$ erhalten wir also für jede Wahl eines Ursprungs $O \in \mathbb{A}$ und jede Wahl einer Basis $(v_1,...,v_n)$ von V ein affines Koordinatensystem $(O; v_1,...,v_n)$ von $\mathbb{A}$.

Ist $\mathbb{A}$ ein euklidischer Raum und ist $(v_1,...,v_n)$ eine ONB von V, so sprechen wir von einem *kartesischen Koordinatensystem* $(O; v_1,...,v_n)$ von $\mathbb{A}$, und die affinen Koordinaten eines Punktes X heißen dann entsprechend *kartesische Koordinaten.*

Es seien nun wieder L ein k–dimensionaler affiner Unterraum von $\mathbb{A}$ und $(P_0; \overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k})$ ein affines Koordinatensystem in L. Dann gilt für die Ortsvektoren der Punkte $P \in L$ bezüglich eines beliebigen Ursprungs $O \in \mathbb{A}$

$$(*) \qquad \overrightarrow{OP} = \overrightarrow{OP_0} + a_1 \overrightarrow{P_0P_1} + ... + a_k \overrightarrow{P_0P_k},$$

wobei $a_1,...,a_k$ die affinen Koordinaten von P bezüglich des ausgezeichneten Koordinatensystems sind. Wir nennen (*) eine *Parameterdarstellung 1.Art* von L mit den *Richtungsvektoren* $\overrightarrow{P_0P_1},...,\overrightarrow{P_0P_k}$ und den *Parametern* $a_1,...,a_k$ (vgl. S.127).

In dieser Parameterdarstellung von L ist der Punkt P_0 vor den anderen Punkten $P_1,...,P_k$ ausgezeichnet. Eine symmetrische Darstellung können wir erreichen, indem wir (*) etwas umformen. Wir setzen $a_0 := 1 - a_1 -...- a_k$ und erhalten

$$\overrightarrow{OP} = (a_0 +...+ a_k)\overrightarrow{OP_0} + a_1 \overrightarrow{P_0P_1} + ... + a_k \overrightarrow{P_0P_k}.$$

Also gilt

$$(**) \qquad \overrightarrow{OP} = a_0 \overrightarrow{OP_0} + a_1 \overrightarrow{OP_1} + \ldots + a_k \overrightarrow{OP_k} \ , \ a_i \in \mathbb{K}, \ a_0 + \ldots + a_k = 1 \ .$$

Wir nennen (**) eine *Parameterdarstellung 2.Art* von L. Die eindeutig bestimmten Skalare $a_0,\ldots,a_k$, die wiederum unabhängig von der Wahl von O sind, heißen *Schwerpunktskoordinaten* oder *baryzentrische Koordinaten* des Punktes P bezüglich $P_0,\ldots,P_k$. Ist char $\mathbb{K}$ kein Teiler von $k+1$, so heißt der Punkt S mit den baryzentrischen Koordinaten $a_0 = \ldots = a_k = \frac{1}{k+1}$ *Schwerpunkt* (*Baryzentrum*) von $P_0,\ldots,P_k$.

Die Darstellung (**) des Ortsvektors $\overrightarrow{OP}$ bezeichnet man auch als *Affinkombination* der Vektoren $\overrightarrow{OP_0},\ldots,\overrightarrow{OP_k}$. Entsprechend heißt $P \in \mathbb{A}$ *Affinkombination* der Punkte $P_0,\ldots,P_k \in \mathbb{A}$, wenn es einen Punkt $O \in \mathbb{A}$ gibt und Zahlen $a_0,\ldots,a_k \in \mathbb{K}$ mit $a_0 + \ldots + a_k = 1$, für die (**) erfüllt ist. In diesem Fall gilt (**) dann für alle $O \in \mathbb{A}$. Die Punkte $P_0,\ldots,P_k$ müssen dabei nicht affin unabhängig sein. Sind sie es aber, so sind die Zahlen $a_0,\ldots,a_k$ die baryzentrischen Koordinaten von P bezüglich $P_0,\ldots,P_k$, also eindeutig bestimmt.

Bemerkungen. **(a)** $P \in \mathbb{A}$ ist genau dann Affinkombination der Punkte $P_0,\ldots,P_k \in \mathbb{A}$, wenn es Zahlen $a_0,\ldots,a_k \in \mathbb{K}$ gibt mit $a_0 + \ldots + a_k = 1$ und

$$a_0 \overrightarrow{PP_0} + \ldots + a_k \overrightarrow{PP_k} = o \ .$$

(b) Sei $P \in \mathbb{A}$ und seien $P_0,\ldots,P_k \in \mathbb{A}$ affin unabhängig. Dann ist P genau dann Affinkombination der Punkte $P_0,\ldots,P_k$, wenn $P_0,\ldots,P_k,P$ affin abhängig sind.

(c) Sei $C \subset \mathbb{A}$ nichtleer. Dann ist die affine Hülle von C die Menge aller Affinkombinationen von Punkten aus C.

(Beweise als Übung)

Als nächstes wollen wir uns mit metrischen Problemen wie Orthogonalität und Abstandsbestimmungen beschäftigen. Hierzu setzen wir voraus, daß $\mathbb{A}$ ein endlich dimensionaler euklidischer Raum ist.

Orthogonale Unterräume

Bei der Definition orthogonaler Unterräume orientieren wir uns an den Verhältnissen im Anschauungsraum. Um Randfälle zu vermeiden, die unserer Anschauung widersprechen, setzen wir voraus, daß die betrachteten Unterräume L_1 und L_2 nicht parallel sind.

Es wäre naheliegend, L_1 und L_2 orthogonal zu nennen, wenn ihre zugehörigen Richtungsräume U_1 und U_2 orthogonal sind. Hierzu müßte $U_1 \cap U_2 = \{o\}$ gelten. Aus geometrischen Gründen wollen wir aber in gewissen Fällen, in denen $U_1 \cap U_2 \neq \{o\}$ gilt, ebenfalls von orthogonalen Unterräumen sprechen. Hierzu betrachten wir die orthogonalen Komplemente $W_1 = (U_1 \cap U_2)^\perp \cap U_1$ bzw. $W_2 = (U_1 \cap U_2)^\perp \cap U_2$ des Schnittes $U_1 \cap U_2$ in U_1 bzw. in U_2. Weil nach Voraussetzung weder U_1 in U_2 noch U_2 in U_1 gelegen ist, gilt $W_i \neq \{o\}$, $i = 1,2$.

Wir nennen nun die Unterräume L_1 und L_2 von $\mathbb{A}$ *orthogonal*, in Zeichen $L_1 \perp L_2$, falls $W_1 \perp W_2$ gilt.

Ist L_1 speziell eine Gerade und $L_1 \cap L_2 \neq \emptyset$, so heißt L_1 auch ein *Lot* auf L_2 .

Beispiel. Im euklidischen Raum $\mathbb{R}^3$ sind bei einem kartesischen Koordinatensystem sowohl die Koordinatenachsen als auch die Koordinatenebenen paarweise orthogonal. Jede Koordinatenachse ist orthogonal zu jeder Koordinatenebene, die durch die beiden anderen Koordinatenachsen aufgespannt wird, und sie ist auch orthogonal zu jeder Geraden in dieser Ebene.

Bemerkungen. **(a)** Für $U_1 \cap U_2 = \{o\}$ ist $W_1 = U_1$ und $W_2 = U_2$, also sind L_1 und L_2 in diesem Fall genau dann orthogonal, wenn es ihre Richtungsräume sind.

(b) Orthogonale Unterräume sind nach Definition nie parallel, müssen sich allerdings auch nicht schneiden. So gibt es, wie wir gesehen haben, im euklidischen $\mathbb{R}^3$ orthogonale windschiefe Geraden.

(c) Jedes Lot g auf einen Unterraum L hat mit diesem genau einen Punkt gemeinsam; dieser Punkt heißt dann der *Lotfußpunkt* von g auf L.

Abstand euklidischer Unterräume

Im euklidischen Fall übertragen sich die metrischen Eigenschaften des Vektorraumes V direkt auf die Punkte von $\mathbb{A}$.

Definition. Unter dem *Abstand* der Punkte $X, Y \in \mathbb{A}$ verstehen wir die Länge ihres Verbindungsvektors,

$$d(X,Y) := \| \overrightarrow{XY} \| .$$

Der euklidische Raum $\mathbb{A}$ wird damit ein metrischer Raum. Ist $\mathbb{A}$ n–dimensional und sind $(x_1,...,x_n)$ und $(y_1,...,y_n)$ die kartesischen Koordinaten der Punkte X und Y bezüglich eines kartesischen Koordinatensystems $(O; v_1,...,v_n)$, so gilt

$$d(X,Y) = \Big(\sum_{i=1}^{n} (y_i - x_i)^2 \Big)^{1/2} .$$

Wir wollen die Definition des Abstandes auf beliebige Teilmengen M_1 , M_2 von $\mathbb{A}$ übertragen.

Definition. Unter dem *Abstand* der Teilmengen M_1, M_2 des euklidischen Raumes $\mathbb{A}$ verstehen wir die Zahl

$$d(M_1,M_2) := \inf \{ d(X_1,X_2) \mid X_1 \in M_1 , X_2 \in M_2 \}.$$

Speziell für Unterräume eines endlich dimensionalen euklidischen Raumes läßt sich der Abstand mit Hilfe von Orthogonalprojektionen konkret angeben.

Satz 5. *Es seien L_1 , L_2 euklidische Unterräume mit zugehörigen Richtungsräumen U_1 , U_2 , und $P_1 \in L_1$ sowie $P_2 \in L_2$ seien beliebig gewählt. Dann gilt*

$$d(L_1,L_2) = d(\overrightarrow{P_1P_2} , U_1 + U_2) = \| \overrightarrow{P_1P_2} - \pi_{U_1+U_2} (\overrightarrow{P_1P_2}) \| .$$

Beweis. Für alle Punkte $X_1 \in L_1$ und $X_2 \in L_2$ gilt

$$d(X_1,X_2) = \| \overrightarrow{X_1X_2} \| = \| \overrightarrow{X_1P_1} + \overrightarrow{P_1P_2} + \overrightarrow{P_2X_2} \| = \| \overrightarrow{P_1P_2} - \overrightarrow{P_1X_1} + \overrightarrow{P_2X_2} \| ,$$

also

$$d(L_1,L_2) = \inf_{\substack{x_1 \in U_1 \\ x_2 \in U_2}} \| \overrightarrow{P_1P_2} - x_1 + x_2 \| = \| \overrightarrow{P_1P_2} - \pi_{U_1+U_2}(\overrightarrow{P_1P_2}) \| . \qquad \blacksquare$$

Bemerkungen. **(a)** Die Formel von Satz 5 gibt nur den Abstand zwischen L_1 und L_2 an, aber keine Punkte $Q_1 \in L_1$ und $Q_2 \in L_2$, für die der Abstand angenommen wird, für die also $d(Q_1,Q_2) = d(L_1,L_2)$ gilt. Diese erhalten wir folgendermaßen:

Ist $\pi_{U_1+U_2}(\overrightarrow{P_1P_2}) = x_1 + x_2$, $x_1 \in U_1$, $x_2 \in U_2$, so seien Q_1 bzw. Q_2 die eindeutig bestimmten Punkte mit $x_1 = \overrightarrow{P_1Q_1}$ bzw. $-x_2 = \overrightarrow{P_2Q_2}$. Es gilt dann $d(L_1,L_2) = \| \overrightarrow{P_1P_2} - \overrightarrow{P_1Q_1} + \overrightarrow{P_2Q_2} \| = \| \overrightarrow{Q_1Q_2} \|$.

(b) Sind die Unterräume L_1 und L_2 disjunkt, so ist $Q_1 \neq Q_2$, und die Gerade Q_1Q_2 ist ein gemeinsames Lot von L_1 und L_2:

Aus $d(Q_1,Q_2) = d(L_1,L_2)$ folgt nämlich mit Satz 5 und dem Satz von Pythagoras

$$\| \overrightarrow{Q_1Q_2} \|^2 = \| \overrightarrow{Q_1Q_2} - \pi_{U_1+U_2}(\overrightarrow{Q_1Q_2}) \|^2 = \| \overrightarrow{Q_1Q_2} \|^2 - \| \pi_{U_1+U_2}(\overrightarrow{Q_1Q_2}) \|^2 .$$

Also ist $\pi_{U_1+U_2}(\overrightarrow{Q_1Q_2}) = o$ und somit $\overrightarrow{Q_1Q_2} \perp U_1 + U_2$. Daraus folgt $Q_1Q_2 \perp L_1$ und $Q_1Q_2 \perp L_2$.

Umgekehrt gilt für die Lotfußpunkte Q_1 und Q_2 eines beliebigen gemeinsamen Lotes von L_1 und L_2 offensichtlich $d(Q_1,Q_2) = d(L_1,L_2)$.

(c) Das Punktepaar $(Q_1,Q_2) \in L_1 \times L_2$ der Lotfußpunkte ist genau dann eindeutig bestimmt, wenn die Summe $U_1 + U_2$ direkt ist. Dies wiederum ist genau dann der Fall, wenn L_1 und L_2 keine parallelen Geraden enthalten (Übungsaufgabe).

(d) Speziell für den Abstand eines Punktes X von einem Unterraum L erhalten wir nach Wahl eines beliebigen Punktes $P \in L$

$$d(X,L) = \| \overrightarrow{PX} - \pi_{U_L}(\overrightarrow{PX}) \| .$$

Beispiel. Im euklidischen $\mathbb{R}^5$ seien die Gerade $g = x_0 + [x_1]$ und die Ebene

$L = y_0 + [x_2,x_3]$ gegeben mit

$$x_0 = \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \; y_0 = \begin{bmatrix} 3 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \; x_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \; x_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 2 \end{bmatrix}, \; x_3 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Wir wollen den Abstand $d(g,L)$ berechnen sowie Punkte $x \in g$ und $y \in L$ bestimmen mit $d(g,L) = d(x,y) = \| x - y \|$.

Die Verbindungsgerade von x und y muß notwendigerweise orthogonal zu g und zu L sein, also muß $x - y \in [x_1,x_2,x_3]^\perp$ sein. Für die gesuchten Punkte x und y gilt $x = x_0 + a_1x_1$ und $y = y_0 + a_2x_2 + a_3x_3$ mit $a_1,a_2,a_3 \in \mathbb{R}$. Die Bedingung $x - y \in [x_1,x_2,x_3]^\perp$ ist dann äquivalent zu

$$<x_0 - y_0 + a_1x_1 - a_2x_2 - a_3x_3 \, , \, x_i> \; = \; 0. \qquad (i = 1,2,3).$$

Dies ist ein inhomogenes LGS mit den Unbekannten a_1,a_2,a_3 und der erweiterten Matrix

$$\left[\begin{array}{ccc|c} 6 & -1 & 4 & -2 \\ 1 & -6 & -3 & 1 \\ -4 & -3 & -6 & 2 \end{array}\right]$$

Als Lösung erhalten wir $a_1 = -\frac{1}{2}$, $a_2 = -\frac{1}{3}$, $a_3 = \frac{1}{6}$. Damit ergibt sich

$$x = \frac{1}{2}\begin{bmatrix} 3 \\ 0 \\ 1 \\ -2 \\ 0 \end{bmatrix}, \; y = \frac{1}{6}\begin{bmatrix} 16 \\ 2 \\ 6 \\ -5 \\ -3 \end{bmatrix},$$

also $d(g,L) = \sqrt{2}$.

Besonders einfach läßt sich der Abstand eines Punktes X von einer Hyperebene L berechnen, wenn diese in Hessescher Normalform gegeben ist. Darunter verstehen wir folgendes:

Wir wählen einen beliebigen Ursprung O in $\mathbb{A}$ und in L einen beliebigen Punkt P. Der Orthogonalraum zu dem Richtungsraum U von L ist eindimensional, also gibt es in $U^{\perp}$ genau zwei Vektoren der Länge 1. Wir wählen einen davon aus, $y \in U^{\perp}$, $\| y \| = 1$, und erhalten

$$L = \{X \in \mathbb{A} \mid \overrightarrow{OX} = \overrightarrow{OP} + u\,,\ u \in U\} = \{X \in \mathbb{A} \mid <\overrightarrow{OX},y> = <\overrightarrow{OP},y> =: a\}\,.$$

Die letztere Darstellung von L heißt *Hessesche Normalform* der Hyperebene L, und y heißt *Normalenvektor* von L.

Ist nun $X \in \mathbb{A}$ ein beliebiger Punkt, so ist der Vektor $\overrightarrow{PX} - \pi_U(\overrightarrow{PX})$ ein Vielfaches des Normalenvektors y von L,

$$\overrightarrow{PX} - \pi_U(\overrightarrow{PX}) = b\,y\,,\ b \in \mathbb{R}\,.$$

Daraus folgt $<\overrightarrow{PX},y> = b$ und wegen $<\overrightarrow{OX},y> = <\overrightarrow{OP},y> + <\overrightarrow{PX},y> = a + b$ erhalten wir schließlich für den Abstand des Punktes X zur Hyperebene L:

$$d(X,L) = |b| = |<\overrightarrow{OX},y> - a|\,.$$

Beispiel. Im euklidischen Raum $\mathbb{R}^4$ sei die Hyperebene L durch die Gleichung

$$2\,x_1 + x_3 + 2\,x_4 = 6$$

mit den Variablen $x_1,\ldots,x_4$ gegeben. Wir wollen den Abstand des Punktes

$$x = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \end{bmatrix}$$

von L bestimmen und stellen dazu L in der Hesseschen Normalform dar:

$$2\,x_1 + x_3 + 2\,x_4 = 6 \iff <\begin{bmatrix} 2 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}> = 6\,.$$

Somit besitzt L die Hessesche Normalform $L = \{x' \mid <x',y> = a\}$ mit

$$y = \frac{1}{3}\begin{bmatrix} 2 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \quad a = 2 \ .$$

Daraus folgt für den gesuchen Abstand: $d\,(x,L) = |\ <x,y> - a\ | = \frac{5}{3}$.

Speziell in einem dreidimensionalen euklidischen Raum gibt es noch eine einfache Methode um den Abstand zweier windschiefer Geraden zu bestimmen, bei der das sogenannte *vektorielle Produkt* eine Rolle spielt. Dieses spezielle Produkt ist auch für andere Anwendungen, etwa in der Physik, nützlich. Wir geben deshalb zum Abschluß dieses Paragraphen eine knappe Einführung in das Vektorprodukt.

Das Vektorprodukt

In einem dreidimensionalen euklidischen Raum $\mathbb{A}$ mit zugehörigem Vektorraum V besitzt ein Parallelogramm $OPQR$ (mit $\overrightarrow{OP} = \overrightarrow{RQ}$ und $\overrightarrow{OR} = \overrightarrow{PQ}$) den Flächeninhalt $\| \overrightarrow{OP} \| \cdot \| \overrightarrow{OR} \| \cdot |\sin \omega|$, wobei ω der Winkel zwischen $\overrightarrow{OP}$ und $\overrightarrow{OR}$ ist. Wir wollen die Parallelogrammfläche durch einen Vektor beschreiben, der zu $\overrightarrow{OP}$ und $\overrightarrow{OR}$ orthogonal ist und dessen Länge mit dem Flächeninhalt übereinstimmt.

Seien nun x und y beliebige Vektoren aus V und (e_1,e_2,e_3) eine feste ONB von V. Dann wird durch

$$z \longmapsto \Delta\,(x,y,z) := \det(\hat{x} \mid \hat{y} \mid \hat{z})$$

auf V eine Linearform definiert. Nach Satz 5.11 gibt es einen eindeutig bestimmten Vektor $z_0 \in V$ mit

$$(*) \qquad \Delta\,(x,y,z) = <z_0,z>$$

für alle $z \in V$. Dieser Vektor z_0 heißt *Vektorprodukt* (*Kreuzprodukt, äußeres Produkt*) der Vektoren x und y; wir schreiben dafür $z_0 = x \times y$.

Bemerkung. Das Vektorprodukt ist eine Abbildung $\times : V \times V \longrightarrow V$ mit folgenden Eigenschaften:

(a) $x \times y = -y \times x$,

(b) $(x + x') \times y = x \times y + x' \times y$,

(c) $(a\,x) \times y = a\,(x \times y)$,

(d) $<x \times y, x> = <x \times y, y> = 0$,

(e) x, y linear abhängig $\iff x \times y = o$,

(f) $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$.

(g) Seien $x = x_1\,e_1 + x_2\,e_2 + x_3\,e_3$ und $y = y_1\,e_1 + y_2\,e_2 + y_3\,e_3$. Dann gilt:

$$x \times y = (x_2\,y_3 - x_3\,y_2)\,e_1 + (x_3\,y_1 - x_1\,y_3)\,e_2 + (x_1\,y_2 - x_2\,y_1)\,e_3 .$$

Merkschema: Man "entwickle" folgenden Ausdruck formal wie eine Determinante nach der ersten Zeile:

$$\begin{vmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

(h) $\| x \times y \| = \sqrt{(\| x \|^2 \| y \|^2 - <x,y>^2)} = \| x \| \cdot \| y \| \cdot | \sin \omega |$,

wo ω der Winkel zwischen x und y ist $(x \neq o,\ y \neq o)$.

Die Beweise überlassen wir als Übungsaufgabe.

Der Flächeninhalt des Parallelogramms $OPQR$ ist nach (h) also $\| \overrightarrow{OP} \times \overrightarrow{OR} \|$.

Nun wollen wir noch, wie angekündigt, mit Hilfe des Vektorproduktes den Abstand zweier windschiefer Geraden berechnen:

Es seien g und h windschiefe Geraden. Ihre Richtungen seien $U_g = [x]$ und $U_h = [y]$. Dann ist $x \times y$ der Richtungsvektor des gemeinsamen Lotes, und für die Lotfußpunkte $P_1 \in g$ und $Q_1 \in h$ gilt $\overrightarrow{P_1Q_1} = a\,(x \times y)$, $a \in \mathbb{R}$. Wir wollen a bestimmen. Hierzu wählen wir beliebige Punkte $P \in g$ und $Q \in h$ und erhalten

$$\overrightarrow{P_1Q_1} = \overrightarrow{P_1P} + \overrightarrow{PQ} + \overrightarrow{QQ_1} .$$

Daraus folgt

$$a \,\|\, x \times y \,\|^2 \;=\; \langle \overrightarrow{P_1 Q_1}\,,\, x \times y \rangle \;=\; \langle \overrightarrow{PQ}\,,\, x \times y \rangle\,.$$

Also ist

$$d\,(g,h) \;=\; \|\, \overrightarrow{P_1 Q_1}\,\| \;=\; |\; a\;| \cdot \|\, x \times y \,\| \;=\; \frac{|\langle \overrightarrow{PQ}\;,\; x \times y \rangle|}{\|\; x \times y \;\|}\,.$$

Beispiel. Im $\mathbb{R}^3$ seien die windschiefen Geraden

$$g = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + [\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}] \;\;,\;\; h = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} + [\begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}]$$

gegeben, ihr Abstand soll berechnet werden. In diesem Fall ist

$$\overrightarrow{PQ} \;=\; \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$$

und

$$x \times y \;=\; (2 \cdot 3 - 1 \cdot 0)\; e_1 + (1 \cdot 2 - (-1) \cdot 3)\; e_2 + ((-1) \cdot 0 - 2 \cdot 2)\; e_3 \;=\; \begin{bmatrix} 6 \\ 5 \\ -4 \end{bmatrix}.$$

Also folgt

$$\|\, x \times y \,\| \;=\; \sqrt{77}\;\;,\;\; \langle \overrightarrow{PQ}, x \times y \rangle \;=\; 16\,,$$

und wir erhalten $d\,(g,h) \;=\; \dfrac{16}{\sqrt{77}}$.

§ 2 Affine Abbildungen und Bewegungen

Nun wollen wir auch affine Abbildungen in allgemeinen affinen Räumen betrachten.

Definition. Es seien $\mathbb{A}$ und $\mathbb{B}$ affine Räume mit zugehörigen $\mathbb{K}$–Vektorräumen V und W. Dann heißt $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ eine *affine Abbildung*, falls eine lineare Abbildung Φ aus Hom (V, W) existiert mit

$$\Phi(\overrightarrow{PQ}) = \overrightarrow{\varphi(P)\varphi(Q)}$$

für alle Punkte $P, Q \in \mathbb{A}$. Jede bijektive affine Selbstabbildung $\varphi : \mathbb{A} \longrightarrow \mathbb{A}$ heißt *Affinität*.

Bemerkung. Nach Auszeichnung eines Ursprungs $O \in \mathbb{A}$ und eines Ursprungs $O' \in \mathbb{B}$ gilt für alle Punkte $X \in \mathbb{A}$:

$$\overrightarrow{O'\varphi(X)} = \overrightarrow{\varphi(O)\varphi(X)} + \overrightarrow{O'\varphi(O)} = \Phi(\overrightarrow{OX}) + w$$

mit $w = \overrightarrow{O'\varphi(O)} \in W$. Damit genügt es meist, den Fall $\mathbb{A} = V$, $\mathbb{B} = W$ und

$$\begin{aligned} \varphi : V &\longrightarrow W \\ x &\longmapsto \Phi(x) + w \end{aligned}$$

$\Phi \in \mathrm{Hom}(V, W)$, $w \in W$, zu betrachten (vgl. § 3.4).

w heißt *Translationsvektor* zur affinen Abbildung φ. Er ist ebenso wie die lineare Abbildung Φ durch φ eindeutig bestimmt.

Beispiele affiner Abbildungen haben wir schon in § 3.4 kennengelernt:

Die Translationen sind diejenigen Abbildungen $\tau : \mathbb{A} \longrightarrow \mathbb{A}$, für die $\overrightarrow{\tau(P)\tau(Q)} = \overrightarrow{PQ}$ für alle $P, Q \in \mathbb{A}$ gilt. Die zugehörige lineare Abbildung ist die Identität.

Die Streckungen mit Zentrum P sind diejenigen Abbildungen $\delta : \mathbb{A} \longrightarrow \mathbb{A}$, für die es ein $c \in \mathbb{K}$, $c \neq 0$, gibt mit $\overrightarrow{P\delta(Q)} = c\,\overrightarrow{PQ}$ für alle $Q \in \mathbb{A}$. Die zugehörigen linearen Abbildungen sind also von der Form $\Phi = c\,\mathrm{id}_V$, $c \neq 0$. Für char $\mathbb{K} \neq 2$ und

$c = -1$ ist δ die Punktspiegelung an P.

Translationen und Streckungen haben die Eigenschaft, daß sie jede Gerade auf eine dazu parallele Gerade abbilden. Solche Affinitäten nennt man auch *Homothetien.* Umgekehrt läßt sich zeigen, daß jede Homothetie, die von der Identität verschieden ist, entweder eine Translation oder eine Streckung ist (Beweis als Übungsaufgabe).

Affine Abbildungen sind durch ein affines Koordinatensystem eindeutig festgelegt.

Satz 6. *Es seien* $\mathbb{A},\mathbb{B}$ *affine Räume über demselben Körper und* dim $\mathbb{A} = n$. *Weiter seien* $P_0,...,P_n$ *affin unabhängige Punkte in* $\mathbb{A}$ *sowie* $Q_0,...,Q_n$ *beliebige Punkte in* $\mathbb{B}$. *Dann gibt es genau eine affine Abbildung* $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ *mit* $\varphi(P_i) = Q_i$ *für* $i = 0,...,n$.

Beweis. Seien V bzw. W die zu $\mathbb{A}$ bzw. $\mathbb{B}$ gehörenden Vektorräume. Wegen der affinen Unabhängigkeit der Punkte $P_0,...,P_n$ sind die Vektoren $\overrightarrow{P_0P_i} \in V$, $i = 1,...,n$ linear unabhängig, bilden also eine Basis von V. Nach Satz 3.4 gibt es genau eine lineare Abbildung $\Phi : V \longrightarrow W$ mit $\Phi(\overrightarrow{P_0P_i}) = \overrightarrow{Q_0Q_i}$, $i = 1,...,n$. Die durch $X \longmapsto \varphi(X)$ mit $\overrightarrow{Q_0\varphi(X)} = \Phi(\overrightarrow{P_0X})$, $X \in \mathbb{A}$, erklärte Abbildung $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ ist affin und bildet P_i auf Q_i ab für $i = 0,...,n$. Die Abbildung φ ist eindeutig bestimmt. Aus $\varphi(P_i) = \varphi'(P_i)$ folgt nämlich $\Phi(\overrightarrow{P_0P_i}) = \Phi'(\overrightarrow{P_0P_i})$ für $i = 1,...,n$. Daraus erhalten wir $\Phi = \Phi'$ und somit $\overrightarrow{Q_0\varphi(X)} = \Phi(\overrightarrow{P_0X}) = \Phi'(\overrightarrow{P_0X}) = \overrightarrow{Q_0\varphi'(X)}$ für alle $X \in \mathbb{A}$, also ergibt sich $\varphi = \varphi'$. ∎

Beispiel. Im $\mathbb{R}^3$ seien die Punkte $x_0,...,x_3$ gegeben und im $\mathbb{R}^4$ die Punkte $y_0,...,y_3$. Dabei seien

$$x_0 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, x_1 = \begin{bmatrix} 4 \\ 3 \\ 3 \end{bmatrix}, x_2 = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix},$$

$$y_0 = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix}, y_1 = \begin{bmatrix} 5 \\ 4 \\ 4 \\ 5 \end{bmatrix}, y_2 = \begin{bmatrix} 4 \\ 3 \\ 4 \\ 4 \end{bmatrix}, y_3 = \begin{bmatrix} 1 \\ 4 \\ 0 \\ -3 \end{bmatrix}.$$

Wir suchen eine affine Abbildung $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ mit $\varphi(x_i) = y_i$, $i = 0,...,3$. Zunächst stellen wir fest, daß die Punkte $x_0,...,x_3$ affin unabhängig sind, da die Matrix

$$S = (x_1 - x_0 \mid x_2 - x_0 \mid x_3 - x_0) = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

den Rang 3 hat.

Die gesuchte affine Abbildung hat die Form $x \longmapsto A\,x + w$ mit $A \in \mathbb{R}^{4\times 3}$ und $w \in \mathbb{R}^4$. Es gilt

$$\varphi(x_i) - \varphi(x_0) = A\,(x_i - x_0)\ ,\ i = 1,2,3 \iff y_i - y_0 = A\,(x_i - x_0)\ ,\ i = 1,2,3$$

$$\iff (y_1 - y_0 \mid y_2 - y_0 \mid y_3 - y_0) = A\,S \iff A = (y_1 - y_0 \mid y_2 - y_0 \mid y_3 - y_0) \cdot S^{-1}$$

$$= \begin{bmatrix} 3 & 2 & -1 \\ 2 & 1 & 2 \\ 1 & 1 & -3 \\ 4 & 3 & -4 \end{bmatrix} \begin{bmatrix} -1 & 2 & -5 \\ 2 & -3 & 8 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ 2 & -1 & 0 \end{bmatrix}$$

Für den Translationsvektor w gilt $w = y_0 - A\,x_0$, also

$$w = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 0 \end{bmatrix}.$$

Aus Satz 6 folgt speziell, daß in $\mathbb{A}$ jeder Wechsel eines affinen Koordinatensystems durch eine Affinität beschrieben wird.

Umgekehrt bildet jede Affinität $\varphi : \mathbb{A} \longrightarrow \mathbb{A}$ mit zugehörigem Vektorraumisomorphismus Φ ein Koordinatensystem $(O;\, v_1,...,v_n)$ auf ein neues Koordinatensystem $(\varphi(O);\, \Phi(v_1),...,\Phi(v_n))$ ab. Für jeden Punkt $X \in \mathbb{A}$ stimmen dann die affinen Koordinaten von X im alten und die von $\varphi(X)$ im neuen Koordinatensystem überein. Aus

$$\overrightarrow{OX} = x_1\,v_1 + ... + x_n\,v_n$$

folgt nämlich

$$\overrightarrow{\varphi(O)\varphi(X)} = \Phi(\overrightarrow{OX}) = x_1 \Phi(v_1) + \ldots + x_n \Phi(v_n).$$

Insbesondere stimmen also die affinen Koordinaten von X im neuen System mit denen von $\varphi^{-1}(X)$ im alten System überein.

Affine Abbildungen lassen die geometrische Struktur des affinen Raumes invariant, denn sie bilden affine Unterräume auf affine Unterräume ab und erhalten die Parallelität (vgl. § 3.4). Bijektive affine Abbildungen erhalten darüber hinaus auch die Dimension. Sie bilden also Geraden auf Geraden ab, insbesonderere parallele Geraden auf ebensolche. Man spricht dann von der *Geradentreue* und der *Parallelentreue* dieser Abbildungen.

Umgekehrt reichen diese Eigenschaften aber nicht aus, um die bijektiven affinen Abbildungen zu charakterisieren.

Beispiel. In der affinen Ebene $\mathbb{C}^2$ ist die Abbildung

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \longmapsto \begin{bmatrix} \overline{z}_1 \\ \overline{z}_2 \end{bmatrix}$$

bijektiv, geradentreu und parallelentreu, aber nicht affin.

Zur geometrischen Beschreibung affiner Abbildungen benötigen wir noch einen weiteren wichtigen Begriff der affinen Geometrie, das Teilverhältnis.

Definition. Sind P,Q,R drei kollineare Punkte (d.h. Punkte, die auf einer Geraden liegen) und ist $P \neq Q$, so heißt die Zahl $c \in \mathbb{K}$ mit $\overrightarrow{PR} = c\,\overrightarrow{PQ}$ das *Teilverhältnis* von R bezüglich P und Q, Schreibweise: TV $(P,Q;R) = c$.

Bemerkungen. **(a)** Für $c = 0$ ist $R = P$, für $c = 1$ ist $R = Q$ und für char $\mathbb{K} \neq 2$ und $c = 1/2$ ist R der *Mittelpunkt* von P und Q.

(b) Affine Abbildungen $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ lassen das Teilverhältnis invariant, d.h. es gilt TV $(P,Q;R)$ = TV $(\varphi(P),\varphi(Q);\varphi(R))$, falls $\varphi(P) \neq \varphi(Q)$ ist.

Satz 7. *Jede bijektive Abbildung $\varphi : \mathbb{A} \longrightarrow \mathbb{A}$, die geradentreu, parallelentreu und teilverhältnistreu ist, ist eine Affinität.*

Beweis. Wir betrachten zunächst den Fall, daß dim $\mathbb{A} = 1$ gilt, $\mathbb{A}$ also eine Gerade ist. Wir wählen in $\mathbb{A}$ zwei verschiedene Punkte O und P. Dann sind auch $\varphi(O)$ und $\varphi(P)$ verschieden und es gibt genau eine Affinität $\alpha : \mathbb{A} \longrightarrow \mathbb{A}$ mit $\alpha(\varphi(O)) = O$ und $\alpha(\varphi(P)) = P$. Aus der Teilverhältnistreue von $\alpha \circ \varphi$ folgt nun sofort $\alpha \circ \varphi = \mathrm{id}_{\mathbb{A}}$, also ist $\varphi = \alpha^{-1}$ eine Affinität.

Sei nun dim $\mathbb{A} \geq 2$. Da φ bijektiv, geradentreu und parallelentreu ist, bildet φ nichtausgeartete Parallelogramme auf ebensolche ab. Für alle $P,Q,R,S \in \mathbb{A}$ folgt somit aus $\overrightarrow{PQ} = \overrightarrow{RS}$ auch $\overrightarrow{\varphi(P)\varphi(Q)} = \overrightarrow{\varphi(R)\varphi(S)}$. Setzen wir also für beliebiges $x = \overrightarrow{PQ} \in V$ als Bildvektor $\Phi(x) := \overrightarrow{\varphi(P)\varphi(Q)}$, so wird dadurch eine bijektive Abbildung $\Phi : V \longrightarrow V$ definiert. Φ ist linear:

Die Homogenität von Φ ergibt sich unmittelbar aus der Teilverhältnistreue von φ: Seien $x = \overrightarrow{PQ} \in V$ und $c \in \mathbb{K}$ beliebig. Dann gilt mit $\overrightarrow{PR} = c\,\overrightarrow{PQ}$

$$\Phi(c\,x) = \Phi(c\,\overrightarrow{PQ}) = \Phi(\overrightarrow{PR}) = \overrightarrow{\varphi(P)\varphi(R)} = c\,\overrightarrow{\varphi(P)\varphi(Q)} = c\,\Phi(\overrightarrow{PQ}) = c\,\Phi(x)\,.$$

Die Additivität von Φ folgt für linear abhängige $x,y \in V$ aus der Homogenität von Φ, für linear unabhängige $x,y \in V$ folgt sie aus der Eigenschaft von φ, echte Parallelogramme in ebensolche abzubilden. ∎

Die Voraussetzungen von Satz 7 lassen sich abschwächen. Wir verweisen hierzu auf die Literatur (etwa: H. Schaal: Lineare Algebra und Analytische Geometrie, Teil I).

Die Affinitäten von $\mathbb{A}$ bilden eine Gruppe. Mit ihrer Hilfe lassen sich die Teilmengen von $\mathbb{A}$ so in Klassen einteilen, daß sich Elemente aus derselben Klasse bezüglich ihrer affingeometrischen Eigenschaften nicht mehr unterscheiden lassen.

Wir nennen M_1, $M_2 \subset \mathbb{A}$ *affin–äquivalent*, wenn es eine Affinität φ von $\mathbb{A}$ gibt mit $\varphi(M_1) = M_2$.

Dieses Einteilungsprinzip wird bei der affinen Klassifikation der Quadriken eine wichtige Rolle spielen.

Zum Schluß betrachten wir nun auch noch affine Abbildungen in euklidischen

Räumen.

Definition. Eine affine Abbildung $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ zweier euklidischer Räume $\mathbb{A}$ und $\mathbb{B}$ heißt *Isometrie*, falls die zugehörige lineare Abbildung Φ eine Isometrie ist.

Für $\mathbb{A} = \mathbb{B}$ heißt eine Isometrie φ auch *Bewegung*, und zwar *eigentliche Bewegung*, wenn $\det \Phi = 1$ und *uneigentliche Bewegung*, wenn $\det \Phi = -1$ ist.

Isometrien lassen den Abstand unverändert, denn für alle $X, Y \in \mathbb{A}$ gilt

$$d(\varphi(X),\varphi(Y)) = \| \overrightarrow{\varphi(X)\varphi(Y)} \| = \| \Phi(\overrightarrow{XY}) \| = \| \overrightarrow{XY} \| = d(X,Y).$$

Diese Eigenschaft genügt nun auch schon, die Isometrien zu charakterisieren.

Satz 8. *Es seien $\mathbb{A}$ und $\mathbb{B}$ euklidische Räume und $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ eine Abbildung. Genau dann ist φ eine Isometrie, wenn*

$$d(\varphi(X),\varphi(Y)) = d(X,Y)$$

für alle Punkte $X, Y \in \mathbb{A}$ gilt.

Beweis. Zu zeigen ist nur noch die eine Richtung. Seien o.B.d.A. $\mathbb{A} = V$ und $\mathbb{B} = W$, und gelte $d(\varphi(x),\varphi(y)) = d(x,y)$ für alle $x,y \in V$. Wir betrachten die Abbildung $\Phi : V \longrightarrow W$, die durch

$$\Phi(x) := \varphi(x) - \varphi(o) \quad , x \in V,$$

erklärt ist. Es gilt dann:

(i) $$\| \Phi(x) - \Phi(y) \| = \| x - y \| \text{ für alle } x,y \in V,$$

insbesondere wegen $\Phi(o) = o$

(ii) $$\| \Phi(x) \| = \| x \| \text{ für alle } x \in V.$$

Daraus folgt wegen

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

und

$$\| \Phi(x) - \Phi(y) \|^2 - \| \Phi(x) \|^2 - \| \Phi(y) \|^2 = -2 \langle \Phi(x),\Phi(y) \rangle$$

schließlich

(iii) $\quad <\Phi(x),\Phi(y)> \;=\; <x,y>$ für alle $x,y \in V$.

Weiter gilt:

(iv) $\quad \Phi(x+y) = \Phi(x) + \Phi(y)$ für alle $x,y \in V$,

denn $\| \Phi(x+y) - \Phi(x) - \Phi(y) \|^2 = \| \Phi(x+y) \|^2 + \| \Phi(x) \|^2 + \| \Phi(y) \|^2 -$
$2 <\Phi(x+y),\Phi(x)> - 2 <\Phi(x+y),\Phi(y)> + 2 <\Phi(x),\Phi(y)> = \| x+y \|^2 + \| x \|^2 +$
$\| y \|^2 - 2 <x+y,x> - 2 <x+y,y> + 2 <x,y> = \| (x+y) - x - y \| = 0$.

(v) $\quad \Phi(a\,x) = a\,\Phi(x)$ für alle $x \in V$, $a \in \mathbb{R}$,

denn $\| \Phi(a\,x) - a\,\Phi(x) \|^2 = \| \Phi(a\,x) \|^2 + a^2 \| \Phi(x) \|^2 - 2\,a <\Phi(a\,x),\Phi(x)> =$
$\| a\,x \|^2 + a^2 \| x \|^2 - 2\,a <a\,x,x> = 2\,a^2 \| x \|^2 - 2\,a^2 \| x \|^2 = 0.$

Aus (iv) und (v) folgt nun, daß Φ linear ist. Wegen (iii) ist Φ sogar Isometrie. Also ist auch die Abbildung $\varphi : \mathbb{A} \longrightarrow \mathbb{B}$ wegen $\varphi(x) = \Phi(x) + w$ mit $w = \varphi(o)$ eine Isometrie. ∎

Die Bewegungen eines euklidischen Raumes $\mathbb{A}$ bilden ebenfalls eine Gruppe. Sie ist eine Untergruppe der Gruppe der Affinitäten, weshalb sich mit ihrer Hilfe die Einteilung der Teilmengen von $\mathbb{A}$ in affine Äquivalenzklassen weiter verfeinern läßt.

Wir nennen die Teilmengen M_1, $M_2 \subset \mathbb{A}$ *kongruent*, wenn es eine Bewegung φ von $\mathbb{A}$ gibt mit $\varphi(M_1) = M_2$.

Dieses Einteilungsprinzip wird in § 6.4 bei der euklidischen Klassifikation der Quadriken angewendet werden.

§ 3 Quadriken in affinen Räumen

Bisher haben wir in affinen Räumen als Punktmengen hauptsächlich die affinen Unterräume untersucht. Sie lassen sich bei gegebenem Koordinatensystem durch lineare Gleichungssysteme beschreiben. Nun wollen wir einen Schritt weitergehen und Punktmengen betrachten, die sich durch die einfachsten nichtlinearen Gleichungen, also durch quadratische Gleichungen beschreiben lassen; dies sind die Quadriken.

Zu den Quadriken zählen die elementaren geometrischen Gebilde, die man vielleicht noch von der Schule her kennt, wie Kreis, Ellipse, Hyperbel, Kugel, Ellipsoid, Paraboloid usw. Auf solche Gebilde stößt man vielfach, wenn man Abstandsaufgaben in der Ebene oder im Raum lösen will. Dadurch spielen Quadriken auch bei vielen technischen Konstruktionen eine Rolle. Quadriken erhält man auch, wenn differenzierbare Flächen lokal approximiert werden sollen. In erster Näherung ergibt sich die Tangentialhyperebene, in zweiter Näherung eine Quadrik, die die Krümmung beschreibt.

Quadriken können in beliebigen affinen Räumen definiert werden, im Hinblick auf die obengenannten Anwendungen beschränken wir uns aber hier auf reelle affine Räume und beginnen mit dem $\mathbb{R}^n$.

Die allgemeinste Form einer quadratischen Gleichung in $\mathbb{R}^n$ ist

$$\sum_{i,\,j=1}^{n} a_{ij}\, x_i\, x_j \;+\; 2 \sum_{j=1}^{n} b_j\, x_j \;+\; c \;=\; 0$$

mit reellen Variablen x_j und reellen Koeffizienten a_{ij}, b_j, c, wobei mindestens einer der Ausdrücke $a_{ij} + a_{ji}$ von Null verschieden ist. Setzen wir

$$\tilde{a}_{ij} := (a_{ij} + a_{ji})/\,2 \quad , \quad i,j = 1,\ldots,n \,,$$

und

$$b \;=\; \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \quad x \;=\; \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

so läßt sich die obige Gleichung auch in der Form

$$(*) \qquad x^\top A\, x \;+\; 2\, b^\top x \;+\; c \;=\; 0$$

schreiben, wobei $A = (\!(\tilde{a}_{ij})\!)$ eine symmetrische Matrix ist und $A \neq O$ gilt.

Die allgemeine quadratische Gleichung im $\mathbb{R}^n$ besteht also aus einem reinquadratischen Anteil, gegeben durch die von der Nullabbildung verschiedene symmetrische Bilinearform $\beta : (x,y) \longmapsto x^\top A\, y$, aus einem linearen Anteil, gegeben durch die Linearform $\Phi : x \longmapsto b^\top x$, und aus einer Konstanten $c \in \mathbb{R}$. Damit können wir die Gleichung (*) auch so schreiben:

$$(**) \qquad \beta(x,x) \;+\; 2\,\Phi(x) \;+\; c \;=\; 0\,.$$

Interpretieren wir die Lösungen von (*) bzw. (**) als Punkte des affinen Standardraumes $\mathbb{R}^n$, so haben wir stillschweigend das affine Standardkoordinatensystem $(o;e_1,\ldots,e_n)$ ausgezeichnet und fassen $x_1,\ldots,x_n$ als Koordinaten des Punktes x bezüglich dieses Koordinatensystems auf. Ist die so beschriebene Punktmenge nicht leer, so sprechen wir von einer Quadrik in $\mathbb{R}^n$.

Im allgemeinen Fall eines beliebigen n–dimensionalen reellen affinen Raumes $\mathbb{A}$ mit zugehörigem Vektorraum V legen wir entsprechend eine Gleichung der Form (**) zugrunde und fassen jede Lösung $x \in V$ als Ortsvektor $x = \overrightarrow{OX}$ eines Punktes $X \in \mathbb{A}$ bezüglich eines ausgezeichneten Ursprungs $O \in \mathbb{A}$ auf. Damit wird durch (**) eine Punktmenge in $\mathbb{A}$ beschrieben, die wir dann ebenfalls eine Quadrik nennen.

Definition. Es seien $\mathbb{A}$ ein reeller n–dimensionaler affiner Raum mit zugehörigem Vektorraum V, $\beta : V \times V \longrightarrow \mathbb{R}$ eine von der Nullabbildung verschiedene symmetrische Bilinearform, $\Phi : V \longrightarrow \mathbb{R}$ eine Linearform und $c \in \mathbb{R}$ eine Konstante. Ferner sei ein Punkt $O \in \mathbb{A}$ als Ursprung ausgezeichnet. Ist die Punktmenge

$$\mathcal{Q} \;=\; \{X \in \mathbb{A} \mid \overrightarrow{OX} = x\,,\, \beta(x,x) \;+\; 2\,\Phi(x) \;+\; c \;=\; 0\,\}$$

nicht leer, so heißt sie eine *Quadrik*. In der affinen Ebene heißt eine Quadrik auch *Kegelschnitt*. Ist die affine Hülle von $\mathcal{Q}$ der gesamte Raum $\mathbb{A}$, so heißt $\mathcal{Q}$ eine

eigentliche Quadrik.

Wir werden uns hauptsächlich mit den eigentlichen Quadriken beschäftigen. Quadriken, die nicht eigentlich sind, liegen in einer Hyperebene von $\mathbb{A}$, es wird sich sogar herausstellen, daß sie selbst affine Unterräume von $\mathbb{A}$ sind, also durch lineare Gleichungen beschrieben werden können.

Bemerkung. Die definierende Gleichung eine Quadrik $\mathcal{Q}$ ist nicht eindeutig bestimmt. Multiplizieren wir sie etwa mit $a \neq 0$, so erhalten wir dieselbe Punktmenge.

Zeichnen wir in V eine Basis $(v_1,\ldots,v_n)$ aus, so erfüllen die Koordinaten der Punkte von $\mathcal{Q}$ bezüglich des affinen Koordinatensystems $(O; v_1,\ldots,v_n)$ eine Gleichung der Form (*) mit $A = (\!(\beta(v_i,v_j))\!)$ und $b^\top = (\Phi(v_1)\cdots\Phi(v_n))$. Diesen Zusammenhang werden wir im folgenden häufig ausnutzen.

Das weitere Programm ist nun vorgezeichnet. Wir suchen alle Äquivalenzklassen affin–äquivalenter Quadriken in $\mathbb{A}$ und in jeder Äquivalenzklasse einen Repräsentanten, der durch eine möglichst einfache quadratische Gleichung beschrieben wird. Diese nennen wir dann die affine Normalform von $\mathcal{Q}$.

Wir beginnen mit dem Nachweis, daß Quadriken unter bijektiven affinen Abbildungen in Quadriken übergehen.

Satz 9. *Es seien $\mathcal{Q} \subset \mathbb{A}$ eine Quadrik und $\varphi : \mathbb{A} \longrightarrow \mathbb{A}$ eine Affinität. Dann ist auch $\varphi(\mathcal{Q})$ eine Quadrik.*

Beweis. $\mathcal{Q}$ werde bezüglich des Ursprungs O durch die quadratische Gleichung

$$\beta(x,x) + 2\,\Phi(x) + c = 0$$

beschrieben und für die Affinität φ gelte für alle $X \in \mathbb{A}$

$$\overrightarrow{O\varphi(X)} = \Psi(\overrightarrow{OX}) + v$$

mit einem Isomorphismus $\Psi \in \mathrm{Hom}(V,V)$ und $v \in V$. Dann gilt:

$$Y = \varphi(X) \in \varphi(\mathcal{Q}) \iff X = \varphi^{-1}(Y) \in \mathcal{Q}\,.$$

Durch Einsetzen von $x = \overrightarrow{OX} = \Psi^{-1}(\overrightarrow{OY}) - \Psi^{-1}(v)$ in die obige Gleichung erhalten wir für die Ortsvektoren $y = \overrightarrow{OY}$ der Punkte Y von $\varphi(\mathcal{Q})$

$$\beta(\Psi^{-1}(y) - \Psi^{-1}(v), \Psi^{-1}(y) - \Psi^{-1}(v)) \; + \; 2\,\Phi(\Psi^{-1}(y) - \Psi^{-1}(v)) \; + \; c \; = \; 0\,,$$

also

$$\beta(\Psi^{-1}(y), \Psi^{-1}(y)) \; + \; 2\,\{\Phi \circ \Psi^{-1}(y) - \beta(\Psi^{-1}(y), \Psi^{-1}(v))\}$$

$$+ \; \beta(\Psi^{-1}(v), \Psi^{-1}(v)) \; - 2\,\Phi \circ \Psi^{-1}(v) \; + \; c \; = 0\,.$$

Dies ist eine quadratische Gleichung, also ist $\varphi\,(\mathcal{Q})$ eine Quadrik. ∎

Bemerkung. Ist $\hat{x}^{\mathsf{T}} A\,\hat{x} \; + \; 2\,b^{\mathsf{T}}\,\hat{x} \; + \; c \; = \; 0$ die Matrixdarstellung von $\mathcal{Q}$ bezüglich eines affinen Koordinatensystems $(O;\, v_1,\ldots,v_n)$ mit dem ausgezeichneten Punkt O als Ursprung und sind entsprechend

$$\hat{x} \; \longmapsto \; B\,\hat{x} + \hat{v} \;\; , \;\; \hat{x} \in \mathbb{R}^n,$$

bzw.

$$\hat{y} \; \longmapsto \; B^{-1}\,\hat{y} - B^{-1}\,\hat{v} \;\; , \;\; \hat{y} \in \mathbb{R}^n,$$

die Matrixdarstellungen von φ bzw. φ^{-1}, so gilt für die Gleichung von $\varphi\,(\mathcal{Q})$ bezüglich $(O;\, v_1,\ldots,v_n)$

$$(***) \qquad \hat{y}^{\mathsf{T}} C^{\mathsf{T}} A\; C\,\hat{y} \; + \; 2\,(b^{\mathsf{T}} C + \hat{u}^{\mathsf{T}} A\; C)\,\hat{y} \; + \; (\hat{u}^{\mathsf{T}} A\,\hat{u} + 2\,b^{\mathsf{T}}\,\hat{u} + c) \; = \; 0\,,$$

wobei $C = B^{-1}$ und $\hat{u} = -B^{-1}\,\hat{v}$ ist.

Fassen wir die Affinität φ als Koordinatenwechsel auf, so wird die Quadrik $\mathcal{Q}$ bezüglich des neuen Koordinatensystems $(\varphi^{-1}(O);\, \Psi^{-1}(v_1),\ldots,\Psi^{-1}(v_n))$ durch die Gleichung (***) dargestellt (vgl. S.298 unten).

Ist φ insbesondere eine Translation, also $\Psi = \mathrm{id}_V$ und damit $B = C = E_n$, so hat (***) die Form

$$\hat{y}^{\mathsf{T}} A\,\hat{y} \; + \; 2\,(b^{\mathsf{T}} + \hat{u}^{\mathsf{T}} A)\,\hat{y} \; + \; (\hat{u}^{\mathsf{T}} A\,\hat{u} + 2\,b^{\mathsf{T}}\,\hat{u} + c) \; = \; 0\,,$$

d.h. bei einer Verschiebung des Urprungs ändert sich an dem quadratischen Teil der Gleichung von $\mathcal{Q}$ nichts, wohl aber an dem linearen Teil und an der Konstanten.

Bleibt der Ursprung fest, so bleibt die Konstante unverändert.

Unser Ziel wird nun sein, den Koordinatenwechsel so vorzunehmen, daß die Gleichung von $\mathcal{Q}$ eine möglichst einfache Gestalt annimmt, oder anders interpretiert, eine Affinität φ so zu wählen, daß die Gleichung von $\varphi(\mathcal{Q})$ möglichst einfach wird.

Satz 10 (Satz über die affine Hauptachsentransformation von Quadriken). *Es sei* $\mathbb{A}$ *ein* n*–dimensionaler reeller affiner Raum. Dann läßt sich jede Quadrik* $\mathcal{Q} \subset \mathbb{A}$ *bezüglich eines geeigneten affinen Koordinatensystems durch eine der folgenden Gleichungen beschreiben:*

$$\text{(I)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 0 \;, \qquad 0 < p \leq r \leq n,\ r-p \leq p \;,$$

$$\text{(II)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 1 \;, \qquad 0 < p \leq r \leq n \;,$$

$$\text{(III)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 2\, x_n \;, \qquad 0 < p \leq r < n,\ r-p \leq p \;.$$

Beweis. Die Quadrik $\mathcal{Q}$ werde nach Auszeichnung eines Koordinatensystems $(O; v_1,\ldots,v_n)$ durch die Gleichung

$$\hat{x}^{\mathsf{T}} A\, \hat{x} + 2\, b^{\mathsf{T}}\, \hat{x} + c = 0$$

dargestellt, wo $A = (\!(\beta(v_i,v_j))\!)$ eine symmetrische Matrix ist. Nach Satz 5.13 gibt es eine orthogonale Matrix S, so daß $A' = S^{\mathsf{T}} A\, S$ Diagonalgestalt hat. In der Diagonalen stehen dann die Eigenwerte $c_1,\ldots,c_n$ von A.

Für den Vektorraum V bedeutet dies die Existenz einer Basis $(v_1',\ldots,v_n')$, so daß die Matrix $A' = (\!(\beta(v_i',v_j'))\!)$ Diagonalgestalt hat mit $\beta(v_i',v_i') = c_i$, $i = 1,\ldots,n$. Wegen $\beta \neq O$ ist mindestens einer der Eigenwerte c_i von Null verschieden. Sei o.B.d.A. $c_1,\ldots,c_p > 0$, $c_{p+1},\ldots,c_r < 0$ und $c_{r+1} = \ldots = c_n = 0$. Es ist $r = \operatorname{Rg} A' = \operatorname{Rg} A > 0$, die Zahl r hängt also nicht von der Wahl der Basis ab, sondern nur von der Bilinearform β. Ersetzen wir die Basisvektoren v_i' durch neue Vektoren $\overline{v}_i$,

$$\overline{v}_i = \frac{1}{\sqrt{|c_i|}}\, v_i' \;, \quad i = 1,\ldots,r \;,$$

$$\overline{v}_i = v_i' \;, \quad i = r+1,\ldots,n \;,$$

so nimmt die Gleichung von $\mathcal{Q}$ bezüglich des neuen Koordinatensystems $(O; \overline{v}_1,...,\overline{v}_n)$ die Form

$$\overline{x}_1^2 + \cdots + \overline{x}_p^2 - \overline{x}_{p+1}^2 - \cdots - \overline{x}_r^2 + 2\, b_1\, \overline{x}_1 + \cdots + 2\, b_n\, \overline{x}_n + c = 0$$

an mit neuen Konstanten $b_1,...,b_k$ und den (neuen) Variablen $\overline{x}_j$. Durch quadratische Ergänzung erhalten wir weiter

$$(\overline{x}_1 + b_1)^2 + \cdots + (\overline{x}_p + b_p)^2 - (\overline{x}_{p+1} - b_{p+1})^2 - \cdots - (\overline{x}_r - b_r)^2$$
$$+ 2\, b_{r+1}\, \overline{x}_{r+1} + \cdots + 2\, b_n\, \overline{x}_n + c = 0\,,$$

wobei sich jetzt auch die Konstante c verändert haben kann. Nun verschieben wir den Ursprung O durch eine Translation, die bezüglich $(O; \overline{v}_1,...,\overline{v}_n)$ durch die folgenden Gleichungen definiert ist:

$$y_i = \overline{x}_i + b_i\,, \quad i = 1,...,p\,,$$

$$y_i = \overline{x}_i - b_i\,, \quad i = p+1,...,r\,,$$

$$y_i = \overline{x}_i\,, \quad i = r+1,...,n\,.$$

Wir erhalten

$$y_1^2 + \cdots + y_p^2 - y_{p+1}^2 - \cdots - y_r^2 + 2\, b_{r+1}\, y_{r+1} + \cdots + 2\, b_n\, y_n + c = 0\,.$$

1. Fall: $b_{r+1} = ... = b_n = c = 0$. Dann besitzt $\mathcal{Q}$ bezüglich des zuletzt betrachteten affinen Koordinatensystems die Gleichung

(I) $$y_1^2 + \cdots + y_p^2 - y_{p+1}^2 - \cdots - y_r^2 = 0\,.$$

Dabei gilt o.B.d.A. $r - p \leq p$, denn andernfalls multiplizieren wir die Gleichung mit -1.

2. Fall: $b_{r+1} = ... = b_n = 0$, $c \neq 0$. Hier können wir entsprechend annehmen, daß $c < 0$ gilt. Wir dividieren durch $-c$ und nach einer einer weiteren Koordinatentransformation

$$z_i = (-c)^{-\frac{1}{2}}\, y_i\,, \quad i = 1,...,r\,,$$

$$z_i = y_i\,, \quad i = r+1,...,n\,,$$

erhalten wir die Gleichung

$$\text{(II)} \qquad z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2 = 1 .$$

Wegen $\mathcal{Q} \neq \emptyset$ ist dann $p > 0$.

3. Fall: Mindestens einer der Koeffizienten b_i , $i = r+1,\ldots,n$, ist von 0 verschieden. Dann ist $r < n$. Wir können annehmen, daß $b_n \neq 0$ gilt, andernfalls numerieren wir um, was wiederum einem Wechsel des Koordinatensystems entspricht. Setzen wir jetzt

$$z_i = y_i \ , \ i = 1,\ldots,n-1 ,$$

$$z_n = -(b_{r+1}\, y_{r+1} + \ldots + b_n\, y_n) - \frac{c}{2} ,$$

so erhalten wir die Gleichung

$$\text{(III)} \qquad z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2 = 2\, z_n$$

mit $0 < r < n$. Weiter können wir erreichen, daß $r - p \leq p$ gilt, indem wir gegebenenfalls die Gleichung mit -1 multiplizieren und

$$\overline{z}_i = z_i \ , \ i = 1,\ldots,n-1 ,$$

$$\overline{z}_n = -z_n$$

setzen. ∎

Bezeichnungen und Bemerkungen. **(a)** Die Gleichungen (I), (II), (III) heißen, im Vorgriff auf Satz 15, die *affinen Normalformen* einer Quadrik. Zwei Quadriken $\mathcal{Q}$ und $\tilde{\mathcal{Q}}$, die bezüglich geeigneter Koordinatensysteme dieselbe Normalform besitzen, sind offensichtlich affin äquivalent.

(b) Die Zahl r hängt, wie wir gesehen haben, nur von β ab. Für den Untervektorraum

$$\operatorname{Rad} \beta := \{x \in V \mid \beta(x,y) = 0 \text{ für alle } y \in V\},$$

der das *Radikal* oder der *Ausartungsraum* von β heißt, gilt $\dim \operatorname{Rad} \beta = n - r$.

(c) Mit Hilfe der affinen Normalformen können wir nun leicht klären, welche der Quadriken eigentlich sind und welche nicht. Der Einfachheit wegen bezeichnen wir

das ausgezeichnete Koordinatensystem in allen drei Fällen mit $(O; v_1,...,v_n)$.

Fall (I): Hier ist $\mathcal{Q}$ für $p = r$ ein affiner Unterraum der Dimension $n - r$. Für $p < r$ ist $\mathcal{Q}$ eine eigentliche Quadrik, denn auf $\mathcal{Q}$ liegen die $n + 1$ affin unabhängigen Punkte $O, Y_1,..., Y_n$, wobei $Y_1,..., Y_n$ durch

$$\overrightarrow{OY_i} = v_i - v_r \ , \quad i = 1,...,p \ ,$$

$$\overrightarrow{OY_i} = v_1 + v_i \ , \quad i = p+1,...,r \ ,$$

$$\overrightarrow{OY_i} = v_i \ , \quad i = r+1,...,n \ .$$

gegeben sind.

Fall (II): $\mathcal{Q}$ ist hier stets eine eigentliche Quadrik, denn es liegen auf $\mathcal{Q}$ die $n + 1$ affin unabhängigen Punkte $Y_0,..., Y_n$ mit den Ortsvektoren

$$\overrightarrow{OY_0} = -v_1 \ ,$$

$$\overrightarrow{OY_i} = v_i \ , \quad i = 1,...,p \ ,$$

$$\overrightarrow{OY_i} = \sqrt{2}\, v_1 + v_i \ , \quad i = p+1,...,r \ ,$$

$$\overrightarrow{OY_i} = v_1 + v_i \ , \quad i = r+1,...,n \ .$$

Fall (III): Auch in diesem Fall ist $\mathcal{Q}$ immer eigentlich, denn die $n + 1$ Punkte $O, Y_1,..., Y_n$ mit

$$\overrightarrow{OY_i} = v_i + \tfrac{1}{2} v_n \ , \quad i = 1,...,p \ ,$$

$$\overrightarrow{OY_i} = v_i - \tfrac{1}{2} v_n \ , \quad i = p+1,...,r \ ,$$

$$\overrightarrow{OY_i} = v_i \ , \quad i = r+1,...,n-1 \ ,$$

$$\overrightarrow{OY_n} = 2\,(v_1 + v_n) \ .$$

liegen auf $\mathcal{Q}$ und sind affin unabhängig.

Beispiele. Für $n = 1,2,3$ ergeben sich aus Satz 10 die folgenden affinen Normalformen von Quadriken:

$\boldsymbol{n = 1}$: (I) $x_1^2 = 0$, (*Punkt*, $r = 1$, $p = 1$)

(II) $x_1^2 = 1$. (*Punktepaar*, $r = 1$, $p = 1$)

$n = 2$:

(I) $x_1^2 + x_2^2 = 0$, (*Punkt* , $r = 2, p = 2$)

$x_1^2 - x_2^2 = 0$, (*Paar sich schneidender Geraden* , $r = 2, p = 1$)

$x_1^2 = 0$, (*Gerade* , $r = 1, p = 1$)

(II) $x_1^2 + x_2^2 = 1$, (*Ellipse* , $r = 2, p = 2$)

$x_1^2 - x_2^2 = 1$, (*Hyperbel* , $r = 2, p = 1$)

$x_1^2 = 1$, (*Paar paralleler Geraden* , $r = 1, p = 1$)

(III) $x_1^2 = 2\, x_2$. (*Parabel* , $r = 1, p = 1$)

$n = 3$:

(I) $x_1^2 + x_2^2 + x_3^2 = 0$, (*Punkt* , $r = 3, p = 3$)

$x_1^2 + x_2^2 - x_3^2 = 0$, (*Kegel* , $r = 3, p = 2$)

$x_1^2 + x_2^2 = 0$, (*Gerade* , $r = 2, p = 2$)

$x_1^2 - x_2^2 = 0$, (*Paar sich schneidender Ebenen* , $r = 2, p = 1$)

$x_1^2 = 0$, (*Ebene* , $r = 1, p = 1$)

(II) $x_1^2 + x_2^2 + x_3^2 = 1$, (*Ellipsoid* , $r = 3, p = 3$)

$x_1^2 + x_2^2 - x_3^2 = 1$, (*einschaliges Hyperboloid* , $r = 3, p = 2$)

$x_1^2 - x_2^2 - x_3^2 = 1$, (*zweischaliges Hyperboloid* , $r = 3, p = 1$)

$x_1^2 + x_2^2 = 1$, (*elliptischer Zylinder* , $r = 2, p = 2$)

$x_1^2 - x_2^2 = 1$, (*hyperbolischer Zylinder* , $r = 2, p = 1$)

$x_1^2 = 1$, (*Paar paralleler Ebenen* , $r = 1, p = 1$)

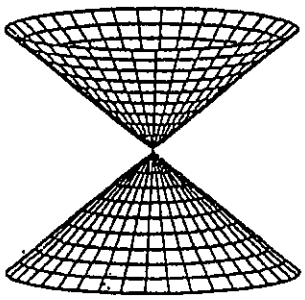
(III) $x_1^2 + x_2^2 = 2\, x_3$, (*elliptisches Paraboloid* , $r = 2, p = 2$)

$x_1^2 - x_2^2 = 2\, x_3$, (*hyperbolisches Paraboloid* , $r = 2, p = 1$)

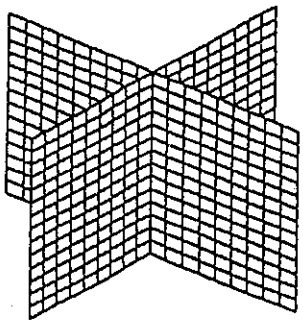
$x_1^2 = 2\, x_3$. (*parabolischer Zylinder* , $r = 1, p = 1$)

Eigentliche Quadriken im dreidimensionalen reellen affinen Raum.

I Kegel

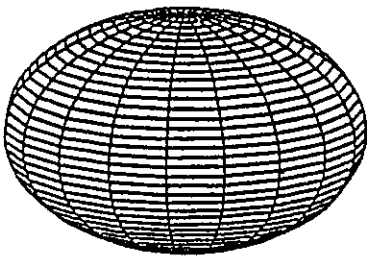


Kegel

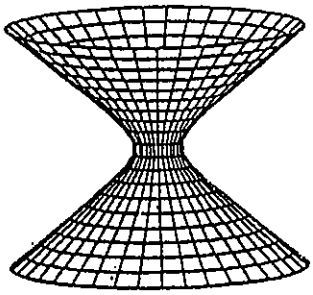


Paar sich schneidender Ebenen

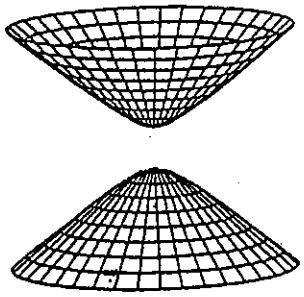
II Echte Mittelpunktsquadriken



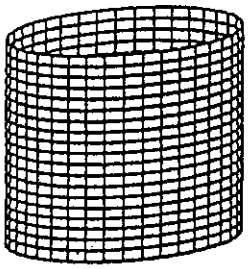
Ellipsoid



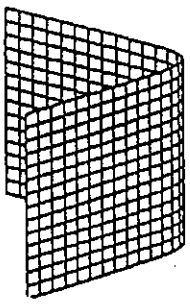
einschaliges Hyperboloid

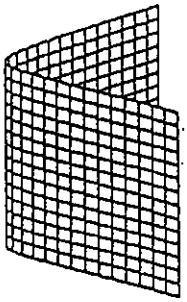


zweischaliges Hyperboloid

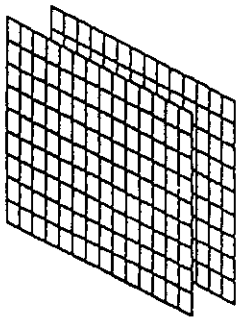


elliptischer Zylinder



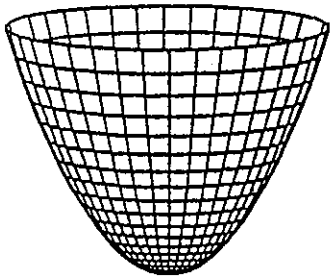


hyperbolischer Zylinder

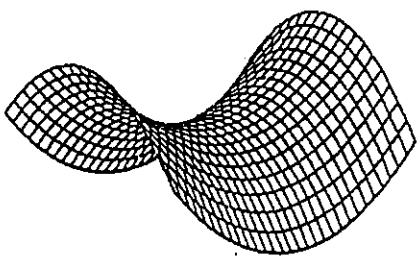


Paar paralleler Ebenen

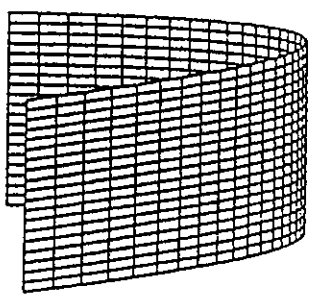
III Paraboloide



elliptisches Paraboloid



hyperbolisches Paraboloid



parabolischer Zylinder

Für die praktische Herleitung der Gleichungen (I), (II), (III) gibt es ein einfaches Verfahren, das mit quadratischer Ergänzung arbeitet. Statt dieses Verfahren allgemein zu behandeln, wollen wir es an zwei Beispielen demonstrieren.

Beispiele. **(a)** Im affinen Raum $\mathbb{R}^3$ sei bezüglich eines affinen Koordinatensystems $(O; v_1, v_2, v_3)$ für jedes $t \in \mathbb{R}$ eine Quadrik $\mathcal{Q}_t$ durch die Gleichung

$$2\,x_1^2 + 4\,x_1\,x_2 - 6\,x_1\,x_3 + 3\,x_2^2 - 8\,x_2\,x_3 + \tfrac{5}{2}\,x_3^2 + 2\,t\,x_3 + t^2 \;=\; 0$$

gegeben. Wir beginnen mit $2\,x_1^2$ und betrachten alle anderen Summanden, in denen x_1 als Faktor auftritt. Durch quadratische Ergänzung erhalten wir

$$2\,[x_1^2 + 2\,x_1\,(x_2 - \tfrac{3}{2}\,x_3) + (x_2 - \tfrac{3}{2}\,x_3)^2] - 2\,(x_2 - \tfrac{3}{2}\,x_3)^2$$

$$+ 3\,x_2^2 - 8\,x_2\,x_3 + \tfrac{5}{2}\,x_3^2 + 2\,t\,x_3 + t^2 \;=\; 0$$

und somit

$$2\,(x_1 + x_2 - \tfrac{3}{2}\,x_3)^2 + x_2^2 - 2\,x_2\,x_3 - 2\,x_3^2 + 2\,t\,x_3 + t^2 = 0\,.$$

In einer weiteren quadratischen Ergänzung berücksichtigen wir x_2^2 und alle Summanden, in denen x_2 als Faktor vorkommt:

$$2\,(x_1 + x_2 - \tfrac{3}{2}\,x_3)^2 + (x_2^2 - 2\,x_2\,x_3 + x_3^2) - 3\,x_3^2 + 2\,t\,x_3 + t^2 = 0\,.$$

Eine letzte quadratische Ergänzung ergibt dann

$$2\,(x_1 + x_2 - \tfrac{3}{2}\,x_3)^2 + (x_2 - x_3)^2 - 3\,(x_3^2 - \tfrac{2}{3}\,t\,x_3 + \tfrac{1}{9}\,t^2) + \tfrac{1}{3}\,t^2 + t^2 \;=\; 0\,.$$

Setzen wir

$$y_1 \;=\; \sqrt{2}\,(x_1 + x_2 - \tfrac{3}{2}\,x_3)\,,$$

$$y_2 \;=\; x_2 - x_3\,,$$

$$y_3 \;=\; \sqrt{3}\,(x_3 - \tfrac{1}{3}\,t)\,,$$

so erhalten wir für $\mathcal{Q}_t$ die Gleichung

$$y_1^2 + y_2^2 - y_3^2 + \frac{4}{3}t^2 = 0 .$$

Für $t = 0$ hat diese Gleichung Normalgestalt (I). Es handelt sich um einen Kegel mit $p = 2$ und $r = 3$.

Für $t \neq 0$ dividieren wir durch $-\frac{4}{3}t^2$, setzen

$$z_1 = \frac{\sqrt{3}}{2t} y_3 \ , \ z_2 = \frac{\sqrt{3}}{2t} y_2 \ , \ z_3 = \frac{\sqrt{3}}{2t} y_1$$

und erhalten die Normalform

$$z_1^2 - z_2^2 - z_3^2 = 1 .$$

Dies ist eine Quadrik vom Typ (II), $p = 1$, $r = 3$, also ein zweischaliges Hyperboloid.

Die obige Rechnung liefert uns für jeden Fall auch gleichzeitig die Affinität φ, für die $\varphi(\mathcal{Q})$ Normalform hat, bzw. das neue affine Koordinatensystem $(O';v_1',v_2',v_3')$, bezüglich dessen die Gleichung von $\mathcal{Q}$ Normalform annimmt.

Für $t = 0$ ist die gesuchte affine Abbildung φ durch die Gleichungen

$$z_1 = \sqrt{2}\, x_1 + \sqrt{2}\, x_2 - \frac{3}{\sqrt{2}} x_3 ,$$

$$z_2 = x_2 - x_3 ,$$

$$z_3 = \sqrt{3}\, x_3$$

gegeben, hat also bezüglich des Koordinatensystems $(O; v_1,v_2,v_3)$ die Matrixdarstellung

$$\hat{x} \longmapsto \begin{bmatrix} \sqrt{2} & \sqrt{2} & -3/\sqrt{2} \\ 0 & 1 & -1 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \hat{x} \ , \ \hat{x} \in \mathbb{R}^3 .$$

Dann hat φ^{-1} die Form

$$\hat{x} \longmapsto \begin{bmatrix} 1/\sqrt{2} & -1 & 1/2\sqrt{3} \\ 0 & 1 & 1/\sqrt{3} \\ 0 & 0 & 1/\sqrt{3} \end{bmatrix} \hat{x} \ , \ \hat{x} \in \mathbb{R}^3 .$$

Für das neue Koordinatensystem $(O'; v_1', v_2', v_3')$ gilt somit $O' = \varphi^{-1}(O) = O$ und

$$v_1' = 1/\sqrt{2}\, v_1 ,$$

$$v_2' = -v_1 + v_2 ,$$

$$v_3' = 1/2\sqrt{3}\, v_1 + 1/\sqrt{3}\, v_2 + 1/\sqrt{3}\, v_3 .$$

Für $t \neq 0$ ist φ durch die Gleichungen

$$v_1' = \frac{3}{2t} v_3 - \frac{1}{2},$$

$$v_2' = \frac{\sqrt{3}}{2t} v_2 - \frac{\sqrt{3}}{2t} v_3 ,$$

$$v_3' = \frac{\sqrt{6}}{2t} v_1 + \frac{\sqrt{6}}{2t} v_2 - \frac{3\sqrt{6}}{4t} v_3$$

gegeben. Hier hat φ bezüglich $(O; v_1, v_2, v_3)$ die Matrixdarstellung

$$\hat{x} \longmapsto \frac{\sqrt{3}}{2t} \begin{bmatrix} 0 & 0 & \sqrt{3} \\ 0 & 1 & -1 \\ \sqrt{2} & \sqrt{2} & -3/\sqrt{2} \end{bmatrix} \hat{x} + \begin{bmatrix} -1/2 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{x} \in \mathbb{R}^3.$$

Für φ^{-1} erhalten wir somit die Darstellung

$$\hat{x} \longmapsto \frac{2t}{\sqrt{3}} \begin{bmatrix} 1/2\sqrt{3} & -1 & 1/\sqrt{2} \\ 1/\sqrt{3} & 1 & 0 \\ 1/\sqrt{3} & 0 & 0 \end{bmatrix} \hat{x} + \begin{bmatrix} t/6 \\ t/3 \\ t/3 \end{bmatrix}, \quad \hat{x} \in \mathbb{R}^3$$

Also ist $\overrightarrow{OO'} = \frac{t}{6} v_1 + \frac{t}{3} v_2 + \frac{t}{3} v_3$ und

$$v_1' = \frac{t}{3} v_1 + \frac{2t}{3} v_2 + \frac{2t}{3} v_3,$$

$$v_2' = -\frac{2t}{\sqrt{3}} v_1 + \frac{2t}{\sqrt{3}} v_2 ,$$

$$v_3' = \frac{2t}{\sqrt{6}} v_1 .$$

(b) Im affinen Raum $\mathbb{R}^3$ sei bezüglich eines affinen Koordinatensystems eine Quadrik $\mathcal{Q}$ durch die Gleichung

$$x_1 x_2 + x_1 x_3 + x_2 x_3 + 2 x_1 - 1 = 0$$

gegeben. Um hier mit quadratischer Ergänzung arbeiten zu können, müssen wir uns zunächst geeignete quadratische Terme verschaffen. Dies geschieht mit Hilfe der folgenden Koordinatentransformation

$$\begin{aligned} x_1 &= y_1 - y_2 \\ x_2 &= y_1 + y_2 \\ x_3 &= y_3 \end{aligned}$$

Die neue Gleichung für $\mathcal{Q}$ hat dann die Form

$$y_1^2 - y_2^2 + 2 y_1 y_3 + 2 y_1 - 2 y_2 - 1 = 0 ,$$

auf die wir nun das Verfahren der quadratischen Ergänzung anwenden können. Es ergibt sich als Normalform

$$z_1^2 + z_2^2 - z_3^2 = 0,$$

also ist $\mathcal{Q}$ ein Kegel.

§ 4 Affine Klassifikation der Quadriken

Mit Satz 10 ist die affine Klassifikation der Quadriken noch nicht vollständig. Dazu muß noch gezeigt werden, daß zwei Quadriken genau dann affin äquivalent sind, wenn ihre Gleichungen dieselbe Normalform besitzen.

Die eine Richtung ist einfach. Besitzen $\mathcal{Q}$ und $\tilde{\mathcal{Q}}$ bezüglich geeigneter affiner Koordinatensysteme dieselbe Normalform, so bildet die Affinität φ, die diese Koordinatensysteme aufeinander abbildet, auch $\mathcal{Q}$ auf $\tilde{\mathcal{Q}}$ ab oder umgekehrt $\tilde{\mathcal{Q}}$ auf $\mathcal{Q}$.

Die andere Richtung ist komplizierter. Gilt $\varphi(\mathcal{Q}) = \tilde{\mathcal{Q}}$ mit einer Affinität φ, so müssen wir zeigen, daß die Normalformen von $\mathcal{Q}$ und von $\tilde{\mathcal{Q}}$ zur gleichen Klasse (I), (II) oder (III) gehören und daß innerhalb dieser Klassen die Normalformen auch noch in den Größen p und r übereinstimmen.

Für die Zahlen r und $\tilde{r}$ haben wir die Gleichheit schon bewiesen, denn es sind die Ränge der symmetrischen Matrizen A bzw. $\tilde{A}$, welche die quadratischen Anteile von $\mathcal{Q}$ bzw. $\tilde{\mathcal{Q}}$ beschreiben. Nach (***), S. 306, gilt $\tilde{A} = C^{\mathrm{T}} A\, C$ mit einer regulären Matrix C, also ist $\tilde{r} = r$. Da weiterhin $\operatorname{Rg} A = \operatorname{Rg} (a\, A)$ für alle $a \neq 0$ gilt, ist r auch invariant gegenüber Multiplikationen der definierenden Gleichung von $\mathcal{Q}$ mit Zahlen $a \neq 0$.

Für die beiden anderen Behauptungen genügt es nicht mehr, die Gleichungen nur algebraisch umzuformen, vielmehr müssen jetzt auch geometrische Eigenschaften der Quadriken berücksichtigt werden, die invariant sind gegenüber affinen Abbildungen. Eine solche Eigenschaft ist z.B. die, einen Mittelpunkt zu besitzen.

Definition. $M \in \mathbb{A}$ heißt *Mittelpunkt* einer Quadrik $\mathcal{Q}$, wenn die Punktspiegelung an M die Quadrik $\mathcal{Q}$ auf sich abbildet. Eine Quadrik $\mathcal{Q}$, die einen Mittelpunkt besitzt, heißt *Mittelpunktsquadrik*. Eine Mittelpunktsquadrik $\mathcal{Q}$, bei der kein Mittelpunkt auf $\mathcal{Q}$ liegt, heißt *echte Mittelpunktsquadrik*.

Besitzt $\mathcal{Q}$ einen Mittelpunkt $M \in \mathcal{Q}$, so heißt dieser eine *Spitze* oder ein *Doppelpunkt* und $\mathcal{Q}$ heißt *Hyperkegel* oder kurz *Kegel*.

Quadriken ohne Mittelpunkt heißen *parabolisch.*

Beispiel. Jeder affine Unterraum L ist ein Kegel und genau die Punkte von L sind die Spitzen dieser nichteigentlichen Quadrik.

Für eigentliche Quadriken beantwortet der folgende Satz die Frage nach der Existenz von Mittelpunkten.

Satz 11. *Es sei* $\mathcal{Q} = \{X \in \mathbb{A} \mid \overrightarrow{OX} = x\ ,\ \beta(x,x) + 2\,\Phi(x) + c = 0\}$ *eine eigentliche Quadrik. Dann gilt:*

(a) $M \in \mathbb{A}$ *ist genau dann Mittelpunkt von* $\mathcal{Q}$*, wenn* $\beta(\cdot, \overrightarrow{OM}) + \Phi$ *die Nullform ist. Die Menge der Mittelpunkte von* $\mathcal{Q}$ *ist entweder leer oder ein affiner Unterraum von* $\mathbb{A}$ *mit Richtungsraum* Rad β.

(b) $M \in \mathbb{A}$ *ist genau dann eine Spitze von* $\mathcal{Q}$*, wenn* $\beta(\cdot, \overrightarrow{OM}) + \Phi$ *die Nullform ist und* $\Phi(\overrightarrow{OM}) + c = 0$ *gilt. In diesem Fall ist jeder Mittelpunkt von* $\mathcal{Q}$ *eine Spitze.*

Beweis. (a) $M \in \mathbb{A}$ ist genau dann Mittelpunkt von $\mathcal{Q}$, wenn für jedes $X \in \mathcal{Q}$ auch der an M gespiegelte Punkt $\sigma(X)$ zu $\mathcal{Q}$ gehört. Der Ortsvektor von $\sigma(X)$ ist $\overrightarrow{O\sigma(X)} = -\overrightarrow{OX} + 2\ \overrightarrow{OM}$. Somit ist M genau dann Mittelpunkt von $\mathcal{Q}$, wenn für alle Punkte $X \in \mathbb{A}$ aus

$$\beta(\overrightarrow{OX},\overrightarrow{OX}) + 2\,\Phi(\overrightarrow{OX}) + c = 0$$

stets auch

$$\beta(-\overrightarrow{OX} + 2\ \overrightarrow{OM}, -\overrightarrow{OX} + 2\ \overrightarrow{OM}) + 2\,\Phi(-\overrightarrow{OX} + 2\ \overrightarrow{OM}) + c = 0$$

folgt. Dazu äquivalent ist, daß für alle $X \in \mathcal{Q}$

$$(\#) \qquad \beta(\overrightarrow{MX},\overrightarrow{OM}) + \Phi(\overrightarrow{MX}) = 0$$

gilt.

Da $\mathcal{Q}$ eine eigentliche Quadrik ist, gibt es $n + 1$ affin unabhängige Punkte $Y_0,\ldots,Y_n$ auf $\mathcal{Q}$. Für einen Mittelpunkt M von $\mathcal{Q}$ gilt dann notwendigerweise

$$\beta(\overrightarrow{MY_i},\overrightarrow{OM}) + \Phi(\overrightarrow{MY_i}) = 0\ ,\ i = 0,\ldots,n\,.$$

Durch Subtraktion erhalten wir daraus

$$\beta(\overrightarrow{Y_0Y_i},\overrightarrow{OM}) + \Phi(\overrightarrow{Y_0Y_i}) = 0 \quad , \quad i = 1,\ldots,n \, .$$

Da $(\overrightarrow{Y_0Y_1},\ldots,\overrightarrow{Y_0Y_n})$ eine Basis von V ist, muß also die Linearform $\beta(\cdot,\overrightarrow{OM}) + \Phi$ die Nullform sein. Ist umgekehrt dies der Fall, so ist M wegen (#) auch tatsächlich ein Mittelpunkt von $\mathcal{Q}$.

Sei nun L die Menge der Mittelpunkte von $\mathcal{Q}$. Existiert ein Mittelpunkt M von $\mathcal{Q}$, so gilt für jeden anderen Mittelpunkt M' entsprechend $\beta(\cdot,\overrightarrow{OM'}) + \Phi = O$. Durch Subtraktion erhalten wir $\beta(\cdot,\overrightarrow{MM'}) = O$, also $\overrightarrow{MM'} \in \operatorname{Rad}\beta$. Somit gilt

$$L \subset \{X \in \mathbb{A} \mid \overrightarrow{MX} \in \operatorname{Rad}\beta\}.$$

Umgekehrt ist jeder Punkt X mit $\overrightarrow{MX} \in \operatorname{Rad}\beta$ wegen

$$\beta(\cdot,\overrightarrow{OX}) + \Phi \;=\; \beta(\cdot,\overrightarrow{OM}) + \beta(\cdot,\overrightarrow{MX}) + \Phi \;=\; O$$

ein Mittelpunkt von $\mathcal{Q}$, also gilt Gleichheit, und L ist somit ein affiner Unterraum mit Richtungsraum $\operatorname{Rad}\beta$.

(b) Der Mittelpunkt M liegt genau dann auf $\mathcal{Q}$, wenn $\beta(\overrightarrow{OM},\overrightarrow{OM}) + 2\,\Phi(\overrightarrow{OM}) + c = 0$ gilt. Dies ist wegen $\beta(\overrightarrow{OM},\overrightarrow{OM}) + \Phi(\overrightarrow{OM}) = 0$ gleichwertig mit $\Phi(\overrightarrow{OM}) + c = 0$.

Ist M eine Spitze von $\mathcal{Q}$, so gilt für jeden anderen Mittelpunkt S wegen $M \in \mathcal{Q}$ und $\overrightarrow{MS} \in \operatorname{Rad}\beta$:

$$\beta(\overrightarrow{OS},\overrightarrow{OS}) + 2\,\Phi(\overrightarrow{OS}) + c$$

$$= (\beta(\overrightarrow{OM},\overrightarrow{OM}) + 2\,\Phi(\overrightarrow{OM}) + c) + 2\,(\beta(\overrightarrow{MS},\overrightarrow{OM}) + \Phi(\overrightarrow{MS})) + \beta(\overrightarrow{MS},\overrightarrow{MS}) \;=\; 0,$$

also $S \in \mathcal{Q}$. ∎

Bemerkung. Hat die eigentliche Quadrik $\mathcal{Q}$ bezüglich eines affinen Koordinatensystems die Darstellung

$$\hat{x}^{\mathsf{T}} A\,\hat{x} \;+\; 2\,b^{\mathsf{T}}\,\hat{x} \;+\; c \;=\; 0\,,$$

so ist M genau dann Mittelpunkt von $\mathcal{Q}$, wenn sein Koordinatenvektor $\hat{m} \in \mathbb{R}^n$ eine Lösung des inhomogenen LGS $A\,\hat{x} = -b$ ist, und M ist eine Spitze, wenn außerdem $b^{\mathsf{T}}\,\hat{m} + c = 0$ gilt. Ist die Lösungsmenge von $A\,\hat{x} = -b$ nicht leer, so hat sie nach

Satz 11 die Dimension $d = \dim \operatorname{Rad} \beta = n - \operatorname{Rg} A$.

Beispiel. Im affinen Raum $\mathbb{R}^3$ sei bezüglich eines affinen Koordinatensystems für jedes $t \in \mathbb{R}$ eine Quadrik $\mathcal{Q}_t$ gegeben durch die Gleichung

$$x_1^2 - 8\,x_1\,x_2 + 6\,x_1\,x_3 - 9\,x_2^2 - 24\,x_2\,x_3 + t\,x_3^2 + 20\,x_2 + 3\,x_3 + 1 \;=\; 0\,.$$

In Matrixschreibweise lautet diese Gleichung

$$x^{\mathsf{T}} A\,x \;+\; 2\,b^{\mathsf{T}}\,x \;+\; c \;=\; 0$$

mit

$$x \;=\; \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ und } A \;=\; \begin{bmatrix} 1 & -4 & 3 \\ -4 & -9 & -12 \\ 3 & -12 & t \end{bmatrix}\,,\; b \;=\; \frac{1}{2}\begin{bmatrix} 0 \\ 20 \\ 3 \end{bmatrix}\,,\; c = 1\,.$$

$\mathcal{Q}_t$ besitzt genau dann einen Mittelpunkt mit Ortsvektor m, wenn m Lösung des inhomogenen linearen Gleichungssystems $A\,x = -\,b$ ist.

Im vorliegenden Fall lautet die zugehörige erweiterte Matrix

$$\left[\begin{array}{ccc|c} 1 & -4 & 3 & 0 \\ -4 & -9 & -12 & -10 \\ 3 & -12 & t & -3/2 \end{array}\right]$$

Dieses LGS ist genau für $t \neq 9$ lösbar. Also ist $\mathcal{Q}_t$ für $t = 9$ eine Quadrik ohne Mittelpunkt, d.h. parabolisch. Durch das Verfahren der quadratischen Ergänzung erhalten wir in diesem Fall die Normalform

$$z_1^2 - z_2^2 \;=\; 2\,z_3\,,$$

mit $p = 1$ und $r = 2$. $\mathcal{Q}_9$ ist ein hyperbolisches Paraboloid.

Für $t \neq 9$ besitzt $\mathcal{Q}_t$ einen Mittelpunkt und die Menge der Mittelpunkte ergibt sich als Lösung des obigen linearen Gleichungssystems. Wegen Rang $A = 3$ gibt es für jedes $t \neq 9$ genau einen Mittelpunkt mit zugehörigem Ortsvektor

$$m = \begin{bmatrix} (16t - 99)/10(t-9) \\ 2/5 \\ -3/2(t-9) \end{bmatrix}.$$

Genau für $t = \frac{189}{20}$ ist der Mittelpunkt eine Spitze der Quadrik. In diesem Fall ist $\mathcal{Q}$ ein Kegel mit der Normalform

$$z_1^2 + z_2^2 - z_3^2 = 0.$$

Für $t \neq 9$, $t \neq \frac{189}{20}$ erhalten wir echte Mittelpunktsquadriken, und zwar für $9 < t < \frac{189}{20}$ zweischalige Hyperboloide und für $t < 9$ oder $\frac{189}{20} < t$ einschalige Hyperboloide (Beweis als Übungsaufgabe).

Bemerkung. Die Eigenschaft, Mittelpunkt zu sein, ist invariant unter Affinitäten φ.

Beweis. Ist σ die Punktspiegelung an M, so ist $\varphi\sigma\varphi^{-1}$ die Punktspiegelung an $\varphi(M)$. Nun ist M genau dann Mittelpunkt von $\mathcal{Q}$, wenn aus $X \in \mathcal{Q}$ stets auch $\sigma(X) \in \mathcal{Q}$ folgt. Gleichwertig hierzu ist die Bedingung, daß aus $\varphi(X) \in \varphi(\mathcal{Q})$ stets $\varphi\sigma(X) \in \varphi(\mathcal{Q})$ folgt, d.h. $\varphi\sigma\varphi^{-1}(\varphi(X)) \in \varphi(\mathcal{Q})$ gilt. Ist also M Mittelpunkt von $\mathcal{Q}$, so ist $\varphi(M)$ Mittelpunkt von $\varphi(\mathcal{Q})$ und umgekehrt. ∎

Wir wenden nun Satz 11 bzw. die anschließende Bemerkung auf die Quadriken mit den Gleichungen der Form (I), (II) oder (III) an.

Fall (I): Hier gilt Rg $A = r$, $b = o$, $c = 0$ und der Ursprung ist eine Spitze, es handelt sich also um Kegel. Sind die Quadriken eigentlich, so ist nach Satz 11 die Menge der Spitzen jeweils ein affiner Unterraum der Dimension $n-r$. Sind sie nicht eigentlich, was nur für $p = r$ eintreten kann, so handelt es sich um affine Unterräume der Dimension $n - r$, und jeder Punkt des Unterraumes ist eine Spitze.

Der Name Kegel rührt daher, daß jede Spitze M die Eigenschaft hat, daß die Verbindungsgerade mit einen beliebigen Punkt $X \in \mathcal{Q}$, $X \neq M$, ganz auf $\mathcal{Q}$ liegt. Eine solche Gerade heißt dann eine *Mantellinie* von $\mathcal{Q}$.

Umgekehrt ist jede Quadrik $\mathcal{Q}$, die einen Punkt M enthält, der $MX \subset \mathcal{Q}$ für alle

$X \in \mathcal{Q}$, $X \neq M$, erfüllt, ein Kegel (Beweis als Übungsaufgabe).

Fall (II): Hier gilt Rg $A = r$, $b = o$, $c = 1$ und der Ursprung ist ein Mittelpunkt, aber keine Spitze, also handelt es sich um echte Mittelpunktsquadriken.

Fall (III): Hier gilt Rg $A = r < n$, $b = -e_n$ und $c = 0$. Alle Quadriken sind eigentlich. Damit ein Mittelpunkt existiert, muß das LGS $A\,\hat{x} = -b = e_n$ lösbar sein. Wegen Rg $(A \mid e_n) = r + 1 \neq$ Rg A ist dies aber nicht der Fall, also existiert kein Mittelpunkt, die Quadriken sind somit parabolisch.

Als weiteres Teilergebnis können wir somit festhalten, daß affin äquivalente Quadriken $\mathcal{Q}$ und $\tilde{\mathcal{Q}}$ aufgrund ihres Mittelpunktverhaltens zur selben Klasse (I), (II) oder (III) gehören.

Um schließlich die Gleichheit von p und $\tilde{p}$ zu beweisen, zeigen wir, daß mit jeder Quadrik affine Unterräume verbunden sind, deren Dimension im wesentlichen durch diese Zahlen gegeben sind.

Satz 12. *Besitzt die Quadrik $\mathcal{Q} \subset \mathbb{A}$ die Normalform*

(I) $$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 0$$

mit $0 < p \leq r \leq n$, $r-p \leq p$, so ist $n-p$ die maximale Dimension aller ganz auf $\mathcal{Q}$ liegenden affinen Unterräume.

Beweis. Für $p = r$ ist $\mathcal{Q}$ selbst ein $(n-p)$-dimensionaler affiner Unterraum und die Behauptung ist bewiesen. Seien nun $p < r$ und L ein affiner Unterraum auf $\mathcal{Q}$ mit zugehörigem Richtungsraum U_L. Wir betrachten den affinen Unterraum $L' = L \vee O$ und zeigen, daß auch L' auf $\mathcal{Q}$ liegt:

Weil der Ursprung O Spitze von $\mathcal{Q}$ ist, liegen alle Verbindungsgeraden $O \vee X$, $X \in L$, ganz auf $\mathcal{Q}$. Weiterhin liegt jeder Punkt X aus dem zu L parallelen Unterraum durch O auf $\mathcal{Q}$. Wählen wir nämlich in L einen beliebigen Punkt P, so liegt die Gerade durch P mit Richtung $\overrightarrow{OX}$ wegen $\overrightarrow{OX} \in U_L$ in L, also auf $\mathcal{Q}$. Somit gilt für alle $t \in \mathbb{R}$

$$0 = \beta(\overrightarrow{OP} + t\,\overrightarrow{OX}\,,\ \overrightarrow{OP} + t\,\overrightarrow{OX}) = 2\,t\,\beta(\overrightarrow{OP},\overrightarrow{OX}) + t^2\,\beta(\overrightarrow{OX},\overrightarrow{OX})\,.$$

Daraus folgt $\beta(\overrightarrow{OX},\overrightarrow{OX}) = 0$, also ist $X \in \mathcal{Q}$ und es gilt somit $L' \subset \mathcal{Q}$.

Wir schneiden L' mit dem affinen Unterraum L_1, der durch das LGS

$$x_{p+1} = \ldots = x_n = 0$$

gegeben ist. Es gilt $\dim L_1 = p$ und $L_1 \cap \mathcal{Q} = \{O\}$. Damit gilt auch $L_1 \cap L' = \{O\}$. Aus der Dimensionsformel für affine Unterräume folgt nun

$$\dim L' + \dim L_1 \;=\; \dim(L' \vee L_1) + \dim(L' \cap L_1) \;\leq\; n\,,$$

also

$$\dim L \;\leq\; \dim L' \;\leq\; n - p\,.$$

Daß die Dimension $n - p$ auch tatsächlich vorkommt, zeigt das Beispiel des affinen Unterraumes $L \subset \mathcal{Q}$, der durch das LGS

$$\begin{array}{lll} x_1 & + x_{p+1} & = 0 \\ \quad\ddots & \quad\ddots & \\ x_{r-p} & + x_r & = 0 \\ x_{r-p+1} & & = 0 \\ \quad\ddots & & \\ x_p & & = 0 \end{array}$$

gegeben ist. ∎

Satz 13. *Besitzt die Quadrik $\mathcal{Q} \subset \mathbb{A}$ die Normalform*

(III) $$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 \;=\; 2\,x_n$$

mit $0 < p \leq r < n$, $r-p \leq p$, so ist $n-p-1$ die maximale Dimension aller ganz auf $\mathcal{Q}$ liegenden affinen Unterräume.

Beweis. Sei L ein affiner Unterraum, der ganz auf $\mathcal{Q}$ gelegen ist. Wir schneiden L mit dem Unterraum L_1, der durch das LGS

$$x_{p+1} = \ldots = x_r = 0\,,\ x_n = -1$$

gegeben ist. Es gilt $\dim L_1 = n + p - r - 1$ und $L_1 \cap \mathcal{Q} = \emptyset$. Damit gilt auch $L_1 \cap L = \emptyset$ und wir erhalten aus der Dimensionsformel

$$\dim L + \dim L_1 = \dim(L \vee L_1) + \dim(U_L \cap U_{L_1}) - 1$$

die Ungleichung

$$\dim L \leq r - p + \dim(U_L \cap U_{L_1}) .$$

Sei nun $Y \in L$ fest gewählt. Dann liegt für jeden Vektor $x \in U_L \cap U_{L_1}$, $x \neq o$, die Gerade durch Y mit der Richtung $[x]$ in L und somit ganz auf $\mathcal{Q}$. Für die Ortsvektoren der Punkte X auf dieser Gerade gilt $\overrightarrow{OX} = \overrightarrow{OY} + \overrightarrow{YX} = \overrightarrow{OY} + t\,x$, $t \in \mathbb{R}$.

Weil der Richtungsraum U_{L_1} durch das homogene LGS

$$x_{p+1} = \ldots = x_r = 0 \,,\; x_n = 0$$

gegeben ist, hat der Koordinatenvektor $\hat{x}$ des Vektors $x \in U_L \cap U_{L_1}$ die Form

$$\hat{x} = (x_1, \ldots, x_p, 0, \ldots, 0, x_{r+1}, \ldots, x_{n-1}, 0) .$$

Somit erfüllt für jedes $t \in \mathbb{R}$ das n–Tupel

$$(y_1 + t\,x_1, \ldots, y_p + t\,x_p, y_{p+1}, \ldots, y_r, y_{r+1} + t\,x_{r+1}, \ldots, y_{n-1} + t\,x_{n-1}, y_n)$$

die Gleichung von $\mathcal{Q}$. Speziell für $t = 1$ und $t = -1$ erhalten wir nach Addition die Gleichung

$$x_1^2 + \ldots + x_p^2 = 0 ,$$

also $x_1 = \ldots = x_p = 0$. Damit gilt

$$\dim(U_L \cap U_{L_1}) \leq n - (r+1) = n - r - 1$$

und somit

$$\dim L \leq r - p + (n - r - 1) = n - p - 1.$$

Daß die Dimension $n - p - 1$ tatsächlich erreicht wird, zeigt das Beispiel des affinen Unterraumes $L \subset \mathcal{Q}$, der durch das LGS

$$
\begin{array}{lll}
x_1 & + x_{p+1} & = 0 \\
\ddots & \ddots & \\
x_{r-p} & + x_r & = 0 \\
x_{r-p+1} & & = 0 \\
\ddots & & \\
x_p & & = 0 \\
x_n & & = 0
\end{array}
$$

gegeben ist. ∎

In den Fällen (I) und (III) folgt somit aus der Äquivalenz der Quadriken Q und $\tilde{Q}$, daß $p = \tilde{p}$ gilt.

Im Fall (II) läßt sich dieser Schluß nicht durchführen. Ohne Beweis bemerken wir, daß hier für die maximale Dimension d der auf Q gelegenen affinen Unterräume

$$d = n - p \quad \text{für} \quad p > r - p$$

gilt bzw.

$$d = n - (r - p) - 1 \quad \text{für} \quad p \leq r - p \,.$$

Aus der Gleichheit der maximalen Dimensionen d und $\tilde{d}$ für Q und $\tilde{Q}$ können wir nun nicht mehr allgemein auf $p = \tilde{p}$ schließen. So gilt zum Beispiel im $\mathbb{R}^2$ sowohl für Ellipsen als auch für Hyperbeln $d = 0$, aber im ersten Fall ist $p = 2$, im zweiten $p = 1$.

Hier müssen wir andere affine Unterräume betrachten, die mit Q affin invariant verbunden sind. Am Beispiel der Ellipsen und Hyperbeln erkennen wir, daß die maximale Dimension der Unterräume durch den Mittelpunkt O, die Q nicht treffen, 0 bzw. 1 ist, und daß diese Zahlen jeweils mit der Zahl $n - p$ übereinstimmen.

Satz 14. *Besitzt die Quadrik $Q \subset \mathbb{A}$ die Normalform*

(II) $$x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 1$$

mit $0 < p \leq r \leq n$, so ist $n-p$ die maximale Dimension aller affinen Unterräume, die den Mittelpunkt O enthalten und mit Q keinen Punkt gemeinsam haben.

Beweis. Der affine Unterraum L, der durch das LGS $x_1 = \ldots = x_p = 0$ definiert ist,

hat die Dimension $n-p$, enthält den Ursprung O und hat mit $\mathcal{Q}$ keinen Punkt gemeinsam.

Sei nun L_1 ein affiner Unterraum durch O mit $\dim L_1 > n-p$. Wir schneiden L_1 mit dem Unterraum L_2, der durch das LGS $x_{p+1} = \ldots = x_n = 0$ gegeben ist. Es gilt $O \in L_2$ und $\dim L_2 = p$. Wegen $L_1 \cap L_2 \neq \emptyset$ folgt für die Dimension des Schnittes

$$\dim (L_1 \cap L_2) = \dim L_1 + \dim L_2 - \dim (L_1 \vee L_2)$$

$$\geq \dim L_1 + \dim L_2 - n > (n-p) + p - n = 0.$$

Es gibt also in $L_1 \cap L_2$ eine Gerade OX. Für die Koordinaten von X gilt dann

$$x_{p+1} = \ldots = x_n = 0 \text{ und } x_1^2 + \ldots + x_p^2 > 0.$$

Somit liegt der Punkt Y mit dem Ortsvektor

$$\overrightarrow{OY} = 1/(x_1^2 + \ldots + x_p^2)^{\frac{1}{2}} \overrightarrow{OX}$$

sowohl in L_1 als auch auf $\mathcal{Q}$, also ist $L_1 \cap \mathcal{Q} \neq \emptyset$. Die Dimension $n-p$ von L ist daher maximal. ∎

Wegen Satz 14 folgt nun auch im verbliebenen Fall (II) für affin äquivalente Quadriken die Gleichheit von p und $\tilde{p}$. Wir fassen zusammen:

Satz 15. *Zwei Quadriken sind genau dann affin äquivalent, wenn ihre affinen Normalformen vom gleichen Typ* (I), (II) *oder* (III) *sind und außerdem in den Zahlen p und r übereinstimmen.*

Zum Abschluß dieses Paragraphen wollen wir noch die Frage klären, ob es möglich ist, eine Quadrik bezüglich eines festen Ursprungs durch verschiedene Gleichungen zu beschreiben, die sich nicht nur um einen Faktor $a \neq 0$ unterscheiden. Der folgende Satz zeigt, daß für eigentliche Quadriken diese Situation nicht eintreten kann. Er wird im letzen Abschnitt dieses Kapitels noch eine wichtige Rolle spielen.

Satz 16. *Es sei $\mathcal{Q} \subset \mathbb{A}$ eine eigentliche Quadrik. Wird $\mathcal{Q}$ bezüglich eines festen Ursprungs $O \in \mathbb{A}$ jeweils durch die Gleichungen*

$$\beta(x,x) + 2\,\Phi(x) + c = 0 \quad \textit{und} \quad \beta'(x,x) + 2\,\Phi'(x) + c' = 0$$

dargestellt, so gibt es eine reelle Zahl $a \neq 0$ mit $\beta' = a\,\beta$, $\Phi' = a\,\Phi$ und $c' = a\,c$.

Beweis. Wir unterscheiden 3 Fälle, je nachdem ob die affine Normalform der ersten Gleichung von $\mathcal{Q}$ vom Typ (I), (II) oder (III) ist.

1.Fall: Bezüglich eines geeigenten affinen Koordinatensystems $(O;v_1,\ldots,v_n)$ hat die erste Gleichung die Normalform

$$\text{(I)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 0\,, \qquad 0 < p \leq r \leq n\,,\ r-p \leq p\,,$$

und die zweite Gleichung geht über in die Form

$$\hat{x}^{\mathsf{T}} A\,\hat{x} + 2\,b^{\mathsf{T}}\hat{x} + c = 0\,.$$

Der Ursprung O ist Mittelpunkt von $\mathcal{Q}$, also ist o Lösung des inhomogenen LGS $A\,\hat{x} = -b$ (vgl. S.319). Somit gilt $b = o$. Weil O außerdem auf $\mathcal{Q}$ liegt, folgt auch $c = 0$. Wir müssen noch zeigen, daß A eine Diagonalmatrix ist mit $a_{ii} = a \neq 0$ für $i = 1,\ldots,p$, $a_{ii} = -a$ für $i = p{+}1,\ldots,r$ und $a_{ii} = 0$ für $i = r{+}1,\ldots,n$.

Da die Mittelpunkte von $\mathcal{Q}$ einen affinen Unterraum von $\mathbb{A}$ bilden, der ganz auf $\mathcal{Q}$ liegt und dessen Richtungsraum $[v_{r+1},\ldots,v_n]$ ist, gilt $A\,\hat{v}_i = o$ für $i = r{+}1,\ldots,n$, also $a_{ij} = 0$ für $i > r$ oder $j > r$. Der weitere Beweis erfolgt dadurch, daß wir die Koordinaten genügend vieler geeigneter Punkte von $\mathcal{Q}$ in die quadratische Gleichung $\hat{x}^{\mathsf{T}} A\,\hat{x} = 0$ einsetzen.

Sei $1 \leq i \leq p$, $p+1 \leq j \leq r$. Dann liegen die Punkte Y mit $\overrightarrow{OY} = v_i \pm v_j$ auf $\mathcal{Q}$. Also ist $(\hat{v}_i \pm \hat{v}_j)^{\mathsf{T}} A\,(\hat{v}_i \pm \hat{v}_j) = 0$, woraus durch Subtraktion zunächst $a_{ij} = \hat{v}_i^{\mathsf{T}} A\,\hat{v}_j = 0$ und dann $a_{ii} = \hat{v}_i^{\mathsf{T}} A\,\hat{v}_i = -\,\hat{v}_j^{\mathsf{T}} A\,\hat{v}_j = a_{jj}$ folgen. Somit gilt $a_{11} = \cdots = a_{pp} = a$ und $a_{p+1,p+1} = \cdots = a_{rr} = -a$ mit $a \neq 0$.

Für $1 \leq i < j \leq p$ liegen die Punkte Y mit $\overrightarrow{OY} = v_i + v_j + \sqrt{2}\,v_r$ auf $\mathcal{Q}$. Daraus erhalten wir unter Verwendung der schon bekannten a_{ij} in diesen Fällen $a_{ij} = 0$.

Analog zeigen wir für $p+1 \leq i < j \leq r$ mit Hilfe der durch die Ortsvektoren $\overrightarrow{OY} = \sqrt{2}\, v_1 + v_i + v_j$ gegebenen Punkte $Y \in \mathcal{Q}$, daß auch hier $a_{ij} = \hat{v}_i^{\mathsf{T}} A\, \hat{v}_j = 0$ gilt.

Damit ist A eine Diagonalmatrix von der gewünschten Form, und die Behauptung des Satzes ist in diesem Fall bewiesen.

2.Fall: $\mathcal{Q}$ ist eine echte Mittelpunktsquadrik und die erste Gleichung hat nun die Normalform

$$\text{(II)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 1\,, \qquad 0 < p \leq r \leq n\,.$$

Dieser Fall wird mit entsprechenden Methoden bewiesen. Zunächst folgt wieder $b = o$ und $a_{ij} = 0$ für $i > r$ oder $j > r$. Weil hier der Ursprung O nicht auf $\mathcal{Q}$ liegt, ist jetzt $c' \neq 0$ und wir können o.B.d.A. annehmen, daß die zweite Gleichung von $\mathcal{Q}$ bezüglich des ausgezeichneten Koordinatensystems $(O;v_1,\ldots,v_n)$ die Form $\hat{x}^{\mathsf{T}} A\, \hat{x} = 1$ hat.

Sei $1 \leq i \leq j \leq p$. Für diese i liegen die Einheitspunkte des gegebenen Koordinatensystems auf $\mathcal{Q}$, also gilt $a_{ii} = 1$. Für $i \neq j$ betrachten wir die Punkte $Y \in \mathcal{Q}$ mit den Ortsvektoren $y = 1/\sqrt{2}\ (v_i + v_j)$ und erhalten aus $\hat{y}^{\mathsf{T}}\ A\ \hat{y} = 1$ und den schon bekannten Ergebnissen $a_{ij} = 0$.

Für $1 \leq i \leq p$ und $p+1 \leq j \leq r$ wählen wir die Punkte $Y \in \mathcal{Q}$ mit $\overrightarrow{OY} = \sqrt{2}\, v_i \pm v_j$ und erhalten $a_{ij} = \hat{v}_i^{\mathsf{T}}\ A\ \hat{v}_j = 0$ und $a_{jj} = \hat{v}_j^{\mathsf{T}}\ A\ \hat{v}_j = -1$, und für $p+1 \leq i < j \leq r$ wählen wir $Y \in \mathcal{Q}$ mit $\overrightarrow{OY} = \sqrt{3}\, v_1 + v_i + v_j$ und erhalten $a_{ij} = \hat{v}_i^{\mathsf{T}}\ A\ \hat{v}_j = 0$.

Damit ist die Behauptung auch in diesem Fall bewiesen.

3. Fall: Die Quadrik ist ein Paraboloid. Die erste Gleichung besitzt die Normalform

$$\text{(III)} \qquad x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 = 2\, x_n\,, \qquad 0 < p \leq r < n,\ r-p \leq p\,,$$

und die zweite Gleichung geht wieder über in die Form

$$\hat{x}^{\mathsf{T}} A\, \hat{x} + 2\, b^{\mathsf{T}} \hat{x} + c = 0\,.$$

Wegen $O \in \mathcal{Q}$ ist $c = 0$. Weil $\mathcal{Q}$ keine Mittelpunktsquadrik ist, muß $b \neq o$ sein.

Sei nun $X \neq O$ mit zugehörigem Koordinatenvektor $\hat{x} = (x_1,\ldots,x_{n-1},0)$ beliebig gewählt und sei

$$s := x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2 .$$

Ist $s \neq 0$, so liegen alle Punkte Y mit den Koordinatenvektoren

$$\hat{y} = t\,\hat{x} + (s\,t^2)/2\,\hat{v}_n\,,\ t \in \mathbb{R},$$

auf $\mathcal{Q}$, also gilt $\hat{y}^\top A\,\hat{y} + 2\,b^\top\,\hat{y} = 0$. Somit erhalten wir für alle $t \in \mathbb{R}$:

$$\tfrac{1}{4}\,s^2\,t^4\,\hat{v}_n^\top A\,\hat{v}_n + s\,t^3\,\hat{x}^\top A\,\hat{v}_n + t^2\,(\hat{x}^\top A\,\hat{x} + s\,b^\top\,\hat{v}_n) + 2\,t\,b^\top\,\hat{x} = 0\,.$$

Daraus folgt

$$a_{nn} = 0\ ,\ \hat{x}^\top A\,\hat{v}_n = 0\ ,\ (\hat{x}^\top A\,\hat{x} + s\,b^\top\,\hat{v}_n) = 0\ ,\ b^\top\,\hat{x} = 0.$$

Ist $s = 0$, so liegt die Gerade OX ganz auf $\mathcal{Q}$, woraus

$$t^2\,\hat{x}^\top A\,\hat{x} + 2\,t\,b^\top\,\hat{x} = 0$$

für alle $t \in \mathbb{R}$ folgt. Also muß $\hat{x}^\top A\,\hat{x} = b^\top\,\hat{x} = 0$ sein.

In beiden Fällen ist $b^\top\,\hat{x} = 0$, also ist $b = a\,v_n$, $a \neq 0$.

Wäre $\hat{x}^\top A\,\hat{v}_n \neq 0$ für $s = 0$, so würde

$$\hat{y} = \hat{v}_n + \frac{a}{\hat{x}^\top A\ \hat{v}_n}\,\hat{x}$$

die Gleichung $\hat{y}^\top A\,\hat{y} + 2\,b^\top\,\hat{y} = 0$ erfüllen, der zugehörige Punkt läge also auf $\mathcal{Q}$. Dies kann aber nicht sein, da $\hat{y}$ nicht die Normalform von $\mathcal{Q}$ erfüllt, also ist auch in diesem Fall $\hat{x}^\top A\,\hat{v}_n = 0$.

Somit gilt für alle $\hat{x} = (x_1,\ldots,x_{n-1},0) \neq o$:

$$\hat{x}^\top A\,\hat{v}_n = 0 \quad \text{und} \quad \hat{x}^\top A\,\hat{x} = -a\,s = -a\,(x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2)\,.$$

Ist schließlich $\hat{x} = (x_1,\ldots,x_n)$ beliebig, so schreiben wir $\hat{x}$ in der Form $\hat{x} = (x_1,\ldots,x_{n-1},0) + x_n\,\hat{v}_n$ und erhalten ebenfalls

$$\hat{x}^\top A\,\hat{x} = -a\,(x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2)\,.$$

Somit unterscheiden sich die beiden definierenden Gleichungen von $\mathcal{Q}$ nur um einen Faktor $-a \neq 0$. ∎

§ 5 Quadriken in euklidischen Räumen

In einem euklidischen Raum spielt die affine Äquivalenz der Quadriken keine Rolle mehr, da bei affinen Abbildungen metrische Eigenschaften im allgemeinen verlorengehen; so ist zum Beispiel die Ellipse affin äquivalent zum Kreis und das Ellipsoid ist affin äquivalent zur Kugel. Da jetzt die Gestalt einer Quadrik unverändert bleiben soll, darf man unter den affinen Abbildungen nur noch die Bewegungen zulassen. Die affine Einteilung der Quadriken wird dadurch verfeinert. Wir erhalten innerhalb der Typen (I), (II), (III) eine weitere Aufteilung in Kongruenzklassen.

Die zugehörigen euklidischen Normalformen ergeben sich in der gleichen Weise wie beim Beweis von Satz 10. Wir müssen nur feststellen, welche der dort verwendeten Koordinatentransformationen Bewegungen entsprechen.

Satz 17 (Satz über die euklidische Hauptachsentransformation von Quadriken). *Es sei $\mathbb{A}$ ein n-dimensionaler euklidischer Raum. Dann läßt sich jede Quadrik $Q \subset \mathbb{A}$ bezüglich eines geeigneten kartesischen Koordinatensystems durch eine der folgenden Gleichungen beschreiben:*

$$\text{(I)} \qquad \frac{x_1^2}{a_1^2} + \cdots + \frac{x_p^2}{a_p^2} - \frac{x_{p+1}^2}{a_{p+1}^2} - \cdots - \frac{x_r^2}{a_r^2} = 0 \ , \qquad 0 < p \leq r \leq n \, , \, p - r \leq p \, ,$$

$$\text{(II)} \qquad \frac{x_1^2}{a_1^2} + \cdots + \frac{x_p^2}{a_p^2} - \frac{x_{p+1}^2}{a_{p+1}^2} - \cdots - \frac{x_r^2}{a_r^2} = 1 \ , \qquad 0 < p \leq r \leq n \, ,$$

$$\text{(III)} \qquad \frac{x_1^2}{a_1^2} + \cdots + \frac{x_p^2}{a_p^2} - \frac{x_{p+1}^2}{a_{p+1}^2} - \cdots - \frac{x_r^2}{a_r^2} = 2\, x_n \ , \qquad 0 < p \leq r < n, \, r - p \leq n \, ,$$

mit positiven reellen Zahlen $a_1, \ldots, a_r$.

Beweis. Die Quadrik $\mathcal{Q}$ sei durch die Gleichung $\beta(x,x) + 2\,\Phi(x) + c = 0$ gegeben. Wie im Beweis von Satz 10 beginnen wir mit dem quadratischen Anteil der Gleichung von $\mathcal{Q}$, beschrieben durch die Bilinearform β. Nach Wahl einer ONB in V wird β durch eine symmetrische Matrix A beschrieben. Diese läßt sich wieder mit Hilfe einer orthogonalen Matrix auf Diagonalgestalt transformieren. In der Diagonalen stehen dann die Eigenwerte $c_1,\ldots,c_n$ von A.

Für den Vektorraum V bedeutet dies die Existenz einer ONB $(v_1,\ldots,v_n)$, so daß die Matrix $\overline{A} = (\!(\beta(v_i,v_j))\!)$ Diagonalgestalt hat mit $\beta(v_i,v_i) = c_i$, $i = 1,\ldots,n$. Wir können o.B.d.A. annehmen, daß $c_1,\ldots,c_p > 0$, $c_{p+1},\ldots,c_r < 0$ und $c_{r+1} = \ldots = c_n = 0$ gilt. Dabei ist $r = \operatorname{Rg} A = n - \dim \operatorname{Rad} \beta > 0$.

Indem wir nun wieder den Ursprung O durch eine Translation geeignet verschieben, erhalten wir ein kartesisches Koordinatensystem $(O';\, v_1,\ldots,v_n)$, bezüglich dessen die Gleichung von $\mathcal{Q}$ die Form

$$c_1\, y_1^2 + \cdots + c_r\, y_r^2 + 2\, b_{r+1}\, y_{r+1} + \cdots + 2\, b_n\, y_n + c \;=\; 0$$

annimmt mit geeigneten $b_{r+1},\ldots,b_n$, einer neuen Konstanten c und Variablen $y_1,\ldots y_n$.
1. Fall: $b_{r+1} = \ldots = b_n = 0$. Dann besitzt $\mathcal{Q}$ bezüglich des zuletzt betrachteten kartesischen Koordinatensystems die Gleichung

$$c_1\, y_1^2 + \cdots + c_r\, y_r^2 + c \;=\; 0\,.$$

2. Fall: Einer der Koeffizienten $b_{r+1},\ldots,b_n$ ist von 0 verschieden. Dann ist $r < n$. Wir können o.E. annehmen, daß $b_n \neq 0$ gilt, andernfalls numerieren wir um. Jetzt wechseln wir die Basis in dem Untervektorraum $[v_{r+1},\ldots,v_n]$ und konstruieren mit Hilfe des Schmidtschen Orthogonalisierungsverfahrens, ausgehend von dem Vektor

$$v_n' \;:=\; -\frac{b_{r+1}}{b'}\, v_{r+1} - \cdots - \frac{b_n}{b'}\, v_n$$

mit $b' := (b_{r+1}^2 + \cdots + b_n^2)^{\frac{1}{2}} \neq 0$, eine weitere Orthonormalbasis $(v_{r+1}',\ldots,v_n')$ von $[v_{r+1},\ldots,v_n]$. Dann gilt für $i = r+1,\ldots,n-1$

$$v_i' = a_{ir+1}\, v_{r+1} + \cdots + a_{in}\, v_n$$

und für $i = n$

$$v_n' = -\frac{b_{r+1}}{b'}\, v_{r+1} - \cdots - \frac{b_n}{b'}\, v_n .$$

Die zugehörige Koeffizientenmatrix ist orthogonal. Setzen wir noch $v_i' = v_i$ für $i = 1,\ldots,r$, so ist insgesamt $(v_1',\ldots,v_n')$ eine ONB von V und die Gleichung von $\mathcal{Q}$ hat bezüglich des neuen Koordinatensystems $(O'; v_1',\ldots,v_n')$ die Gestalt

$$c_1 z_1^2 + \cdots + c_r z_r^2 - 2 z_n + c = 0 .$$

Verschieben wir jetzt noch den Ursprung O' um den Vektor $\frac{c}{2} v_n'$, so erhalten wir schließlich die Gleichung

$$c_1 u_1^2 + \cdots + c_r u_r^2 = 2 u_n .$$

Um die Normalformen (I), (II) und (III) zu erhalten, setzen wir im 1. Fall (für $c = 0$) bzw. im 2. Fall

$$a_i = \sqrt{1/c_i} \quad \text{für } 1 \le i \le p \quad \text{und} \quad a_i = \sqrt{-1/c_i} \quad \text{für } p+1 \le i \le r$$

und erhalten

$$\text{(I)} \qquad \frac{y_1^2}{a_1^2} + \cdots + \frac{y_p^2}{a_p^2} - \frac{y_{p+1}^2}{a_{p+1}^2} - \cdots - \frac{y_r^2}{a_r^2} = 0 , \qquad 0 < p \le r \le n ,$$

bzw.

$$\text{(III)} \qquad \frac{u_1^2}{a_1^2} + \cdots + \frac{u_p^2}{a_p^2} - \frac{u_{p+1}^2}{a_{p+1}^2} - \cdots - \frac{u_r^2}{a_r^2} = 2 u_n , \qquad 0 < p \le r < n ,$$

wobei wir $p - r \le p$ annehmen können.

Ist im 1. Fall $c \neq 0$, so können wir wieder annehmen, daß $c < 0$ gilt. Wir dividieren durch $-c$, setzen

$$a_i = \left(-c / |c_i|\right)^{\frac{1}{2}} \quad \text{für } i = 1,\ldots,r$$

und erhalten die Normalform (II). ∎

Bemerkungen und Bezeichnungen. Die Zahlen $a_1,\dots,a_r$ sind in den Fällen (II) und (III) eindeutig bestimmt, im Fall (I) erreichen wir die Eindeutigkeit durch die Normierung $a_r = 1$. Durch Vertauschen der Koordinatenachsen können wir sogar noch erreichen, daß

$$a_1 \geq a_2 \geq \cdots \geq a_p > 0\,,$$

$$0 < a_{p+1} \leq a_{p+2} \leq \cdots \leq a_r$$

gilt. Die Koordinatenachsen des kartesischen Koordinatensystems, für das die Gleichung von Q Normalform hat, heißen die *Hauptachsen* von Q. Im Fall (II) geben die Zahlen $2\,a_i$ die *Hauptachsenlängen* an.

Für $n = 2,3$ ergeben sich aus Satz 16 die folgenden euklidischen Normalformen von Quadriken.

$\boldsymbol{n = 2}$: In der euklidischen Ebene gibt es fünf Klassen von eigentlichen Kegelschnitten.

(I) $\dfrac{x_1^2}{a_1^2} - \dfrac{x_2^2}{a_2^2} = 0\,,$ (*Paar sich schneidender Geraden*)

(II) $\dfrac{x_1^2}{a_1^2} + \dfrac{x_2^2}{a_2^2} = 1\,,$ (*Ellipse, Kreis für* $a_1 = a_2$)

$\dfrac{x_1^2}{a_1^2} - \dfrac{x_2^2}{a_2^2} = 1\,,$ (*Hyperbel*)

$\dfrac{x_1^2}{a_1^2} = 1\,,$ (*Paar paralleler Geraden*)

(III) $\dfrac{x_1^2}{a_1^2} = 2\,x_2\,.$ (*Parabel*)

$n = 3$: Im dreidimensionalen euklidischen Raum gibt es elf Klassen von Quadriken, die keine Punkte, Geraden oder Ebenen sind.

(I) $\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} - \frac{x_3^2}{a_3^2} = 0$, *(Kegel)*

$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} = 0$, *(Paar sich schneidender Ebenen)*

(II) $\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} = 1$, *(Ellipsoid, Kugel für $a_1 = a_2 = a_3$)*

$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} - \frac{x_3^2}{a_3^2} = 1$, *(einschaliges Hyperboloid)*

$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} - \frac{x_3^2}{a_3^2} = 1$, *(zweischaliges Hyperboloid)*

$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1$, *(elliptischer Zylinder)*

$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} = 1$, *(hyperbolischer Zylinder)*

$\frac{x_1^2}{a_1^2} = 1$, *(Paar paralleler Ebenen)*

(III) $\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 2\,x_3$, *(elliptisches Paraboloid)*

$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} = 2\,x_3$, *(hyperbolisches Paraboloid)*

$$\frac{x_1^2}{a_1^2} = 2\, x_3 \quad . \qquad \textit{(parabolischer Zylinder)}$$

Beispiel. In einem dreidimensionalen euklidischen Raum sei bezüglich eines kartesischen Koordinatensystems eine Quadrik $\mathcal{Q}$ durch die Gleichung

$$8\, x_1^2 + 4\, x_1\, x_2 + 4\, x_1\, x_3 + 5\, x_2^2 - 8\, x_2\, x_3 + 5\, x_3^2 + 54\, x_1 + 72 = 0$$

gegeben. Wir wollen die euklidische Normalform von $\mathcal{Q}$ bestimmen.

Die Gleichung von $\mathcal{Q}$ lautet in Matrixschreibweise

$$\hat{x}^{\mathsf{T}} A\, \hat{x} + 2\, b^{\mathsf{T}}\, \hat{x} + c = 0$$

mit

$$A = \begin{bmatrix} 8 & 2 & 2 \\ 2 & 5 & -4 \\ 2 & -4 & 5 \end{bmatrix}, \quad b = \begin{bmatrix} 27 \\ 0 \\ 0 \end{bmatrix}, \quad c = 72\,.$$

Das charakteristische Polynom von A ist $p = -X\,(9 - X)^2$. Die Eigenwerte von A sind somit $c_1 = c_2 = 9$, $c_3 = 0$.

In den Eigenräumen E_{c_1} und E_{c_3} bestimmen wir jeweils eine ONB. Für E_{c_1} bilden die Vektoren

$$\hat{v}_1 = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \quad \hat{v}_2 = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix},$$

in E_{c_3} der Vektor

$$\hat{v}_3 = \frac{1}{3}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$$

eine ONB. Dann ist $S = (\hat{v}_1 |\, \hat{v}_2 |\, \hat{v}_3)$ eine orthogonale Matrix, die A auf Diagonalgestalt transformiert.

In den neuen Koordinaten y_1, y_2, y_3 , die mit den alten Koordinaten x_1, x_2, x_3 durch die Gleichungen

$$x_1 = \frac{2}{3}y_1 + \frac{2}{3}y_2 - \frac{1}{3}y_3 ,$$

$$x_2 = \frac{2}{3}y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_3 ,$$

$$x_3 = -\frac{1}{3}y_1 + \frac{2}{3}y_2 + \frac{2}{3}y_3$$

verbunden sind, hat Q somit die Gleichung

$$9\,(y_1^2 + y_2^2 + 4\,y_1 + 4\,y_2 - 2\,y_3 + 8) = 0 .$$

Mit Hilfe der Translation

$$z_1 = y_1 + 2 \;,\; z_2 = y_2 + 2 \;,\; z_3 = y_3$$

erhalten wir die euklidische Normalform der Gleichung von Q :

$$z_1^2 + z_2^2 = 2\,z_3 .$$

Q ist also ein elliptisches Paraboloid.

Euklidische Klassifikation der Quadriken

Im Gegensatz zu den affinen Normalformen, gibt es im euklidischen Fall wegen der Nebenbedingungen an die Hauptachsenlängen unendlich viele Typen in jeder Klasse. Außerdem können uneigentliche Quadriken, d.h. euklidische Unterräume, durch verschiedene Normalformen der Klasse (I) dargestellt werden. So beschreiben die Normalformen

$$x_1^2 + x_2^2 = 0 \quad , \quad \frac{x_1^2}{2} + x_2^2 = 0$$

bezüglich eines festen kartesischen Koordinatensystems im $\mathbb{R}^3$ dieselbe Gerade.

Beschränken wir uns auf eigentliche Quadriken, so gilt analog zu Satz 15 der folgende Satz.

Satz 18. *In einem euklidischen Raum sind zwei eigentliche Quadriken genau dann kongruent, wenn ihre euklidischen Normalformen vom gleichen Typ* (I), (II) *oder* (III) *sind und außerdem in den Zahlen p, r und a_i übereinstimmen, wobei im Fall* (I) $a_r = 1$ *gilt.*

Beweis. Besitzen zwei Quadriken Q und $\tilde{Q}$ dieselbe Normalform, so stimmen auch die Größen a_i überein, also bildet die Bewegung φ, die die zugehörigen kartesischen Koordinatensysteme aufeinander abbildet, auch Q auf $\tilde{Q}$ ab oder umgekehrt $\tilde{Q}$ auf Q.

Gilt $\varphi(Q) = \tilde{Q}$ mit einer Bewegung φ, so gehören Q und $\tilde{Q}$ nach Satz 15 zunächst zur selben Klasse (I), (II) oder (III) und stimmen auch bezüglich p und r überein. Es bleibt also nur noch zu zeigen, daß auch die Größen a_i und $\tilde{a}_i$ gleich sind. Dies ist aber nach Definition von a_i und $\tilde{a}_i$ genau dann der Fall, wenn die Eigenwerte c_i und $\tilde{c}_i$ der zu β und $\tilde{\beta}$ gehörenden Matrizen A und $\tilde{A}$ dieselben sind. Nach S. 306 gilt $\tilde{A} = C^T A\ C$, wobei jetzt C sogar orthogonal ist, also sind A und $\tilde{A}$ orthogonal äquivalent und besitzen somit dieselben Eigenwerte. ∎

§ 6 Tangenten und Tangentialhyperebenen von Quadriken

In diesem letzten Abschnitt wollen wir die geometrischen Eigenschaften der Quadriken noch etwas genauer studieren. Hierzu untersuchen wir zunächst die möglichen Lagen, die eine Gerade in bezug auf eine Quadrik $\mathcal{Q}$ einnehmen kann. Da der Fall, daß $\mathcal{Q}$ ein affiner Unterraum ist, schon früher behandelt worden ist, können wir hier ohne Einschränkung annehmen, daß $\mathcal{Q}$ eine eigentliche Quadrik ist. Der zugrunde liegende affine Raum kann auch euklidisch sein.

Eine Gerade g hat entweder mit der Quadrik $\mathcal{Q}$ keinen, einen oder zwei Punkte gemeinsam, oder sie ist ganz auf $\mathcal{Q}$ gelegen. Alle Fälle sind möglich, wie man an den Kegelschnitten im $\mathbb{R}^2$ sieht.

Die Schnittmenge $g \cap \mathcal{Q}$ läßt sich explizit bestimmen, indem man eine Parameterdarstellung $x = x_0 + t\,y$, $t \in \mathbb{R}$, von g in die definierende Gleichung

$$\beta(x,x) + 2\,\Phi(x) + c = 0$$

von $\mathcal{Q}$ einsetzt. Man erhält so für $g \cap \mathcal{Q}$ eine quadratische Gleichung in t,

$$(*) \qquad t^2\beta(y,y) + 2\,t\,\{\beta(x_0,y) + \Phi(y)\} + \beta(x_0,x_0) + 2\,\Phi(x_0) + c = 0\,,$$

die entweder keine oder eine oder zwei oder unendlich viele Lösungen besitzt.

Der Fall $g \cap \mathcal{Q} \neq \emptyset$

Ist $g \cap \mathcal{Q} \neq \emptyset$, so können wir die Parameterdarstellung von g so wählen, daß der Punkt X_0 mit dem Ortsvektor x_0 auf $\mathcal{Q}$ liegt. Dann nimmt die Schnittgleichung (*) die folgende einfache Form an

$$t^2\beta(y,y) + 2\,t\,\{\beta(x_0,y) + \Phi(y)\} = 0\,.$$

Für $\beta(x_0,y) + \Phi(y) \neq 0$ und $\beta(y,y) \neq 0$ gibt es zwei Lösungen. Die Gerade g hat dann mit $\mathcal{Q}$ genau zwei Punkte gemeinsam, sie ist eine *Sekante*.

Für $\beta(x_0,y) + \Phi(y) \neq 0$ und $\beta(y,y) = 0$ gibt es nur die Lösung $t = 0$ mit zugehörigem Punkt X_0. In diesem Fall heißt g eine *Einfach–Schneidende.*

Für $\beta(x_0,y) + \Phi(y) = 0$ und $\beta(y,y) \neq 0$ ist wiederum $t = 0$ die einzige Lösung (jetzt aber mit Vielfachheit 2). In diesem Fall heißt g eine *Tangente* an $\mathcal{Q}$ und der gemeinsame Punkt X_0 heißt der *Berührpunkt* der Tangente.

Für $\beta(x_0,y) + \Phi(y) = 0$ und $\beta(y,y) = 0$ erfüllt jedes $t \in \mathbb{R}$ die Schnittgleichung. Dies bedeutet, daß g ganz auf $\mathcal{Q}$ liegt. Wir nennen die Gerade g dann eine *Erzeugende* von $\mathcal{Q}$.

Bemerkung. Diese Definitionen sind von der speziellen Darstellung von $\mathcal{Q}$ unabhängig, da $\mathcal{Q}$ eine eigentliche Quadrik ist und die Gleichungen von $\mathcal{Q}$ sich dann nach Satz 16 nur um einen von Null verschiedenen Faktor unterscheiden.

Nun betrachten wir die Vereinigung $T_{X_0}(\mathcal{Q})$ aller Tangenten an $\mathcal{Q}$ mit Berührpunkt X_0 und aller Erzeugenden von $\mathcal{Q}$, die X_0 enthalten. Dann gilt:

$X \in T_{X_0}(\mathcal{Q})$

$\Longleftrightarrow$ X liegt auf einer Tangente oder auf einer Erzeugenden durch X_0

$\Longleftrightarrow$ Der Ortsvektor $x = \overrightarrow{OX}$ erfüllt die lineare Gleichung

$$\beta(x_0,x - x_0) + \Phi(x - x_0) = 0,$$

die wegen $X_0 \in \mathcal{Q}$ äquivalent ist zu der linearen Gleichung

$$\beta(x_0,x) + \Phi(x) + \Phi(x_0) + c = 0\,. \tag{**}$$

Also ist $T_{X_0}(\mathcal{Q})$ ein affiner (euklidischer) Unterraum. Er heißt der *Tangentialraum* von $\mathcal{Q}$ im Punkt X_0.

Ist die Linearform $\beta(x_0,\cdot) + \Phi$ in (**) nicht die Nullform, so ist der Tangentialraum eine Hyperebene. Sie heißt die *Tangentialhyperebene* von $\mathcal{Q}$ in X_0 und der Punkt X_0 heißt ein *regulärer Punkt* der Quadrik $\mathcal{Q}$.

Ist die Linearform $\beta(x_0,\cdot) + \Phi$ in (**) die Nullform, so ist der Tangentialraum $T_{X_0}(\mathcal{Q})$ der gesamte Raum $\mathbb{A}$, denn dann ist $\Phi(x_0) + c = \beta(x_0,x_0) + \Phi(x_0) + \Phi(x_0) + c = 0$, und jeder Vektor $x \in V$ erfüllt die Gleichung (**). Der Punkt X_0 wird dann *singulärer Punkt* von $\mathcal{Q}$ genannt.

Singuläre Punkte können nur bei Kegeln auftreten, es sind die Spitzen der Kegel. Quadriken vom Typ (II) oder (III) dagegen besitzen nur reguläre Punkte.

Beispiele. **(a)** Wir betrachten in der euklidischen Ebene eine Ellipse $\mathcal{Q}$ mit der Normalform

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1 .$$

Dann ist

$$\beta(x,y) = \frac{x_1\, y_1}{a_1^2} + \frac{x_2\, y_2}{a_2^2} , \quad \Phi = O , \quad c = -1 .$$

Die Gleichung der Tangente im Punkt $X' \in \mathcal{Q}$ lautet somit

$$\frac{x_1'\, x_1}{a_1^2} + \frac{x_2'\, x_2}{a_2^2} = 1 .$$

(b) $\mathcal{Q}$ sei eine Parabel in der euklidischen Ebene mit der Normalform

$$\frac{x_1^{\,2}}{a_1^2} = 2\, x_2 .$$

Dann gilt

$$\beta(x,y) = \frac{x_1\, y_1}{a_1^2} , \quad \Phi(x) = -\, x_2 , \quad c = 0 ,$$

und die Gleichung der Tangente in $X' \in \mathcal{Q}$ ist

$$\frac{x_1' \; x_1}{a_1^2} = x_2 + x_2' .$$

Der Fall $g \cap \mathcal{Q} = \emptyset$

Hier darf die Schnittgleichung keine Lösung haben. Für $\beta(y,y) \neq 0$ ist dies nur möglich, wenn die Diskriminante der quadratischen Gleichung (*) negativ ist. In diesem Fall nennen wir die Gerade g eine *Passante*.

Für $\beta(y,y) = 0$ muß notwendigerweise

$$\beta(x_0,y) + \Phi(y) = 0 \text{ und } \beta(x_0,x_0) + 2\,\Phi(x_0) + c \neq 0$$

gelten. Dies ist dann auch hinreichend dafür, daß g und $\mathcal{Q}$ keinen Punkt gemeinsam haben. Die Gerade g erfüllt dann die Tangentenbedingung $\beta(x_0,y) + \Phi(y) = 0$, hat aber mit $\mathcal{Q}$ keinen Punkt gemeinsam; sie heißt *Asymptote* von $\mathcal{Q}$. Asymptoten haben die Eigenschaft, daß im euklidischen Raum ihr Abstand zur Quadrik beliebig klein wird. Wir verzichten hier auf einen Beweis dieser Grenzwerteigenschaft.

Ist $\mathcal{Q}$ eine echte Mittelpunktsquadrik und ist X_0 ein Mittelpunkt von $\mathcal{Q}$, so ist $\beta(x_0,\cdot) + \Phi$ die Nullform, also ist eine Gerade $X_0 Y$ genau dann eine Asymptote von $\mathcal{Q}$, wenn $\beta(\overrightarrow{X_0 Y}, \overrightarrow{X_0 Y}) = 0$ gilt. Die Vereinigungsmenge aller Asymptoten durch X_0 ist also selbst wieder eine Quadrik, nämlich ein Kegel. Er wird auch *Asymptotenkegel* von $\mathcal{Q}$ im Punkt X_0 genannt. Gibt es nur einen Mittelpunkt von $\mathcal{Q}$, so spricht man von *dem* Asymptotenkegel der Quadrik $\mathcal{Q}$.

Beispiel. In der euklidischen Ebene besitzt die Hyperbel $\mathcal{Q}$ mit der Normalform

$$\frac{x_1^2}{a_1^2} - \frac{x_2^2}{a_2^2} = 1$$

den Asymptotenkegel

$$[\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}] \cup [\begin{bmatrix} a_1 \\ -a_2 \end{bmatrix}] .$$

Bemerkung. Die Gleichung (**) von S. 339 ist auch für $X_0 \notin \mathcal{Q}$ sinnvoll. Sie beschreibt genau dann eine Hyperebene in $\mathbb{A}$, wenn X_0 kein Mittelpunkt von $\mathcal{Q}$ ist.

Definition. Es seien $\mathcal{Q} = \{X \mid \overrightarrow{OX} = x\,,\ \beta(x,x) + 2\,\Phi(x) + c = 0\}$ eine eigentliche Quadrik und X_0 kein Mittelpunkt von $\mathcal{Q}$. Dann heißt die Hyperebene

$$\{X \mid \overrightarrow{OX} = x\,,\ \beta(x_0,x) + \Phi(x) + \Phi(x_0) + c = 0\}$$

die *Polarhyperebene* von X_0 bezüglich $\mathcal{Q}$ und X_0 heißt der *Pol* dieser Hyperebene.

Für $X_0 \in \mathcal{Q}$ ist die Polarhyperebene auch Tangentialhyperebene und der Pol X_0 ist der Berührpunkt. Hat die Polarhyperebene von X_0 mit der Quadrik $\mathcal{Q}$ einen nichtleeren Schnitt, so ist jede Verbindungsgerade von X_0 mit einem Punkt Y aus dieser Schnittmenge eine Tangente an $\mathcal{Q}$ mit Berührpunkt Y.

Satz 19 (Hauptsatz der Polarentheorie). *Es sei $\mathcal{Q}$ eine eigentliche Quadrik und die Punkte $X, Y \in \mathbb{A}$ seien keine Mittelpunkte von $\mathcal{Q}$. Dann gilt:*

X liegt genau dann in der Polarhyperebene von Y, wenn Y in der Polarhyperebene von X liegt.

Beweis. Der Beweis ist offensichtlich, da die Gleichung (**) symmetrisch in x_0 und x ist. ∎

Literaturverzeichnis

Artmann, B. : Lineare Algebra. Birkhäuser Skripten, 1989 (2. Auflage).

Beiglböck, W.D. : Lineare Algebra. Springer, 1983.

Beutelspacher, A. : Lineare Algebra. Vieweg, 2003 (6., durchgesehene Auflage).

Bosch, S. : Lineare Algebra. Springer, 2006 (3. Auflage).

Brieskorn, E. : Lineare Algebra und anlytische Geometrie I,II. Vieweg, 1983,1985.

Bröcker, Th. : Lineare Algebra und anlytische Geometrie. Birkhäuser, 2004 (2., korrigierte Auflage).

Fischer, G. : Lineare Algebra. Vieweg, 2005 (15., verbesserte Auflage).

Fischer, G. : Analytische Geometrie. Vieweg, 2001 (7., durchgesehene Auflage).

Grauert, H. , *Grunau, H.–C.* : Lineare Algebra und analytische Geometrie. Oldenbourg, 1999.

Greub, W. : Lineare Algebra. Springer, Heidelberger TB, 1976.

Heinhold, J. , *Riedmüller, B.* : Lineare Algebra und Analytische Geometrie I,II. Hanser, 1973, 1980 (1. bzw. 3. Auflage).

Heinhold, J. , *Riedmüller, B.* : Aufgaben und Lösungen zur Linearen Algebra und Analytischen Geometrie I,II. Hanser, 1977 (2. bzw. 3. Auflage).

Jänich, K. : Lineare Algebra. Springer, 2004 (10. Auflage).

Klingenberg, W. : Lineare Algebra und Geometrie. Springer, 1992 (3. Auflage).

Koecher, M. : Lineare Algebra und Analytische Geometrie. Springer, 1997 (4., erg. u. aktualisierte Aufllage).

Kowalsky, H.–J. , *Michler, G.O.* : Lineare Algebra. de Gruyter, 1998 (11. Auflage).

Lamprecht, E. : Lineare Algebra I,II. Birkhäuser, Unitaschenbücher, 1980, 1983.

Lorenz, F. : Lineare Algebra I,II. BI–Hochschultaschenbuch, 1992 (3. Auflage).

Schaal, H. : Lineare Algebra und Analytische Geometrie I,II. Vieweg, 1976.

Schaal, H. , *Glässner, E.* : Lineare Algebra und Analytische Geometrie III. Aufgaben mit Lösungen. Vieweg, 1981 (2. Auflage).

Storch, U. , *Wiebe, H.* : Lehrbuch der Mathematik, Bd II: Lineare Algebra. BI, 1990.

Strang, G. : Lineare Algebra. Springer, 2003.

Stroth, G. : Lineare Algebra. Heldermann, 1995

Wagner, R. : Grundzüge der linearen Algebra. Teubner, MLG–Reihe, 1981.

Walter, R. : Einführung in die lineare Algebra. Vieweg, 1990 (3., verb. Auflage).

Walter, R. : Lineare Algebra und analytische Geometrie. Vieweg, 1993 (2., durchgesehene Auflage).

Symbolverzeichnis

Neben den üblichen lateinischen Buchstaben werden auch Buchstaben anderer Schriften verwendet.

Griechische Buchstaben:

β (Beta) , δ und Δ (Delta) , φ und Φ (Phi) , Ψ (Psi) ,π (Pi) , σ (Sigma) , τ (Tau) , ω (Omega).

Skriptbuchstaben: $\mathcal{M}$, $\mathcal{P}$, $\mathcal{Q}$.

Symbol	Seite
$d(X,Y)$	289
$d(M_1,M_2)$	289
$x \perp y$	224
$A^{\perp}$	224
$A \perp B$	224
$L_1 \perp L_2$	288
$x \times y$	293
l^2	228
$\mathbb{A}$	277
$\dim \mathbb{A}$	278
$\varphi : V \longrightarrow W$	157
$\varphi : \mathbb{A} \longrightarrow \mathbb{B}$	296
$\overrightarrow{PQ}$	93,278
$\mathrm{TV}(P,Q;R)$	299
$L_1 \vee \ldots \vee L_k$	281
P_0P_1	281
$L_1 \parallel L_2$	129,283
$\mathrm{Rad}\ \beta$	309
$\mathcal{Q}$	304
$T_{X_0}(\mathcal{Q})$	339

Stichwortverzeichnis

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z